The Phenomena of Vagueness

by

Delia Ruby Graff

B.A., Harvard University (1991)

Submitted to the Department of Linguistics and Philosophy
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1997

Author.....

/ Department of Linguistics and Philosophy
September 5, 1997

Certified by ....

Robert Stalnaker
Professor of Philosophy
Thesis Supervisor

Accepted by......................

Alex Byrne
Chairman, Departmental Committee on Graduate Students

The Phenomena of Vagueness

by

Delia Ruby Graff

Abstract

Today, "The Sorites paradox" is used to refer to a class of paradoxical arguments having a similar form. An example is: A man weighing 100 lbs. is thin; every man who is thin will remain thin if he gains an ounce. Therefore, a man weighing 100 lbs. will remain thin if he gains 400 lbs. What makes the argument paradoxical is that while it seems both to be valid and to have true premises, it clearly has a false conclusion. It is commonly agreed that the argument shows that we should not accept the second premise—the principle that thinness always tolerates the gaining of a mere ounce. One does not solve the Sorites, though, just by giving up this "tolerance" principle. In the first chapter of the thesis, I say what more is required: 1. If the tolerance principle is not true, why are we unable to say which instance or instances of it are not true?; 2. Why are we so attracted by the tolerance principle in the first place?; 3. Can we, despite the paradox, maintain the thought that vague predicates have borderline cases, even though that thought seems to conflict with the denial of tolerance principles?; 4. Can we, despite the paradox, maintain in some revised form the thought that if two things are similar enough in a certain respect (say, weight), they will have the same semantic status with regard to a related vague predicate (in this case 'thin')?

In the second chapter of the thesis, I present and criticize some going solutions to the paradox. In particular, I argue that none of these solutions can be regarded as complete, since none provides an answer to all four of the questions set out in the first chapter. In the third chapter, I develop my own solution to the Sorites paradox: vague predicates are radically context-dependent, in the sense that they may express different properties on different occasions of use; in any context, a vague predicate expresses a property instantiated by both or neither of two things that are relevantly similar in that context. In this third chapter I show, on the one hand, how my account of the context-dependence of vague predicates contains the resources for providing a complete solution to the Sorites; and on the other hand, how many incomplete solutions could be coherently supplemented with this account. I conclude the thesis with a fourth chapter, in which I consider versions of the Sorites paradox thought to arise from the existence of phenomenal continua, and to which I do not extend the solution developed in chapter three. I argue that phenomenal continua do not provide a series of the sort required to get the paradox going—that is, a series of things, each of which looks the *same* as its neighbor, but not the same as more distant members of the series. The conclusion of the chapter is disjunctive: either there are

no phenomenal continua, or we have infinite powers of discrimination. Either way, we are permitted to accept as true the claim that if two things look the same, then if one looks red (for example), then so does the other.

Thesis Supervisor: Robert Stalnaker
Title: Professor of Philosophy

Acknowledgments

It is difficult to say when I stopped writing precursors to this essay and started writing drafts of it. But since it was later than it should have been, I must thank everyone who has commented on and criticized my written work on vagueness. Save two whom I will mention last, they are: Alex Byrne, Richard Cartwright, Matti Eklund, Michael Fara, Michael Glanzberg, Ned Hall, Richard Heck, Paul Horwich, Vann McGee, Timothy Williamson, and an audience at the fourth annual Harvard-MIT graduate conference. I want to thank especially: Michael Fara for helping with nearly every aspect of the writing of this thesis; and Timothy Williamson for saving me from making many mistakes early on, for his book (Williamson 1994) which proved to be an invaluable resource, and most importantly, for always taking the time to answer my questions and read my work, even long after his appointment as a visiting professor at M.I.T. had ended.

My thesis committte has had two supervisors: George Boolos and Robert Stalnaker. My greatest debt to is to them, not just because they were always so patient with me, probably more so than I deserved, but for teaching me so much. I have never left a meeting with either without feeling as though I'd really learned a *lot*, sometimes to the point of dizziness. With George, I'd also often leave meetings frustrated, feeling as though he had missed the point of what I was trying to say. But of course, he never did. Sometimes it would take me months, but I'd always realize eventually that he'd understood what I was trying to say better than I did myself, and was seeing ways it could be fixed where I couldn't see yet that there was a problem. I am incredibly fortunate to have gotten to know George over the past five years and have him as my teacher, and I will always regard both him and Bob as examples of the kind of philosopher I would like to be.

Contents

Chapter 1

The Sorites Paradox

Take any heap of sand. If you remove just one grain from it, you will still have a heap. It seems crazy to deny this—for it seems that a thing could not be a *heap* unless it were large enough to remain a heap if made just slightly smaller. If we accept this, however, we commit ourselves to the following principle: For any number n, if $n + 1$ grains of sand can form a heap then so can n grains. Unfortunately, this seemingly incontrovertible principle embroils us in paradox:

> 10,000 grains of sand can form a heap
> For any n, if $(n + 1)$ grains of sand can form a heap then so can n grains
>
> ---
>
> Therefore, 1 grain of sand can form a heap

Here we have a valid argument with a false conclusion, and so it cannot be that both of the premises are true. What makes it a paradox is that both of the premises certainly *seem* true. It is generally agreed that the second premise is the culprit, that our seemingly incontrovertible principle is not true. (Following Crispin Wright, we will call the second premise a principle of *tolerance* for 'heap'.[1]) But in order to dispel the paradox, we have to do more than just point our finger at the guilty premise. What more *is* required? I provide

---

[1](Wright 1975, 333f.). The idea is that the applicability of 'heap' to a collection of grains of sand could always survive, or *tolerate*, a unit decrease.

an answer to this question in the second section of this chapter. First, however, I wish to consider the general question of what it is to solve a paradox. On the one hand, this will provide us with a useful frame of reference when we proceed to discuss what is required of a solution to the Sorites in particular. On the other hand, by looking at a few other paradoxes, I aim to demonstrate that the general question has no answer, and hence that it cannot be taken for granted that it is obvious what the answer to the particular question is. It is my view that all attempted solutions to the Sorites paradox suffer from a failure to identify at the outset what questions are being provided answers.

1.1   Solving Paradoxes

What are we to do when confronted with a paradox? Set out to solve it, of course. But what is it to "solve" a paradox? I am unable to find a univocal answer to this question. It seems that what counts as solving a paradox varies widely from case to case, since not all paradoxes are created equal. Consider the *care-taker paradox*:

> There seems no harm in supposing that for any property a person might have, it would be possible for someone to take care of just those people who have that *property. So there seems no harm in supposing it possible that Kind Kate takes* care of just those people who do not take care of themselves. Now does Kind Kate take care of herself or not? Well, if she does she doesn't and if she doesn't *she does. So we are in a bind, because that's not possible.*

It's not much of a bind, though. We "solve" this paradox just by giving up our initial universal principle ("For any property a person might have ..."), which may be called an *abstraction* principle for care-takers. There is no psychological barrier to giving up the principle—as there is in the case of the tolerance principle—since the care-taker paradox, far from leaving us puzzled, *shows* us why we should give up the abstraction principle by handing us precisely that instance of it (the property of not taking care of oneself) which gets us into trouble.

Something similar may be said for the Liar paradox. The starting principle is this: For any indicative sentence p, 'p' is true if and only if p.[2] (We will call this a *disquotation* principle for truth.) But consider the Liar sentence: 'The Liar sentence is not true'. Now is this sentence true or not? Given the disquotation principle plus the identity of the Liar sentence and 'The Liar sentence is not true', we quickly conclude that the Liar is true if and only if it's not true. Thus we must give up our init.al universal principle and, as with the care-taker paradox, the present paradox shows us why. The Liar paradox, like the care-taker paradox, but unlike the Sorites paradox, not only demonstrates *that* its starting universal principle cannot be true, it also provides us with a trouble-making instance, and hence a way of *seeing* that its starting universal principle is not 'rue.

But still, there is an important contrast between the Liar and the care-taker: we do not *solve* the Liar paradox just by giving up the disquotation principle, in the way that we solve the care-taker paradox just by giving up the abstraction principle. One reason for this contrast in what's to count as a solution is that the disquotation principle seems to have a fundamental status as reflecting our understanding of what truth is. If we merely relinquish the principle, we relinquish that understanding. Thus part of what's required to solve the Liar is to investigate and propose consistent alternatives to the disquotation principle; to say what relationship any such alternative bears to our original understanding of truth; or else to argue, though this seems desperate, that we had no coherent understanding of truth, or worse, that there is no coherent notion of truth to be made out. Because the abstraction principle for care-takers has no similarly fundamental status for our understanding of *anything*, abandoning the principle imposes on us no further philosophical burden.

There might be a second reason for the contrast between the Liar and the caretaker: a reason deriving from a difference in the extent to which the two paradoxes lead us to question classical principles of reasoning. Let me explain. Letting 'k' stand for Kate, 'R' for the *takes care of* relation, and 'L' for the Liar sentence, i.e. '$\neg Tr(L)$', we cannot

---

[2]In order for my formulation of the principle to be well-formed, the universal quantifier must be understood as a substitutional one; for it to be plausible, the pair of quotation marks around the substitutional variable must be understood as a term-forming functor.

accept that kRk iff $\neg kRk$ or that $Tr('\neg Tr(L)')$ iff $\neg Tr(L)$, since these are inconsistent; and because they are inconsistent, it seems we *should* accept their negations, which are classically equivalent to instances of the Law of Excluded Middle, namely, to '$kRk \vee \neg kRk$' and '$Tr(L) \vee \neg Tr(L)$', respectively.[3] Now in the case of Kind Kate, it is not hard to accept that she either takes care of herself or she doesn't, because it is not hard to see how one or the other disjunct could be true. But in the case of the Liar, it *is* hard to accept that it is either true or it isn't, because it is very hard to see how it could be the one, or how it could be the other. Surely, we should reason, the Liar cannot be *true*, for if it were true then since it says that it is not true, it wouldn't be. But conversely, it is difficult to accept that the Liar is *not* true because then, it seems, we would be accepting the Liar itself, and would that not just be to accept the Liar as true? Maybe not—maybe we could learn to live with the idea that the Liar, after all, is not true. But short of doing that, we must let go of some of our classical principles. The problem posed by the Liar is to say which, if any, should give.

In another sort of case, just identifying the fallacy involved in a paradox would count as providing a solution to it. I have in mind, as an example, the paradox of the surprise examination:

> A teacher informs his class that at some point during the next month, a surprise examination will be given. The students decide that, having been so informed, no such examination can be given. They reason as follows: the exam cannot be on the last day, for then we would know on the night before that there would be an exam the following day, and hence it would not be a *surprise*. But then it cannot be on the second to last day either, for then, knowing the exam cannot be on the last day, we would know on the night before the second to last day that there would be an exam the following day, and hence *it* would not be a surprise. They reason similarly for each of the remaining days, continuing in reverse order. But then, come the twelfth day, the teacher passes out an exam and, lo and behold, the students are surprised. What, if anything, was wrong with their reasoning?

---

[3]For the equivalence of '$Tr(L) \vee \neg Tr(L)$' and the negation of '$Tr('\neg Tr(L)') \leftrightarrow \neg Tr(L)$', we make use of the identity of L and '$\neg Tr(L)$'.

In the case of the care-taker and Liar paradoxes there could be no question that the abstraction and disquotation principles were the source of contradiction, since the reasoning required to get from these principles to contradiction was so meager, explicit, and unimpeachable.[4] With the surprise exam paradox, however, the reasoning is more complex, more subtle—it is not clear just what principles are involved—and, moreover, it is not strictly logical. It suffices to solve the surprise exam paradox if one identifies the principles involved in the students' reasoning, and says which of those principles is flawed. Of course by this, I do not mean that it is sufficient, having identified the principles involved, merely to *state* which principle is flawed. One must offer a reasoned defense of the claim—one must explain why the flawed principle may not be employed in the envisioned circumstances. It could in the end turn out that the surprise exam paradox is like the Liar paradox in the respect that it could just be that the flawed principle employed by the students (whichever one it is) is so fundamental to our understanding that, like the disquotation principle, it could not be given up without saying what is to replace it. But that remains to be seen.

In any case, the foregoing considerations suggest that there is no one thing which counts as *solving* a paradox. We will say that by a *paradox* we mean a seemingly valid argument, with seemingly true (or mutually consistent) premises but a seemingly false (or contradictory) conclusion. Perhaps standardly one thinks of solving a paradox as *finding the mistake*, as providing an answer to the question 'What's wrong with this argument?'. The surprise exam paradox may fit this mold, but the other paradoxes so far mentioned do not. The arguments that constitute the care-taker, Liar and Sorites paradoxes are each so bare-boned that one "finds" the mistake straight away.

It will not do to revise the standard thought by saying that in order to solve a paradox one must find the mistake *and* resolve any puzzles which consequently arise; such a revision would be too demanding. Above I claimed that it is not difficult to accept that Kate

---

[4]In the case of the care-taker, we appealed only to universal instantiation (including UI in the scope of an existential quantifier); in the case of the Liar, we appealed only to universal instantiation and the substitutivity of identity.

either takes care of herself or she doesn't (an equivalent of the negation of the contradictory 'kRk iff $\neg kRk$') because it is not difficult to see how it could be the one or the other. But this downplayed the problem, for it *is* difficult to accept that it *must* be one or the other. The *takes care of* relation has borderline cases, and for that reason, it seems, it could just be that neither 'Kate takes care of herself' nor 'Kate does not take care of herself' is correct. This is a puzzle—a puzzle which, moreover, arises upon reflection of the care-taker paradox. Yet one need not address this puzzle in order to solve the care-taker. For some reason, it just seems unrelated.

It will now be tempting to say that in order to solve a paradox, one must find the mistake and resolve any *related* puzzles which consequently arise. Even if right, this would be singularly uninformative: in part because it will not always be clear whether a given puzzle is *related* to a given paradox; and in part because it will not always be clear whether, when a given puzzle arises, it is *consequent* upon finding the mistake in a given argument. In any case, I believe this thought is wrong. We have already seen that finding the mistake in a paradoxical argument is not always sufficient for solving the paradox, as it is not in the version of the Liar paradox I set out above. Below it will also emerge that finding the mistake in a paradoxical argument is not always *required* for solving it, for there are formulations of the Sorites paradox in which the mistaken premise cannot be found. I will not consider any further the univocity of paradox solving, but instead will get down to business and commence discussion of what I do think solving the Sorites involves.

1.2   Solving the Sorites

I would like to begin with a little compare and contrast.

I.   First, it is worth repeating, the Sorites paradox shares a feature with the Liar and care-taker paradoxes which the surprise exam paradox does not, namely, that the work of identifying the flaw in its argument comes virtually for free. In light of the Sorites argument (p. 9 above) it is clear that the tolerance principle for 'heap' cannot be true. I *know* that 10,000 grains of sand can form a heap, (and if you are unsure of this, pick any number you

prefer). But this, together with the tolerance principle, and the aid of 9,999 applications of Universal Instantiation and Modus Ponens, licenses the conclusion that one grain of sand can form a heap.[5] The conclusion is not licensed, however, and since MP and UI are surely valid,[6] it must be that the tolerance principle is not true.

II. Restricting our attention now to just the care-taker, Liar and Sorites paradoxes, we may note a second point of contrast. The abstraction, disquotation and tolerance principles are each, as I have formulated them, universal generalizations, none of which can be true. Presumably, then, not every instance of these generalizations is true.[7] For the abstraction and disquotation principles, we have already got our hands on an untrue instance; for the tolerance principle, we do not.

But could we find an untrue instance? We would begin by limiting our search. Start with some obvious facts about heaps of sand: some numbers of grains, for example one grain, are too few to form a heap; other numbers of grains, perhaps a trillion grains, are too many to form a heap, but would form instead a mountain; if a number falls into this latter category, then so will any number greater than it; meanwhile, 10,000 falls into neither category, since 10,000 grains can form a heap. From these facts it follows that any untrue instance of the tolerance principle must also be an instance of the *bounded* generalization 'For any $n < 10,000$, if $n + 1$ grains of sand can form a heap then so can n grains'. We could no doubt choose a lower bound than this, it doesn't matter. What does matter is that the existence of a bound is in an important way misleading. It suggests that there is some specific finite task we could perform, upon completion of which it would be decided whether an untrue instance were discoverable by us or not. It suggests, perhaps, that we need only assess each of the ten-thousand instances of the bounded generalization, and when at the

---

[5]The proof need not be so long, however. As George Boolos (1991) has pointed out, if we help ourselves to two further inference rules—Universal Generalization and transitivity of the conditional, the length of the derivation can be reduced to fewer than seventy lines.

[6]In fact we've only employed a restricted form of Modus Ponens, one whose conditional premise contains no conditionals as constituents—hence one not prey to McGee's (1985) counterexamples.

[7]Kamp (1981) actually rejects this. He proposes a semantics for vague predicates according to which *false* universal generalizations may have none but true instances.

end of this assessment we have not found an untrue instance, it will have been shown that we are unable to find one. But this is not so. Although there is a bound on the number of instances we need consider, there is no bound on the number of times or the length of time we need consider them. Who knows, maybe after the fifth run, or after the sixth, we would hit upon that instance which just struck us, glaringly, as untrue.

If we suppose, however, as it seems reasonable to suppose, that we are in fact unable to discover an untrue instance, then we have a problem. For then it would be a simple matter to formulate an intractable version of the Sorites paradox by replacing the unbounded tolerance principle in the Sorites argument with the finite set of instances of the bounded universal above,[8] yielding:

> 10,000 grains of sand can form a heap
> If 10,000 grains of sand can form a heap then so can 9,999 grains
> If 9,999 grains of sand can form a heap then so can 9,998 grains
> If 9,998 grains of sand can form a heap then so can 9,997 grains
>
> $\vdots$
>
> If 3 grains of sand can form a heap then so can 2 grains
> If 2 grains of sand can form a heap then so can 1 grain
> _____
>
> Therefore, 1 grain of sand can form a heap

This version of the paradox would be intractable in the sense that we would be unable to find the mistake—to say which premise was not true. Thus the similarity in the first respect mentioned, of the Sorites to the Liar and care-taker, is as superficial as its *dis*similarity to them in the second respect is deep.

What is wanted, if the Sorites paradox is intractable, is an explanation of our inability to find an untrue instance of the tolerance principle. This explanation is wanted for two reasons. On the one hand, such an explanation would provide a justification for philosophical approaches to the Sorites. For if it *were* possible for us to discover which instance is untrue, then the right way to go about it, since none of the instances is contradictory, would be to head to the beach and start heaping. Without an advance explanation of why

---

[8]We will, in so doing, dispense with the need for UI.

such an experiment would be unsuccessful, we have no excuse for not going out and getting our hands dirty. The fact that we have not yet found an untrue instance is by itself no guarantee that we will not in the future.

On the other hand, the explanation is wanted just because our inability to find an untrue instance is wholly mysterious without an explanation, and should cast doubt on the idea that we really know what 'heap' means. One would have thought that 'heap' is a word designed by humans to make classifications for human purposes; that in particular, it was designed expressly to require only "casual observation" (as Wright put it),[9] and never investigation, for deciding its applicability; and that consequently, for any given n, the question whether n grains of sand can form a heap could never be a matter of something hidden from our view. One would have thought, moreover, that knowledge of the meaning of 'heap' requires knowledge that its applicability can be decided by only casual observation— someone who uses a microscope or other such aid to decide whether a thing is a heap, even if he always gets it right, cannot be said to know the meaning of the word.[10] Assuming, however, that we can know what 'heap' means, such knowledge cannot require it to be *known* that: for any given n, the question whether n grains of sand can form a heap could never be a matter of something hidden from our view; since if the Sorites is intractable, as we are supposing, this is not even true. For if we could correctly decide whether any such 'heap' sentence were true, then we could also correctly decide whether any complex sentence built up from such 'heap' sentences and logical connectives were true. But this is just what we are supposing not to be the case. I will consider an explanation of our inability to find an untrue instance of 'For any $n < 10,000$, if $n + 1$ grains of sand can form a heap then so can n grains' to be a required component of any solution to the Sorites. I will also take this to involve explanation of how that inability is compatible with the possibility of our knowing what 'heap' means.

III. I would like to discuss one final point of comparison. The abstraction and disquotation

---

[9](Wright 1975, p. 335).

[10]This is somewhat of an overstatement. The use of some aids, for example binoculars, may be appropriate on some occasions, for example when the thing is far away.

principles enjoy some initial appeal, but in light of the paradoxes it is not difficult to accept that while they hold good in many instances, the principles do not hold good in general.[11] The appeal of the tolerance principle, in contrast, seems to be unrelenting. The resistance to giving it up has led some to rather bizarre views, for example:

- The tolerance principle is true; what is not true is that there are heaps! (Unger 1979)
- Well, the tolerance principle is not true, but all of its instances are! (Kamp 1981)
- The tolerance principle is true, and the Sorites paradox is a "genuine" paradox (Dummett 1975)—by which it is difficult to understand Dummett as meaning anything other than a valid argument with true premises and a false conclusion! .

Why is the appeal of the tolerance principle so great? Why does the idea that minute changes in size cannot affect the applicability of 'heap' have such a grip on us? I will consider the provision of an answer to this question to be the second required component of any solution to the Sorites. The hope is that we would be able to rid ourselves of the compulsion to believe tolerance principles, if we could identify the source of our attraction to them.

That ends the comparison. We are not yet finished, though, with our discussion of what is involved in solving the Sorites. A final question to be contended with is, what are we to say of the *denial* of the tolerance principle? If the Sorites paradox leads us to give up the tolerance principle, then it should also command our acceptance of its negation. The reason for this is not that our beliefs must be negation complete—it is not that for every claim, we must believe either it or its negation. Nor is it that in every case where we have reason to give up a belief we once held, we thereby acquire reason to believe the contrary. (We may have reason to give up a belief we once held when we discover, for example, that

---

[11]To find trouble-makers for the disquotation principle, one needn't turn to such exotic sentences as the Liar. 'I am hungry' (uttered now by you) may be true even though I am not hungry. Additionally, as Kripke (1975) has emphasized, whether a sentence is Liar-like or not may depend not just on its form, but also on the facts: Whether 'Everything Nixon said was true' is Liar-like or not depends on what Nixon said. (Suppose Nixon said he once lied.)

our grounds for it were not good grounds.) We are led to give up the tolerance principle, though, because it proves *incompatible* with plain truths: that 10,000 grains of sand can form a heap and that one grain cannot. It is for that reason that we should not only give up the tolerance principle but also outright deny it.

The denial of the tolerance principle, however, is apparently equivalent to the claim that the predicate 'heap' has, as we shall say, *sharp boundaries*—that there is a number n such that $(n + 1)$ grains of sand can form a heap but n grains cannot, which on the face of it seems absurd. How could there be a heap that one could turn into a non-heap just by removing a tiny little grain of sand from it? How could the boundary between being a heap and not being a heap be so precise, so *sharp*?

Suppose you have a heap of sand and remove the grains from it gingerly one by one. Call the result of each removal a *stage*. We are supposed to accept that at some stage we will have a heap, while at the next stage we will not. There are at least two separate reasons we find it difficult to accept the existence of such a stage. I think it important to distinguish them. One reason for the difficulty is that it seems that before getting to a non-heap stage, we will pass through some stages where it is *indeterminate* whether the thing before us is a heap, where there is "no fact of the matter" as to whether the thing before us is a heap, where what we have before us is a *borderline case*. A second reason for the difficulty, not surprisingly, is just our original affinity for tolerance—our belief that if you have a heap at any stage, you will have a heap at the next stage too, not a borderline case, but a *heap*. If we line up the stages in order, then the first thought is that *between* the heaps and the non-heaps will be some things of indeterminate status; the second thought is that any pair of adjacent stages must have the *same* status, whether it be *heap*, *non-heap* or *indeterminate*. These two thoughts conspire to make the sharp-boundaries claim doubly difficult to accept.

Consider an analogy. Suppose someone walks into a third-grade classroom, lines up all the people in it, then tells you that somewhere in the line a girl is standing next to a boy. (We assume there are both boys and girls in the classroom.) You would find it difficult

to accept this claim if you believed, by analogy with the first thought, that the *teacher* was standing *between* the girls and the boys. By analogy with the second thought, you would find it difficult to accept the claim if you believed that anyone standing next to a girl was a girl.

The analogy makes it clear, if it was not clear already, that the second thought cannot be maintained. Given any finite series, and any kinds of status, if any two members of the series do not have the same status, then it cannot be that adjacent members always do have the same status.

There is nothing so *patently* problematic with the first thought, however—with the thought that it can sometimes be indeterminate whether a thing is a heap, and that as you remove the grains one by one from a heap you will get to such things before you get to the things that are not heaps. A pressing question is whether, despite the Sorites paradox, this thought can be maintained. On some accounts we are not required, despite the paradox, to accept the claim that for some n, $(n + 1)$ grains can form a heap while n grains cannot, *because* 'heap' has borderline cases—the existence of borderline cases renders the sharp boundaries claim untrue. On other accounts we are required to accept the claim, but in one way or another it is argued that this is compatible with the predicate's having borderline cases.

There is still another pressing question, less frequently addressed by those who now write about the paradox: the question whether some revised version of the *second* thought can be maintained. It cannot be that heap stages in our series are always directly followed by heap stages. But could there be some *kernel* of truth in the idea? Some acceptable alternative to it which can be recovered? It would be easier to give up the second thought if we could defend an affirmative answer to this question. A revision of the the second thought might take the form of a qualification: All adjacent pairs of stages in our series that meet condition C always have the same status. Another candidate form for the revision is: Every stage bears some special relation R to adjacent stages, which it does not bear to more distant stages—where R should be in some way *like*, though not identical to, 'has the

same status as'. There is no guarantee that a tenable such revision is to be found, but any complete solution to the Sorites should involve a search for one.

Chapter 2

Competing Accounts of Vagueness

In this chapter, I present some going solutions to the Sorites. Before proceeding, however, we should review the list of questions which I claim should be addressed by any complete solution to the Sorites.

> 1. If the tolerance principle is not true, why are we unable to say which instance or instances of it are not true?

> 2. Why are we so attracted by the tolerance principle in the first place?

> 3. Can we, despite the paradox, maintain the thought that 'heap' has borderline cases, even though that thought seems to conflict with the denial of the tolerance principle?

> 4. Can we, despite the paradox, maintain in some revised form the thought that adjacent stages in our series always have the same status?

The accounts to be presented in this chapter share a common feature. Each of them proposes an affirmative answer to question 3. above. That is, each of them upholds, in one way or another, the claim that 'heap' has borderline cases. What the different accounts provide are different ways of understanding the claim that do not lead us back into the paradox. The claim is always assumed to be defensible, since it is assumed that to deny that 'heap' has borderline cases would just be to deny that it is vague. Russell (1923, p. 148) said that the word 'red' is vague since "there are shades of colour concerning which we shall be in doubt whether to call them red or not, not because we are ignorant of the meaning of the word 'red', but because it is a word the extent of whose application

is essentially doubtful." Borderline cases have been taken to be the defining mark of a predicate's vagueness ever since.

After noting the "essential doubtfulness" of certain predications of 'red', Russell immediately went on to say:

> This, of course, is the answer to the old puzzle about the man who went bald. It is supposed that at first he was not bald, that he lost his hairs one by one, and that in the end he was bald; therefore, it is argued, there must have been one hair the loss of which converted him into a bald man. This, of course, is absurd.

No contemporary philosopher follows Russell in taking the paradox to be so easily dispensed with—to be resolved just by noting that predicates subject to the paradox have borderline cases. Nevertheless, none of the accounts to be considered does provide a complete solution to the Sorites in my estimation, since none of them addresses all of questions 1–4. Moreover, as we shall see, they do not even address the same subset of them.

2.1 Truth-Value Gaps and Degrees of Truth

What do we mean when we say that an object is a *borderline case* of a predicate? To pick a concrete example, consider the cover of Dummett's *The Logical Basis of Metaphysics*. (We'll call it *LBM*.) It seems to be a borderline case of 'red'. What does this mean? I take it that the phrase 'borderline case' is not in need of *definition*, since I take it that the the phrase is in common usage. We are not using it as a technical term. The phrase is in need of some analysis, however. As Russell put it, if *LBM* is a borderline case of red, it is because it is "essentially doubtful" whether the applicability of 'red' extends to it. His use of epistemological vocabulary is somewhat misleading. Russell wanted to convey that there is more to *LBM*'s being a borderline case for 'red' than its being unknowable whether it is red. He wanted to convey that here there is no fact to be known. He might have argued as follows: the meaning of 'red' is such that anyone who knows it can tell by looking, if a thing is red, that it is red (assuming visual faculties, lighting conditions, etc. to be favorable); or if it is not red, that it is not red. If when looking at *LBM* we who know what 'red' means

cannot tell whether it is red, then it cannot be a fact that *LBM* is red, and it cannot be a fact that *LBM* is not red; were either a fact, we would be able to tell. Neither '*LBM* is red' nor '*LBM* is not red' is true. So taking the falsity of p to be the truth of its negation, neither sentence is either true or false.

The argument is not confined to observational predicates—predicates whose applicability to a thing depends only on the thing's appearance. Whether 'old' applies to a thing has nothing to do with its appearance, but the argument may be extended to 'old' as follows: the meaning of 'old' is such that anyone who knows it can tell whether a man is old just by learning his age. If you who know what 'old' means cannot tell whether a man is old, even though you know his age, then it cannot be a fact that he is old, nor a fact that he is not old. Were either a fact, you would be able to tell. The argument is rather simplified, since to know whether a thing is old, one has to know not only its age but also general facts about the typical life-span and stages of development of things of its kind, and perhaps other things as well. But the crux of the argument is that once one is in possession of certain sorts of information, one will be in a position to tell whether a thing is old, in the sense that rational reflection on one's information will be sufficient for correctly deciding the question, given knowledge of what 'old' means. Once one is informed, for example, about the degrees of the angles of a four-sided figure, one will be in a position to tell whether the figure is square. Given what is meant by 'in a position to tell', the following is a tautology: if I am in a position to tell whether a thing is old, but still no amount of reflection would enable me to decide that it is old or to decide that it is not old, then it cannot be true that the thing is old, and it cannot be true that the thing is not old. Thus the argument may be extended to any predicate F, as long as one can be in a position to tell whether F applies to some thing, without being able to decide that F does or that it doesn't apply to that thing.

What we now want to know is, how harmless was our simplification? If being informed about a thing's age does not place one in a position to tell whether it is old, (just as being informed about a man's *income*, for example, does not really place one in a position

to tell whether that man is rich--maybe he's paying off huge debts), exactly what other sorts of information are required? Can we be sure that it is indeed possible for someone to be in a position to tell whether a thing is old without being able to decide that it is or that it isn't, when we have not yet specified what sorts of information do place one in a position to tell whether a thing is old? It will not do to say that even in absence of such a specification we can be assured of the envisaged possibility, since in some cases we find ourselves unable to decide whether a thing is old, and no amount of new information would help us to decide. It will not do to say this, unless one can argue against the possibility that in such a situation we are faced with an unknowable truth. The challenge, to my knowledge, has not been met. [1]

Nevertheless, there has been fairly wide-spread consensus among those who write about vagueness that the attribution of a predicate to its borderline cases does lead to truth-value gaps—by which I just mean sentences that are neither true nor false. Proponents of this view must say whether such sentences are to be deemed as having some value other than *true* or *false*; if so, whether there is more than one such value; and in addition, what is the value of a complex sentence containing a gappy sentence as a constituent.

A basic approach would be to say that there is exactly one more value, *Indefinite*, and that the value of a complex sentence is given by filling in each of the blanks newly added to the classical two-valued truth tables with one of True, False or Indefinite:

p	$\neg p$
T	F
I	
F	T

&	T	I	F
T	T		F
I			
F	F		F

∨	T	I	F
T	T		T
I			
F	T		F

→	T	I	F
T	T		F
I			
F	T		T

Clearly, there are many ways the tables could be filled in, and some ways are more interesting than others.[2] Commutativity and associativity, for example, are desirable properties for both conjunction and disjunction, but there are ways of filling in the tables that

---

[1] For a pessimistic view, see Williamson (1996).

[2] For a useful survey of many alternatives that have been adopted, see Bolc & Borowik (1992, ch. 3).

do not respect these properties.[3] If the properties are respected, then the basic approach could be extended to a first-order language by letting the value of a universally (existentially) quantified sentence be equivalent to the value of the conjunction (disjunction) of its instances.

Although there are many ways the tables could be filled in, we may broadly distinguish those which let indefiniteness trump, as in the first group of tables below, from those which let truth or falsity trump where appropriate, of which the second group of tables[4] provides a representative example.

I.

p	$\neg p$
T	F
I	I
F	T

&	T	I	F
T	T	I	F
I	I	I	I
F	F	I	F

∨	T	I	F
T	T	I	T
I	I	I	I
F	T	I	F

→	T	I	F
T	T	I	F
I	I	I	I
F	T	I	T

II.

p	$\neg p$
T	F
I	I
F	T

&	T	I	F
T	T	I	F
I	I	I	F
F	F	F	F

∨	T	I	F
T	T	T	T
I	T	I	I
F	T	I	F

→	T	I	F
T	T	I	F
I	T	I	I
F	T	T	T

The first group of tables seems appropriate when indefiniteness is understood as the value assigned to a sentence that does not express a proposition, because it contains a nonsense word, say, or a non-referring term. If a sentence is defective in one of these ways, it will infect any sentence containing it as a part. 'Grass is green or flook is blandy' is no more meaningful than 'flook is blandy' on its own.

The second group of tables seems more appropriate when dealing with the sort of indeterminacy thought to arise from vagueness, however. If Kate is kind, then 'Kate is kind or Kate is funny' should be true even if she is only borderline funny. The truth of the first disjunct is sufficient for the truth of the whole. There seems no reason to relegate the disjunction to indefiniteness.

---

[3]An operator ∗ is commutative if the value of $A ∗ B$ is always equal to that of $B ∗ A$; it is associative if the value of $(A ∗ B) ∗ C$ is always equal to that of $A ∗ (B ∗ C)$.

[4]Due to Kleene (1952).

However the tables are filled in, what is distinctive about this basic approach is that it admits exactly one value in addition to *true* and *false*, and that it treats the logical connectives (and in an extended sense also the quantifiers) as *truth-functional.*[5] Both of these features may be questioned.

One popular account retains the second feature while jettisoning the first by admitting not just three values, but a continuum of values—represented by the real numbers from 0 to 1.[6] How are we to understand these different values? Goguen (1967) gives some indication:

> ... $S(x)$ is the truth of the statement 'x is short'. So $S(x) = 1$ means 'x is short' and $S(x) = 0$ means 'x is not (at all) short', while $S(x) = \frac{1}{2}$ means 'x is half-short' or ' "x is short" is half-true'.

If we take a series of men ranging from five-feet to six-feet tall, each man one millimeter shorter than his successor in the series, then the idea is that as we move along the series, at some point the value of 'x is short' will begin to gradually decrease—it will get less and less true. The value of 'x is short' will be 1 when the first man in the series is assigned to x and 0 when the last man is. The drop in value witnessed in one move from any man to the next will be at most some very small number ε. (In this case ε might equal $\frac{1}{100}$.) Before we can assess how well this account provides resolution of the Sorites paradox, we must fill in some more details.

On this account, the logical connectives will be degree-functional. The value of a disjunction $p \vee q$, for example, will be a function just of the *values* of p and q, no matter which sentences p and q might be. What we want to know is, *which* function. As in the three-value case, there are again many choices, but there are also plausible constraints. One important constraint is that the functions should *extend* the classical two-valued truth tables

---

[5]To say that conjunction, for example, is truth-functional is to say that if sentences A and A' have the same value and B and B' have the same value, then the conjunction of A and B has the same value as the conjunction of A' and B', whatever A, B, A' and B' may be.

[6]The account to be presented is derived primarily from Łukasiewicz & Tarski (1930), Goguen (1967), Lakoff (1973), and Machina (1976). Unlike Goguen, Lakoff and Machina, Łukasiewicz did not consider how infinite-valued logic might be applied to the problem of vagueness. He also did not take the values to be the real numbers, but rather the rational numbers between 0 and 1.

in the sense that when restricted to the classical values "complete truth" (1) and "complete falsity" (0) they should yield classical results. Again, commutativity and associativity of conjunction and disjunction are also desirable. These constraints do not completely restrict our choice, but further considerations may be brought to bear.[7] We will not go into detail about what these considerations might be, but just present those clauses which are often considered to be the most plausible starting point.

$$[\neg p] \; = \; 1 - [p] \qquad\qquad [p \lor q] \; = \; max\{[p],[q]\}$$
$$[p \,\&\, q] \; = \; min\{[p],[q]\} \qquad [p \to q] \; = \; min\{1, 1 - ([p] - [q])\}$$

$$[\forall x F x] \; = \; min\{[Fx] : x \text{ is assigned an object in the domain}\}$$
$$[\exists x F x] \; = \; max\{[Fx] : x \text{ is assigned an object in the domain}\}$$

Square brackets are used here to denote the value of a sentence: $[p]$ is the value of p. The functions min and max select the least $(\le)$ and greatest $(\ge)$ number, respectively, from a set of numbers.[8] Thus the disjunction of two sentences will have the value of its truest disjunct, while the conjunction of two sentences will have the value of its least true conjunct. Similarly, an existentially quantified sentence will have the value of its truest instance,[9] while a universally quantified sentence will have the value of its least true instance. A conditional will be completely true if its consequent is at least as true as its antecedent; otherwise the value of the conditional will be 1 less the difference between the values of the consequent and antecedent. Thus if $[p] = .6$ and $[q] = .7$, then $[p \to q] = 1$ and $[q \to p] = .9$.[10] Some classical equalities are retained, for example De Morgan's laws, the distributive laws for conjunction and disjunction, double negation elimination, contraposition, and quantifier conversion. Other classical equalities no longer hold, for example:

---

[7]Williamson (1994, §4.8) presents plausible constraints which do uniquely determine what functions conjunction and disjunction may be.

[8]In case there is no least or greatest number, min and max select the greatest lower bound and least upper bound, respectively, from a set of numbers. Here the choice of reals as values rather than rationals becomes important, since min and max are totally defined on sets of the former but not of the latter.

[9]Or the least upper bound of the values of its instances.

[10]If the biconditional is defined as the conjunction of a conditional and its converse, then the value of a biconditional $A \leftrightarrow B$ equals 1 if $[A] = [B]$; otherwise 1 minus the absolute value of $[A] - [B]$.

$$[p \to q] \;=\; [\neg(p \;\&\; \neg q)]$$
$$[p \to q] \;=\; [\neg p \lor q]$$

Let's now see how the degrees-of-truth account is able to handle the Sorites paradox:

> Heap (10,000)
> $\forall n$ (Heap $(n+1) \to$ Heap (n))
> _____
> Therefore, Heap (1)

The first premise is deemed to be completely true, that is, its value is 1. The conclusion is deemed to be completely false, its value 0. Meanwhile, for some n, the value of 'Heap (n)' will be neither 0 nor 1, but somewhere in between. Nevertheless, for each n, the value of 'Heap (n)' will be at most (say) .001 less than the value of 'Heap $(n+1)$'. Hence the value of each instance of the second premise, by our clause for '$\to$', is at least .999; so the value of the second premise (the tolerance principle) is itself at least .999, by our clause for '$\forall x$'.

Is the argument valid on this account? The question remains open. When values other than truth and falsity are introduced, we must ask which are the values that a rule of inference must preserve in order to be deemed valid. We call such values *designated values*. Assuming that a value greater than a designated value is itself a designated value, Universal Instantiation will be valid on this account no matter which set of values are chosen as designated. The choice of designated values does, however, affect the validity of Modus Ponens. If 1 is the only designated value,[11] then Modus Ponens is valid: for if $[p \to q] = 1$, then $[q]$ is at least as great as $[p]$; so if $[p] = 1$ then $[q] = 1$ too. If values less than 1 are also designated, however, Modus Ponens is not valid. For suppose the set of designated values is $\{x : x \geq .9\}$. Then if $[p] = .9$ and $[q] = .8$, the values of p and $p \to q$ are both designated—each is equal to .9—but the value of q is not designated.[12] Either way, the

---

[11] The is the approach adopted by Lukasiewicz.

[12] This seems to be the approach favored by Goguen. Some alternative approaches require no choice of designated values. That favored by Lakoff is: a rule of inference $\frac{\Gamma}{A}$ is valid just in case the value of A is never lower than the value of any member of Γ, even when the values of the members of Γ are themselves very low. Modus Ponens is not validated on this account. The approach favored by Edgington (1992), who rejects degree-functionality, is: a rule of inference $\frac{\Gamma}{A}$ is valid just in case $1 - [A]$ is never greater than $\sum_{S \in \Gamma} 1 - [S]$. Modus Ponens is validated on this account.

verdict is (as it should be) that the Sorites argument is not a good argument: if its premises are acceptable then its reasoning is not; conversely, if its reasoning is acceptable, then its premises are not.

_____

What answers, if any, does the degrees of truth account provide to questions 1–4 above (p. 23)? The answer to question 2. is that we are attracted to tolerance principles because they enjoy a very high degree of truth. If the difference in value between 'Heap $(i + 1)$' and 'Heap (i)' never exceeds ε, then the value of the tolerance principle for 'heap' will be at least $1 - \varepsilon$. Since ε will be very small, $1 - \varepsilon$ will be correspondingly high.

The answer to 3. is: yes we can maintain the thought that 'heap' has borderline cases and, despite the paradox, we are not required to accept the sharp-boundaries claim— that for some n, Heap $(n + 1)$ and ¬Heap (n). There are borderline cases in the sense that for some n, 'Heap (n)' is neither completely true nor completely false, but has instead a value in the intermediary range. In order for the sharp boundaries claim to be completely true, and hence command our acceptance, there would have to be an n such that the value of 'Heap $(n + 1)$' equaled 1 while that of 'Heap (n)' equaled 0. There is no such n since the borderline cases are nestled *between* those n for which 'Heap (n)' is completely true and those for which it is completely false. Since the sharp boundaries claim is not completely true, we are not required to accept it.

The answer to 4. is yes, we have revealed a revised version of the untenable thought that adjacent stages always have the same status. The revision takes the form: every stage does bear a special relation R to adjacent stages, which it does not bear to more distant stages. That relation R is not *has the same status as* but rather, *has a status extremely similar to*.

But what of our first question: If the tolerance principle is not true, why are we unable to find an untrue instance of it? To this the degree-theorist replies, "We *are*, as it turns out, able to find an untrue instance of the tolerance principle. Pick some n for which 'Heap $(n+1)$' and 'Heap (n)' both *obviously* have middling values. 'Heap $(n+1)$' will be truer

than 'Heap (n)', so 'Heap $(n+1) \to$ Heap (n)' will not be (completely) true." The answer is unsatisfactory—it misses the point of our original question. The degree-theorist is able to answer our original question while begging it, since on his account, untrue conditionals need not have true antecedents. So let us rephrase the question. If the tolerance principle is not true, and has only finitely many untrue instances, why can we not identify *all* of its untrue instances? More to the point: if some instance of the tolerance principle has a completely true antecedent but not a completely true consequent, why can we not say which instance that is? As so far presented, the degree theory has no means of answering this question.

In effect what we are now asking the degree-theorist is, why can we not locate the point of transition from complete heaphood to borderline heaphood in our series of stages. Given all we've said so far, the degree theorist seems committed to the existence of such a point, and cannot explain why we cannot find it. The commitment stems from an assumption that the meta-language in which the degree-theoretic semantics for a vague language is formulated is not itself vague: that in particular, sentences of the form '$[A] = x$' are themselves either completely true or completely false. But perhaps the degree-theorist could give up the assumption, and thereby discharge the commitment to the precise point.

The story would run something like this: "In order for us to be able to locate a point of transition in our series from complete heaphood to borderline heaphood, there would have to be an n such that the value of '$[\text{Heap}(n+1)'] = 1$' equaled 1, while the value of '$[\text{Heap}(n)'] < 1$' also equaled 1. But there is no such n. The truth of statements in our meta-language is not an all or nothing matter, it also comes in degrees." Rather than gauge the merit of this reply, I want to step back and make some general comments about the move that has been made here—the introduction of higher-order vagueness as an explanation of our inability to find a sharp boundary between those things that are not borderline case and those things that are are.

What is higher-order vagueness? There has been in recent discussions some equivocation on this point. I have said that a predicate is commonly defined to be vague just in case

it has borderline cases. Let us now call this *first-order vagueness*. On one conception of higher-order vagueness, a predicate is higher-order vague just in case certain higher-order predicates are first-order vague.[13] These higher-order predicates may either be predicates in the meta-language such as 'has the value .5', or they may be complex predicates in the object language containing operators such as 'half', as in 'is half-short'. (Recall that Goguen regarded 'x is half-short' and ' "x is short" is half-true' as equivalent.) Thus 'short' is first-order vague if it has borderline cases; second-order vague if 'half-short' (or 'completely-short' or 'three-eighths-short' or the like) has borderline cases—or equivalently, if for some name α, $\ulcorner \alpha$ is short$\urcorner$ is a borderline case of 'has the value .5' (or 'has the value 1', etc.); 'short' is third-order vague if 'half-short' or the like is second-order vague; $(n+1)^{th}$-order vague if 'half-short' or the like is n^{th}-order vague.[14]

It should be mentioned that the definition of higher-order vagueness can be expressed in a way that is more neutral with respect to competing accounts of vagueness. Instead of employing the operators 'complete(ly)' and 'half' in our definition of higher-order vagueness, we may use the operators 'definitely' and 'borderline' with the specific intent that these words be neutral. Taking 'borderline' as primitive, $\ulcorner$is definitely $F\urcorner$ may be defined by $\ulcorner$is F and not borderline $F\urcorner$. Sense may be given to these operators as is appropriate for a particular view.

The second conception of higher-order vagueness is much more inchoate than the first. If when a predicate F is first-order vague, we are unable to locate the boundary between the F's and the non-F's, then if we are also unable to locate the boundary between the things that are definitely F and those that are borderline F, it must be because F is second-order vague. Thus second-order vagueness is conceived to be whatever it is that is the source of this inability. Third-order vagueness will be whatever it is that is the source of an inability to locate a second-order boundary, and so on.

---

[13]*Cf.* Williamson (1994, p. 157), Fine (1975, p. 287)

[14]To give a rigorous account of the necessary and sufficient conditions for $(n+1)^{th}$-order vagueness, we would have to specify just what is meant by " 'half-short' *or the like*." This is not a trivial problem, and I have not yet reached a conclusion about what exactly the specification should be.

The two conceptions are tacitly assumed to amount to the same thing. It is thought that we are unable to locate the boundary between (for example) the men who are short and the men who are not short *because* there are borderline cases between the men who fall definitely into one or the other category—it is the existence of borderline cases that *explains* our inability. The subsequent assumption is that there must be a parallel explanation of our inability to locate the boundary between the men who are definitely short and those who are borderline cases. The parallel explanation is provided by acceptance of a class of men who do not fall definitely into one of *these* categories—a second-level class of borderline cases.

In my view, the picture is off the mark. What is mistaken about it is its initial suggestion that the existence of borderline cases is what *explains* our inability to locate the first-order boundary. In my view, the existence of borderline cases is not really what explains our inability to locate boundaries at all. We must postpone elaboration of the view, however. We will also leave off for the moment our discussion of higher-order vagueness, with the conclusion that it is probably best understood in accordance with the first conception above, namely, as the vagueness of certain higher-order predicates. It then remains an open question whether there is higher-order vagueness, and whether something other than higher-order vagueness might explain our inability to locate "higher-order" boundaries.

We now resume discussion of the degree-theoretic resolution of the Sorites paradox. We had wondered whether the degree-theorist could provide an answer to our first question: Why can we not discover an instance of the tolerance principle that has a completely true antecedent but not a completely true consequent? The degree-theorist may have the resources to provide an answer to this question by extending his semantics to the language in which that semantics was formulated. But short of doing that, it seems he must remain silent. I will not explore the prospects for a degree-theoretic account of higher-order vagueness,[15]

---

[15]Though for a dim view of the prospects, see Fine (1975, pp. 292f.), Rolf (1984, pp. 222f.), Burns (1991, pp. 72–74), and Williamson (1994, §4.12).

for I think we must first assess whether the answers already provided by the account are adequate.

What has made the degree-theoretic account of vagueness seem so plausible to many is that it provides answers to questions 2. and 4. when other accounts do not. The degree-theorist is able to help us rest more easily with the fact that we must give up the tolerance principle since on the one hand he provides an explanation of why we were so drawn to it in the first place (question 2.), and on the other hand he provides a revision of it that is acceptable (question 4.). The account is only as good as its underpinnings, however. We still do not know what degrees of truth *are*, and so we have no idea what the connection might be between the degree of truth of a sentence and its attractiveness to us. Thus we still have no reason to accept that high-degrees of truth do explain our attraction to certain sentences.

The most common criticism of the approach is that it assigns values to some sentences that are intuitively too high. Although instances of excluded middle are not always assigned the value 1,[16] (and this has in many quarters been deemed to be a good thing), the flip side is that instances of contradictions will not always be assigned the value 0. For example, $[p \& \neg p] = \frac{1}{2}$ when $[p] = \frac{1}{2}$. Attempts have been made to lower the value of contradictions by revising the clause for ' & '.[17] But no degree-functional account can reduce the values of contradictions to 0 without losing all credibility. For if the value of a conjunction is just a function of the *value* of its conjuncts, and in particular is blind to any logical relations these conjuncts might bear to each other, then were $[\neg p]$ to equal $[\neg q]$, $[p \& \neg p]$ would equal $[p \& \neg q]$. This can be seen to be implausible when we let $[p] = \frac{1}{2}$, for then $[\neg p] = [\neg\neg p]$, and so $[p \& \neg\neg p]$ would be 0.

How good is the objection? I say that at this point, we have no more reason to think the objection good than we have reason to accept that possession of a high degree of truth explains our attraction to the tolerance principle. Until we have an explanation of what of

---

[16]For example, when $[p] = \frac{1}{2}$, $[p \lor \neg p] = max\{\frac{1}{2}, (1 - \frac{1}{2})\} = \frac{1}{2}$.
[17]Goguen (1967) for example at one point suggests that we might let $[p \& q] = [p] \times [q]$.

it is to have the value $\frac{1}{2}$, we have no grounds for regarding that value as too high for a contradiction to have. After all, when faced with borderline cases we are sometimes attracted to contradictions, at least in the sense that it may seem appropriate to utter them. Were I to be asked whether *LBM* is red, I might well reply (though in a hesitant, not assertive tone of voice), "Well, ... it's red and it isn't." If degrees of truth are supposed to explain our attraction to certain sentences, and we are sometimes half-attracted to contradictions, then $\frac{1}{2}$ might be just the right value for contradictions in some cases.

The proponent of the degree-theory may not, however, understand degrees of truth as *just being* measures of how attracted we are to a sentence, since that would deprive the account of any explanatory power. For then to say that we are attracted to tolerance principles because they enjoy a high degree of truth, would just be to say that we are attracted to them because we are attracted to them.

The problem is compounded by the fact that the degree of truth assigned to a sentence does not always coincide with the degree of its attractiveness. Moreover, the relative degrees of truth of two sentences may not always coincide with their relative attractiveness. Let's look at an example. At some point, the successive values of 'Heap (i)' will cross the half-way mark. By this I mean that ['Heap(i)'] will be greater than or equal to $\frac{1}{2}$, and ['Heap$(i-1)$'] will be less than $\frac{1}{2}$. At this point, the value of 'Heap(i) & ¬Heap$(i-1)$' will itself be at least $\frac{1}{2}$, and may even exceed $\frac{1}{2}$ by a small margin.[18] Consider a concrete example. Let ['Heap(i)'] $= .51$ and ['Heap$(i-1)$'] $= .49$; then ['Heap(i) & ¬Heap$(i-1)$'] $= min\{.51, (1-.49)\} = .51$.[19] Thus, at the point p at which the half-way mark is crossed, the claim that p is the boundary between the heaps and the non-heaps will have a value at least as high, possibly higher, than the claim that p is not the boundary. This is contrary to the relative attractiveness of the two claims.

It would be misguided, at this point, to pursue a strategy of trying to revise the

---

[18]NB: It follows that the sharp boundaries claim itself may have a value exceeding $\frac{1}{2}$. *Cf.* Wright (1987, p. 251) for a similar point.

[19]More generally, if ['Heap(i)'] $= \frac{1}{2} + \delta$ and ['Heap$(i-1)$'] $= (\frac{1}{2} + \delta) - \varepsilon$, where $0 \leq \delta < \varepsilon$, then ['Heap$(i)$ & ¬Heap$(i-1)$'] $= \frac{1}{2} + min\{\delta, \varepsilon - \delta\}$.

clauses for the various logical connectives—in this case conjunction—with the hope of better approximating degrees of attractiveness with degrees of truth. As long as degree-functionality is retained, the strategy will never be successful, since the degree of attractiveness of a complex sentence is not a function of just the degrees of attractiveness of its parts, but will be a function also of the logical, analytical and also evidentiary relations those parts bear to one another.[20]

The degree theorist is thus faced with the task of elaborating a conception of degrees of truth that meets the following criteria: (i) Degrees of truth do not always coincide with, but may nonetheless sometimes explain, our attraction to certain sentences; (ii) It is possible for a contradiction to have a value greater than 0; and (iii) Degrees of truth greater than 0 but less than 1 are properly understood as neither *true* nor *false*.

One way to understand the degree of truth of a sentence is as a measure of how similar the world actually is to the way that a sentence says the world is; or conversely, as a measure of how different things would have to be in order for the sentence to be true. To illustrate: If there are exactly 250,126,435 permanent residents of the United States, then 'There are 250 million permanent residents of the U.S.' would be very nearly true, since things would not have to be *that* different in order for it to be true; while 'There are three permanent residents of the U.S.' would be very far from true, since things would have to be very different from how they in fact are in order for the sentence to be true. A sentence that is very nearly, but not quite true, may be "true enough" for it to be appropriate to assert it.

It may be that such a conception of a degree of truth meets criterion (i) above, for it may be that sometimes we believe a sentence that is only very nearly true because we do not recognize that things would have to be just *slightly* different in order for it to be true. We could then also understand why high (or low) degrees of truth do not *always* coincide with our attraction (or aversion) to a sentence. Nevertheless, the account fails to

---

[20]An example of what I mean by an evidentiary relation is the relation that a sentence S bears to S' when one could not have evidence for S without also having some evidence for S'.

meet criteria (ii) and (iii). It fails to meet criterion (ii) since on the present conception of a degree of truth, the value 0 should be reserved for all and only contradictions, generously understood. For it makes no sense to say that the world would have to be *more* different in order for 'One grain of sand can form a heap' to be true, than it would have to be in order for a sentence of the form 'p & $\neg p$' to be true. *No* amount of change in the way things are would be enough to verify a sentence of the form 'p & $\neg p$'.[21] The conception fails to meet criterion (iii) since although it does give us reason for thinking of *being nearly, but not quite true* as a way of being not true, the conception gives us no reason for thinking of it as a way of being not false. 'There are 250 million residents of the U.S.' may be very nearly true, but it is not thereby far from being false; it is not *completely* false, in the sense of being widely off the mark, but it is still false.[22]

Mark Sainsbury (1988, §2.3) advocates a different way of understanding degrees of truth—in terms of degrees of confidence. I am more confident that it is raining now, than I am confident that it was raining yesterday morning; but still more confident that it was raining yesterday morning than I am confident that it was raining a week ago. It thus makes sense to think of our confidence in the truth of a statement as coming in degrees. Sainsbury believes that our confidence in the truth of statements containing vague predicates may also come in degrees, but that what is distinctive about the degree of confidence we have in the truth of these statements is that it may not change as we become better informed. He believes that I could know all there is to know, and still have only partial confidence in the truth of '*LBM* is red'; "an omniscient being could do no better. ... Where we have vagueness, there may be no chance of improvement" (pp. 44-5). Thus a degree of truth of a sentence S is conceived to be the degree of confidence in S a fully informed person would have (or perhaps ought to have) on reflection.

---

[21] A qualification: a change in the meaning of ' & ' might be sufficient to verify a sentence of the form 'p & $\neg p$'; what is meant here is that no amount of change in the way things are would be enough to verify a sentence of the form 'p & $\neg p$', given its *actual* truth-conditions. *Cf.* Kripke (1972).

[22] It may be that on some occasions, 'there are' may be used with the same sense as 'there are roughly', and that on such occasions-'There are 250 million residents of the U.S.' will be not merely nearly true, but completely true.

I find it difficult to accept the compatibility of omniscience and partial degrees of confidence. On the one hand it seems that partial *confidence* in the truth of some statement is an attitude someone has only when he believes there to be a *correct* answer to the question 'is this true?', but is not entirely convinced that *he* has gotten the answer right. If there is no correct answer, then a fully informed person will know as much, and hence will himself feel *no* confidence, not even partial confidence, in any answer. In fact, he will feel completely confident that his attitude is the right one to take.

But, even if we could make out a conception of partial confidence which sensibly cohered with omniscience, it still remains problematic that partial confidence must furthermore be construed as coming in varying degrees. To justify the claim that it does, Sainsbury (1988, p. 44) invites us to imagine the following scenario: you wish to poison Jones by mushroom, and you know that all and only red mushrooms are poisonous. It is then claimed that even if the only mushrooms you can get your hands on are neither definitely red nor definitely not red, you may still regard it as more reasonable to use this mushroom rather than that one for the job at hand. "The more confident you are that this mushroom is really red, the more reasonable it is to use it; the less confident, the less reasonable." This is extremely puzzling. If it is true that all and only red mushrooms are poisonous, then presumably the vagueness of 'poisonous' must match that of 'red'. If you are merely partially confident in each of two mushrooms that they will do the job, (and we are permitted to assume that you are fully informed, even omniscient), what room is there for being *more* confident in the one mushroom than in the other. It cannot be that you think it more *likely* that the one will do the job than that the other will. By hypothesis, you are correct to lack full confidence in the truth of either 'this one will do the job' or 'that one will', and you *know* this. You entertain no hope for the truth of either.

Even if sense can be made of the compatibility of omniscience with partial degrees of truth, the account faces the further problem that even when we are fully informed (in whatever the appropriate sense turns out to be), our degree of confidence in the truth of some statement just will not be a function of the degree of our confidence in the truth of its

parts. I may be partially confident that a mushroom is red, and still slightly more confident that it is brown, but I had better be fully confident that it is not *both* red and brown. It is surprising in fact that given Sainsbury's interpretation of degrees of truth as degrees of confidence, he should affirm a degree-functional approach. He writes:

> ... the naturalness of 'It is and it isn't', as a response to the question whether a borderline case mushroom is red, gives at least a preliminary indication that the degree theorist is right to recognize that not all instances of '*p* and not-*p*' are completely false. (45)

But it seems to me that 'It is and it isn't' (said hesitantly) as a response is not really evidence of partial confidence in the truth of 'It is and it isn't'. Rather 'It is and it isn't' seems to be a way of expressing that 'red' has different uses, and that while it could be appropriate or permissible to call the mushroom 'red' it could also be appropriate or permissible to call it 'not red'. The response is uttered hesitantly not to indicate partial confidence, but rather to indicate that it is not to be construed literally as an assertion.

The final conception of a degree of truth that we will consider is the one most commonly affirmed, but which in my mind fares even worse than the preceding two. The idea is to understand the truth of certain comparative claims as providing the basis for varying degrees of truth. If this mushroom is redder, or more poisonous, than that one, then it seems to make sense to say that this mushroom is red, or poisonous, to a greater degree than that one is, and hence that the claim that this mushroom is red, or poisonous, is *true* to a greater degree than the claim that that one is.

One feels that something like this idea must have been underlying Sainsbury's claim that it could be more reasonable to use one mushroom than another to poison Jones, even when neither mushroom is clearly red or clearly poisonous. The idea must have been that if you want to poison Jones, then you want him as poisoned as possible; and that if this mushroom is redder than that one, even if both are borderline cases, then this mushroom would poison Jones more than that one would. But why should we accept this? On the one hand, even if you want Jones done in, you may still prefer a healthy, unsuspecting Jones, to a somewhat sick and suspicious one. On the other hand, it is not at all clear why the

truth of certain comparatives should lead to differing degrees of confidence. If you have two clearly brown—hence clearly not red—mushrooms, it would be unreasonable to use either to poison Jones, even if one is redder than the other. Even if one is redder than the other, you would not be more confident that one is red than that the other one is. Assuming you are fully informed, you will be completely confident that neither is red. Similarly, given two clearly red mushrooms, you will not be more confident that one is red than that the other one is, even if one is redder than the other; you will be completely confident that both are red. Why should the situation be different with borderline cases? It is of course true that if you are *unsure* whether it would be correct to say, about each of two things, that it is red because you are unsure about the proper use of 'red', then you may just be *less* unsure about the redder of the two because you think that the redder of the two has a better *chance* of being in the extension of 'red' than the other. But we have stipulated that in the imagined circumstance, your partial confidence does not stem from lack of information—about the proper use of 'red' or anything else.

Using the truth of certain comparative claims to justify the possibility of varying degrees of partial confidence (even for a fully informed subject), and these in turn to justify varying degrees of truth, seems to raise more questions than it answers. And at any rate, it remains implausible to retain degree-functionality for degrees of confidence. The final conception of degrees of truth we are considering bypasses degrees of confidence, and aims to provide a foundation for degrees of truth by direct appeal to the truth of certain comparatives. Commitment to degree-functionality remains as problematic as before, however. It may be that things can be tall to differing degrees, but how could anything be tall-and-not-tall to any degree? But it is not only the commitment to degree-functionality that creates problems for the view. If we accept degrees of truth in order to account for the truth of certain comparatives, then it becomes difficult to see how degrees of truth could be adapted to account for the semantics of vague predicates. If, for example, we say that 'Michael is taller than Dennis' is true if and only if 'Michael is tall' is truer than 'Dennis is tall', even when Michael and Dennis are *both* tall, then we are precluded from identifying

"complete truth" in the intended sense with the value 1. If degrees of truth are required
to account for the semantics of 'taller than', then for no α will $\ulcorner \alpha$ is tall$\urcorner$ be completely
true, since there could always be a β for which $\ulcorner \beta$ is taller than $\alpha \urcorner$ is true. Similarly, we
are precluded from saying that a is a borderline case of F if and only if (where α names
a) $\ulcorner \alpha$ is $F \urcorner$ has a value between 0 and 1. The predicate 'acute' (as applied to angles) has
no borderline cases, but still it must be the case that for some α, $\ulcorner \alpha$ is acute$\urcorner$ has a value
between 0 and 1, since there could be three angles, the first of which is more acute than the
second, and the second of which is more acute than the third.[23] Thus the account does not
meet my third criterion—namely, that degrees of truth greater than 0 but less than 1 are
properly understood as neither *true* nor *false*. Given a continuum of degrees of truth, as
these must be employed in a semantic theory of comparatives, it remains an open question
how complete truth and complete falsity map onto this continuum. Degrees of truth, on
this conception, do not do the work they were invoked to do.

2.2 Supervaluations

"Either Jim is not taller than Eric, or Jim is tall if Eric is." This disjunction seems to have
the status of an analytic truth. How can we provide a semantic theory that will account
for the truth of the disjunction, even when Jim and Eric are both borderline cases of tall?
In order to illustrate the problem, we will compare the following two sentences:

(1) Either Jim is not taller than Eric, or Jim is tall if Eric is.

(2) Either Jim is not taller than Eric, or Jim is bald if Eric is.

Now suppose that the first disjunct in these sentences is false—that is, Jim *is* taller
than Eric; but that Jim and Eric are each borderline cases of both 'bald' and 'tall'. Any
truth-functional semantic theory—one according to which the truth value of a complex
sentence is a function just of the truth values of its parts—will verify (2) if it verifies (1).
But when Jim and Eric are each borderline cases of 'bald', it does not seem that (2) should

---

[23]This example is borrowed from Williamson (1994, p. 127).

be verified. Due to what Kit Fine calls a "penumbral connection" between certain predicates in the language—in this case between 'is taller than' and 'is tall'—if Jim is taller than Eric, then 'Jim is tall if Eric is' should be *true*, even though 'Jim is *bald* if Eric is' might not be. The commitment to degree-functionality emerged as a problem for the degree-theorist for just this reason. Accounting for "penumbral connections" between vague predicates, while accommodating their indeterminacy, is the motivation for a semantic theory of vagueness which has come to be known as supervaluational semantics.[24]

The defining feature of vagueness, according to Fine (1975), is what he calls "deficiency of meaning." What is meant is illustrated by an artificial predicate of natural numbers, 'nice$_1$', the meaning of which is given by just the following clauses:

- n is nice$_1$ if $n > 15$
- n is not nice$_1$ if $n < 13$

The predicate 'nice$_1$' is vague, claims Fine, in part because it will be the source of truth-value gaps in sentences containing it, for example, '14 is nice$_1$'. But not only because of this. What distinguishes truth-value gaps arising from vagueness from those that arise from other sources is "that they can be closed by an appropriate linguistic decision, viz. an extension, not change, in the meaning of the relevant expression" (p. 267). Thus we could decide to count 14 as nice$_1$, or we could decide to count it as not nice$_1$. Neither decision would constitute a change in, but merely an extension of, the meaning of 'nice$_1$'. In making such a decision, we will be "precisifying" the meaning of 'nice$_1$'. The idea of supervaluational semantics is to employ such precisifications in the truth-definition for the language. A sentence will be true (simpliciter) just in case it would be true on every admissible such precisification.

A supervaluational model will contain a set of "points"—the admissible precisifications of the expressions in the language—each of which may be taken to be a classical

---

[24]The *locus classicus* for the theory I will be presenting, as applied to the problem of vagueness, is Fine (1975). See also, Lewis (1970), Dummett (1975), Dummett (1991, ch. 3). For discussion and criticism of the view, see Williamson (1994, ch. 5) and Sorensen (1988, pp. 236–239). It should be mentioned that Fine presents formal considerations which he takes to provide independent motivation for supervaluation semantics.

model. A sentence will be true in a model just in case it is classically true in every point in the model. Taking the falsity of a sentence to be the truth of its negation, a sentence will be false in a model just in case its negation is true in the model, hence just in case its negation is true in every point in the model, hence just in case it is false in every point in the model. Let's look at an example. Suppose, for simplicity, that 'tall' and 'short' are the only vague predicates in the language, and that 'tall' is true of everyone 6' or taller, false of everyone 5'8 or shorter; and that 'short' is true of everyone 5'4 or shorter, false of everyone 5'9 or taller. A *precisification* of these predicates will close the gap by expanding the extensions and anti-extensions of these predicates to cover all undecided cases, leaving all previously decided cases as before. But there are constraints. In order for a precisification to be *admissible*, it must adhere to what we might call certain *meaning postulates*. For example, the gaps must be closed in an orderly way: Anyone taller (shorter) than someone in the extension (anti-extension) of 'tall' in a precisification must also be in the extension (anti-extension) of 'tall' in that precisification; anyone shorter (taller) than someone in the extension (anti-extension) of 'short' in a precisification must also be in the extension (anti-extension) of 'short' in that precisification. Also, "penumbral connections" between 'tall' and 'short' must be respected: although each of 'tall' and 'short' may be precisified in such a way that a $5'8\frac{1}{2}$ inch person is placed in its extension, the extensions of 'tall' and 'short' in any *single* precisification must be disjoint—'tall' and 'short' are incompatible predicates.

It is easy to see that every theorem of classical logic is validated on this account: a theorem of classical logic will be true in every classical model, hence true in every "point" in a supervaluational model, hence true in every supervaluational model. The account also manages to verify penumbral truths like (1). If Jim is taller than Eric, then 'Jim is tall if Eric is' will be true in every admissible precisification of 'tall'.

What will be the assessment of the Sorites argument?

Heap (10,000)
$\forall n$ (Heap $(n+1) \rightarrow$ Heap (n))

—————————————————

Therefore, Heap (1)

Both Universal Instantiation and Modus Ponens are validated by supervaluational seman-tics. If a universal generalization is true in a model, then it is classically true in each point in the model; hence each of its instances is classically true in each point in the model; hence each of its instances is true in the model. If a conditional and its antecedent are true in a model, hence classically true in each point in the model, then the consequent of the conditional will also be classically true in each point, hence true in the model.[25]

Assuming that 10,000 is in the extension of 'heap' prior to any precisification, that is, assuming that 10,000 grains of sand *definitely* can form a heap, 10,000 will be in the extension of 'heap' in every precisification of the predicate, hence the first premise of the argument is deemed true. Similarly, the conclusion of the argument is false: 1 will not be in the extension of 'heap' in *any* precisification.

What about the second premise—the tolerance principle for 'heap'? It is not deemed as being neither true nor false, as it is by the degree theorist, but rather just plain false. In any admissible precisification of 'heap', there will be some n in the extension of 'heap', while $(n-1)$ is in its anti-extension. It will not be the *same* n in every precisification, but there will always be one. Thus the second premise will have some false instance in every precisification—different instances in different precisifications, but always at least one (in fact at most one as well, if the precisification is admissible). Hence the universal will be false in every precisification, hence false absolutely. The Sorites argument is deemed valid but unsound.

Although supervaluation semantics, if correct, provides a way of understanding why the tolerance principle is not true, it cannot really be counted among the proposed *solutions* to the Sorites, since it provides an answer to only the third of my four questions. That third question was: Does reflection on the Sorites paradox require us to accept sharp boundaries claims for vague predicates—does it require us to accept, for example, that for some n, n

---

[25]We are taking an argument $\frac{\Gamma}{A}$ to be valid just in case A is true in every model in which every member of Γ is true. Williamson (1994) refers to this as "global validity." It may be contrasted with a stricter notion, which he refers to as "local validity," according to which an argument $\frac{\Gamma}{A}$ is valid just in case in every model: A is (classically) true *in every point* in which every member of Γ is (classically) true. Universal Instantiation and Modus Ponens are also both valid on the stricter conception.

grains can form a heap while $(n-1)$ grains cannot? And if so, how can we reconcile this claim with the thought that 'heap' has borderline cases? The supervaluationist does require us to accept the sharp boundaries claim—it is, by his lights, true—but he argues that the truth of this claim is compatible with the existence of borderline cases, things of which we cannot truly say either that they are heaps or that they are not. The argument rests on a distinction between two claims:

- It is true that: $\exists n(\text{Heap}(n+1)\ \&\ \neg\text{Heap}(n))$

- For some n, it is true that: $(\text{Heap}(n+1)\ \&\ \neg\text{Heap}(n))$

The supervaluationist, though he affirms the first claim, rejects the second. Although what we have been calling the sharp boundaries claim for 'heap' is deemed true, none of its instances are. None of its instances are true since between those i for which 'Heap (i)' is true, and those for which '$\neg$ Heap (i)' is true will be some i—the borderline cases—of which neither is true. Similarly, supervaluational semantics declares some disjunctions true that have no true disjuncts. For any n, the claim that n is a heap or it isn't will be true, even though it will not be the case that for any n either the claim that n is a heap will be true or the claim that it isn't will be true. These results are the cost of providing a semantic theory which permits truth-value gaps but also verifies those sentences that express penumbral connections.

Is the cost too high? We cannot help but feel that corresponding to every true disjunction and existential is a question 'Which one?' that has a right answer. We may say to the supervaluationist, "you have defined a perfectly good predicate of sentences, but just because you have spelled it t-r-u-e does not show that it is a truth predicate. That this predicate may apply to an existentially quantified sentence without applying to any of its instances is, moreover, evidence to the contrary." The supervaluationist will respond to this charge, however, by telling us that the phenomenon of vagueness may make it impossible to accommodate *all* of our gut feelings. He has at least accommodated many of them, and thus the burden of proof is on those who would deny the equation of truth with "super-truth".

Further problems for the account surface, however, once an operator 'Definitely' (which will abbreviate with a 'D'), used in the object language to express the supervaluationist's conception of truth, is introduced. A provisionally proposed semantics for this operator is the following: a sentence DA is true at a point in a model just in case A is true at *every* point in the model. The value of DA will remain constant from point to point in a model, and hence will always be either true or false. Thus, given this provisional semantics, (which will eventually be modified in order to account for higher-order vagueness), a sentence of the form $A \to DA$ may not be true. If A, for example, is 'Herbert is bald', where Herbert is a borderline case, A will be true at some points, false at others. But since the value of DA does not vary from point to point, the conditional 'If John is bald then it is definitely the case that John is bald' will be neither true nor false. It is however the case that a sentence of the form DA will always be a *consequence* of A—true whenever A is. But still the gappiness of 'If John is bald then it is definitely the case that John is bald' casts some doubt on the claim that what 'definitely' expresses is truth.[26]

But let us grant, for the sake of argument, that the definiteness operator does indeed have the sense it is intended to have. Still there are problems. With 'definitely' in the language, we will not only have true disjunctions with no disjuncts that are *in fact* true, we will also have true disjunctions with no disjuncts that *could be* true. In giving an example, I will use 'BA' as an abbreviation of '$\neg DA$ & $\neg D \neg A$'. If A is 'Herbert is bald', then BA may be understood as expressing that Herbert is a borderline case.[27] Now consider the following:

(3) $(A \ \& \ BA) \lor (\neg A \ \& \ BA)$

Take A again to be 'Herbert is bald', and suppose that Herbert is a borderline case. Then the sentence is true, since equivalent to $(A \lor \neg A) \ \& \ BA$, the first conjunct of which is

---

[26]McGee & McLaughlin (1995) claim that vagueness forces us to recognize that we have two distinct conceptions of truth, which they call 'truth' and 'definite truth', only the first of which satisfies the disquotation schema.

[27]On the assumption that every instance of $DA \to A$ is valid, this definition of the borderline case operator in terms of the definiteness operator is equivalent to my earlier definition (p. 33) which took the borderline case operator as primitive: $DA \equiv_{def} A \ \& \ \neg BA$.

true, since it is a classical theorem, and the second conjunct of which is true by stipulation. But not only is neither disjunct of (3) true, neither disjunct could be true. Take the first disjunct. It is a conjunction, and so in order for it to be true, each of its conjuncts must be true.[28] But if A is *true*, then DA is true, since a consequence of A. But if DA is true, then BA is false. Similar reasoning applies to the second disjunct of (3).

It is difficult to say how much of a problem this is for the supervaluationist. Offhand it may seem that if he is prepared to countenance the truth of a disjunction when neither of its disjuncts is in fact true, it is not *much* more of a step to countenance the truth of a disjunction when neither of its disjuncts *could* be true. I would hazard a guess, though, that if this does not seem much more of a step, it is because the first step already seems too steep.

———————

Of course, nothing in what I have said so far *refutes* supervaluational semantics. What is of primary concern here is that it does not provide the means for answering those questions that really get to the heart of the Sorites paradox. I offered two reasons for our difficulty in accepting sharp boundaries claims for vague predicates—one is the thought that, given an appropriate series, there will be *between* the positive cases and the negative cases of a vague predicate some things of which one cannot correctly say either that the predicate applies or that it does not apply; the other is the thought that given an appropriate series, any adjacent members of it will have the same *status*, whether it be positive, negative, or indeterminate. Supervaluational semantics helps us deal only with one half of the difficulty, by rendering the first thought compatible with sharp boundaries claims. But it is the second thought which really underlied our acceptance of tolerance principles in the first place. Supervaluation semantics, unlike degree-theoretic semantics, gives no indication of any acceptable revision of the second thought (question 4.); unlike the degree-theory, it gives no explanation of why we were so attracted to tolerance principles in the first place

—————————————

[28] If a conjunction is true in a model, then it is true in every point in the model, hence each of its conjuncts is true in every point in the model, hence each of its conjuncts is true in the model.

(question 2.); and, like the degree-theory, it so far provides no explanation of why we cannot say which instances of a tolerance principle are not true (question 1.).

Fine (1975) recognizes that there is some pressure to explain at least why, if tolerance principles for vague predicates are false, we are nonetheless so attracted to them. His remarks in answer to the question are only very brief, however. I quote those remarks in full:

> I suspect that the temptation to say that the second premiss is true may have two causes. The first is that the value of a falsifying n appears to be arbitrary. This arbitrariness has nothing to do with vagueness as such. A similar case, but not involving vagueness, is: if n straws do not break a camel's back, then nor do $(n + 1)$ straws. The second cause is what one might call truth-value shift. This also lies behind LEM [The Law of Excluded Middle]. Thus $A \vee \neg A$ holds in virtue of a truth that shifts from disjunct to disjunct for different complete specifications, just as the sentence 'for some n a man with n hairs is bald but a man with $(n + 1)$ hairs is not' is true for an n that shifts for different complete specifications. (p. 286)

It is odd that Fine should cite the *arbitrariness* of the falsifying n as a cause of our temptation to say that a tolerance principle is true. In many cases, we readily acknowledge that a tolerance principle is not true, or that a "slippery slope" argument is not a good one, despite the arbitrariness of the bad step. That is why the claim, "That's the straw that breaks the camel's back" has such metaphorical weight. If I tug your ear once, you may not get angry. If I keep tugging your ear, when it is has long ceased to be funny, you will eventually get angry. There will be some arbitrariness in *which* of the tugs pushes you over the edge, but that in no way leads me to believe that if n ear tugs do not provoke your wrath, then neither will $(n + 1)$.

The second cited cause of our temptation to accept tolerance principles seems to be tantamount to the claim we are tempted to accept tolerance principles because we do not recognize supervaluational semantics to be correct. The claim raises a serious methodological question. The project of providing an adequate semantic theory of natural languages is an empirical enterprise. What are we to take as the data of our enterprise? Which are the phenomena against which we measure our theory for adequacy or correctness? What

observed phenomena could prove a theory wrong? One of the main motivations for and attraction of the supervaluationist's theory is that it manages to verify sentences—those which express penumbral connections—that we are inclined to think true, where truth-functional theories that admit truth-value gaps cannot. Why is our inclination to regard true existential and disjunctive claims as having a corresponding 'which one?' question that has a correct answer given any less weight. Fine seems to think that overall best fit with our intuitions is the best we can hope for in a semantic theory, even when sometimes the fit is not at all good at certain places.[29] Oddly, he does not consider the question whether acceptance of bivalence might yield an overall best fit. The methodology, and its application, are unsatisfying.

Nevertheless, we may still wonder whether supervaluational semantics could in some way be supplemented. One prospect would be to supplement supervaluational semantics with degrees of truth, with the aim of reclaiming a more complete solution to the paradox. A sentence would still be deemed true (false) just in case true (false) in all admissible precisifications of the vague expressions in the language. But in addition to the values *true* and *false*, a continuum of intermediary values would be admitted as well. Very roughly, the idea would be to take the value of a sentence to be the proportion of admissible precisifications in which it is true.[30] Suppose that 1,500 is the least number for which '*n* grains of sand can form a heap' is true, and that 500 is the greatest number for which it is false, and that consequently there are 1,000 admissible precisifications of '*n* grains of sand can form a heap'. The predicate will be true of 1,499 in 99.9% of them. The drop in truth value from any number to its predecessor will be at most $\frac{1}{100}$. Thus on the combined account, we could again say, in answer to my fourth question, that although adjacent members in a sorites series do not always have the *same* status, they do always have a very similar status.

Would the combined account also provide an explanation of why we are so attracted to the tolerance principle in the first place? It may not be said, as it could be said on the

---

[29] *Cf.* Fine (1975, p. 286).
[30] *Cf.* Lewis (1970, 1979). See Kamp (1975) and McGee & McLaughlin (1995) for developments.

degree-functional theory presented in section 2.1, that we are attracted to the tolerance principle because it enjoys a very high degree of truth. On the combined account, the tolerance principle will still be perfectly false—true in no admissible precisification. It will be the case, however, that every *instance* of a tolerance principle will be very nearly true. In the case of the tolerance principle for 'heap', every instance will be true in all but one admissible precisification.[31] Does this fact explain why we are inclined to believe the tolerance principle itself? McGee & McLaughlin (1995, 236ff.) claim that it does. But there is a worry. Combining degrees of truth with supervaluations in the way sketched, yields a conception of a degree of truth as a measure of a kind of probability—the probability that a sentence would be true if the language were made precise. Now it is not generally true that whenever we believe each instance of a universal to be highly probable we also tend to believe the universal itself to be highly probable. Suppose you are being dealt a single playing card from a normal deck, and consider the claim that for any x, you will not be dealt an x. (Take the variable to range over the cards in the deck, e.g. the two of clubs, the queen of spades, etc.) We regard each instance as highly probable, but we do not thereby regard the universal itself as highly probable. We know we will be dealt some card or other. The explanation must, then, run instead as follows. Because each instance of the tolerance principle is very nearly true, we *mistake* each instance as being *completely* true. Because we mistake each instance as being completely true, we mistakenly regard the tolerance principle itself as completely true. What is missing, however, is an explanation of why we would mistake near truth for complete truth.

At any rate, we still lack an answer to the first on my list of questions: Why are we unable to say which instances of the tolerance principle are not true? In particular, why can we not say which of its instances has a true antecedent, but an untrue consequent? Put another way, why can't we find a boundary in our sorites series for 'heap' between the stages that are definitely heaps and those that are not definitely heaps? The supervaluationist claims that although for any stage in our heap series, that stage is either a heap or it isn't,

---

[31]We are still taking 'Heap (x)' to abbreviate 'n grains of sand can form a heap'.

we cannot find a boundary between the heap stages and the non-heap stages, since it is not the case that every stage is either *definitely* a heap or *definitely not* one. But given the semantics so far provided for the definiteness operator, it *will* be the case that every stage is either definitely definitely a heap or definitely not definitely a heap—a sentence of the form 'Definitely: n grains of sand can form a heap' will always be either true or false.

In order to account for the possibility that not only 'Heap (n)', but also 'Definitely: Heap (n)', may have borderline cases—in order, that is, to account for the possibility of higher-order vagueness—Fine provides a revised semantics for the definiteness operator.[32] On the new account, the structure of a supervaluational model is enriched by introducing an accessibility relation on points in a model. Instead of saying that DA is true at a point just in case A is true at every point, we now say that DA is true at a point w just in case A is true at every point *accessible* to w. The accessibility relation is not taken as primitive, however. The structure of each point in a model is also enriched, and the accessibility relation on those points is determined by that structure.

Instead of taking each point to be an admissible precisification of the expressions in the language, each point w is, on the new picture, an infinite sequence $w_0, w_1, w_2, \ldots$, the first term of which is a precisification of the expressions in the language, the second term of which is a set of precisifications, the third term of which is a set of sets of precisifications, and so on. A constraint placed on these sequences is that each w_i must be a member of w_{i+1}. The intuitive idea is that if a predicate F is second-order vague, then not only is there more than one admissible way of drawing the "0th-order boundary" between the Fs and the not-Fs, there is also more than one admissible way of drawing the two first-order boundaries—the boundary between the definitely Fs and the borderline Fs and between the borderline Fs and the definitely not-Fs. If F is third-order vague, then there will also be more than one admissible way of drawing the five second-order boundaries; ... if F is $(n+1)^{th}$-order vague, there will be more than one admissible way of drawing the (2^n+1) n^{th}-order boundaries. Each point $w_0, w_1, w_2, \ldots$ in a model can be understood as representing

---

[32]Fine (1975, §5, pp. 293ff.).

a choice of all such boundaries, each w_i representing a choice of boundaries at the ith order.

The accessibility relation is then defined as follows: a point w' is accessible to a point w just in case, for each i, w'_i is a member of w_{i+1}. Given this definition and the constraint on sequences that each term of a sequence be a member of its successor in the sequence, the accessibility relation will always be reflexive. But it need not be either symmetric or transitive. Let's look at an example. For simplification, we will assume that 'n grains of sand can form a heap' is the only vague predicate in the language. I will use boldface numerals to denote precisifications—for example, **1,000** is the precisification in which 1,000 is the *least* number of grains that can form a heap. The model under consideration contains just three points:

$$
\begin{array}{llllll}
w & = & 7, & \{6,7,8\}, & \{\{6,7,8\},\{7,8,9\}\}, & \ldots, & \text{for } i > 1,\ w_{i+1} = \{w_i, w'_i\} \\
w' & = & 8, & \{7,8,9\}, & \{\{7,8,9\},\{8,9,10\}\}, & \ldots, & \text{for } i > 1,\ w'_{i+1} = \{w'_i\} \\
w'' & = & 9, & \{8,9,10\}, & \{\{7,8,9\},\{8,9,10\}\}, & \ldots, & \text{for } i > 1,\ w''_{i+1} = \{w''_i\}
\end{array}
$$

$$w \longrightarrow w' \longleftrightarrow w''$$

Given the definition of the accessibility relation, and the structure of these points, the accessibility relation on the points is represented by the diagram. Each point will be accessible to itself. The arrows indicate which points are accessible to a point in addition to itself.

Now, a 'D'-less sentence is true at a point w just in case it is classically true at w_0. A sentence DA is true at a point just in case A is true at every point accessible to it. A sentence is true in a model, just in case it is "super-true", that is true at every point in the model. Since not every point in a model need be accessible to every other point, a sentence DA may be true at some points but not at others, hence neither (super) true, nor (super) false. For example, in the model given, 'Definitely: eight grains of sand can form a heap' is true at w but false at w', hence neither true nor false absolutely.

The logic which now results from the semantics given for the definiteness operator is analogous to the logic for the necessity operator given by the modal system T. If we exchange 'D's for '$\Box$'s, then the sentences validated by the semantics for 'definitely' are the theorems of T. The consequence relation is not the same for the two systems, however. Given the definition of truth as truth at *every* point in a model, as opposed to truth at some

privileged point in a model (the "actual world"), an inference from A to DA will always be supervaluationally valid. But DA will not always be a consequence of A in the modal system T, but rather only in the special case when A is itself valid.

Vagueness of any order is now a formal possibility. Just as in the modal system T, iterations of 'definitely' are not redundant. Neither $DA \to DDA$ (the analog of the S4 schema), nor $\neg DA \to D\neg DA$ (the analog of the S5 schema) comes out valid, as long as *every* model of the type described is considered relevant for determining validity. The explanation now offered of our inability to locate a boundary between the things which are definitely heaps and the things which are not definitely heaps, is that not everything falls definitely into one of these categories. Between the things that are definitely definitely heaps and those that are definitely not definitely heaps, will be some things which are neither—a class of second-level borderline cases. If higher-order vagueness never runs out, then no matter how many times we iterate the 'definitely' operator, there may always be a class of borderline cases between the things that are *definitely* definitely $\cdots$ definitely heaps, and those that are *definitely not* definitely $\cdots$ definitely heaps.

But we are not interested in *every* model of the type described. We are only interested in those in which the set of points in the model can be construed as the set of *admissible* points. A point $w_0, w_1, w_2, \ldots$ "is admissible," writes Fine, "if each of its terms are" (p. 293). It remains an open question whether considerations about admissibility leave intact those features of the account that are required in order to accommodate higher-order vagueness. I think we should be skeptical about the prospects. Sometimes, certain precisifications cannot be deemed admissible unless certain other precisifications are deemed admissible as well. For example, if **1,000** and **1,100** are admissible precisifications, **1,050** must be an admissible precisification as well. But it may also be that certain precisifications cannot be deemed admissible unless certain other precisifications are *not* deemed admissible. Look again at the following:

$$
\begin{array}{lclllll}
w & = & 7, & \{6,7,8\}, & \{\{6,7,8\},\{7,8,9\}\}, & \ldots, & \text{for } i > 1, w_{i+1} = \{w_i, w_i'\} \\
w' & = & 8, & \{7,8,9\}, & \{\{7,8,9\},\{8,9,10\}\}, & \ldots, & \text{for } i > 1, w_{i+1}' = \{w_i'\} \\
w'' & = & 9, & \{8,9,10\}, & \{\{7,8,9\},\{8,9,10\}\}, & \ldots, & \text{for } i > 1, w_{i+1}'' = \{w_i''\}
\end{array}
$$

If w is admissible then its first term is—**7** represents an admissible way of drawing a boundary between the Fs and the not-Fs. But if **7** is an admissible way of drawing that boundary, how could {**8,9,10**}—the second term of w''—be admissible? For it represents **7** as being an *in*admissible way of drawing that boundary. How could w and w'' be simultaneously admissible? The intended answer, presumably, is that just as what is possible from one perspective may not be possible from another perspective, what is admissible from one perspective may not be admissible from another perspective. But there is an important disanalogy. Possible worlds can be coherently thought of as perspectives, since they are the kind of things we could occupy—they represent a way things could be. But it is difficult to see how these structured points in the supervaluational model can be understood as the type of thing from which there is a point of view. Unlike in the original, simpler models for supervaluations, these structured points do not actually represent ways our language could be if it were made more precise. And given the supervaluationist's rejection of bivalence, they certainly cannot represent ways things could be given that our language is not precise, since every sentence is either or true or false in these points.

Were we to add the constraint on admissibility towards which I gestured—namely, that whenever a point $w_0, w_1, w_2, \ldots$ is in a model, a point $w_0', w_1', w_2', \ldots$ cannot be in that model unless, for each i, w_i is in w_{i+1}'—were we to add this constraint on the class of models relevant for determining validity, the accessibility relation would turn out to be symmetric and transitive after all. The possibility of higher-order vagueness would be ruled out. At the very least, without *some* account of what would make two points simultaneously admissible, we have yet no reason to accept that the possibility of higher-order vagueness is ruled in.

But even if we grant, for the sake of argument, that the supervaluationist *has* accounted for higher-order vagueness, we are still without an answer to the question why are we unable to say which instances of a tolerance principle for a vague predicate are not true. The account of higher-order vagueness provides an explanation only of why we cannot locate boundaries dividing things of this sort from things of that sort—why we cannot locate boundaries dividing things of which 'Definitely F' is true from things of which 'Not

definitely F' is true; why we cannot locate boundaries dividing things of which 'Definitely definitely F' is true from things of which 'Not definitely definitely F' is true. The proposed answer, is that there are always things of a third sort in between. (This is analogous to: Why can we not find a boundary between the girls and the boys? Because the teacher is in between.) But we still have no answer to the question, why can we not locate a boundary dividing things of this sort from *everything else*? In particular, why can we not locate a boundary dividing the things of which 'Definitely F' is true from the things of which it is *not* true. Despite the sophisticated machinery, the supervaluationist is *still* committed to the existence of such a point, and cannot explain why we cannot find it.

The picture he offers us is this. We let A be 'n grains of sand can form a heap'. Vertical lines represent boundaries.

$$
\begin{array}{lll}
\text{10,000 grains} \ldots & & \ldots \text{1 grain} \\
A & \Big| & \Big| \leftarrow \neg A \\
DA & \Big| & \Big| \leftarrow \neg DA \\
DDA & \Big| & \Big| \leftarrow \neg DDA \\
DDDA & \Big| & \Big| \leftarrow \neg DDDA \\
\quad \vdots & \Big| & \quad \vdots \\
DDD\cdots DDDDA & \Big| & \Big| \leftarrow \neg DDD\cdots DDDDA
\end{array}
$$
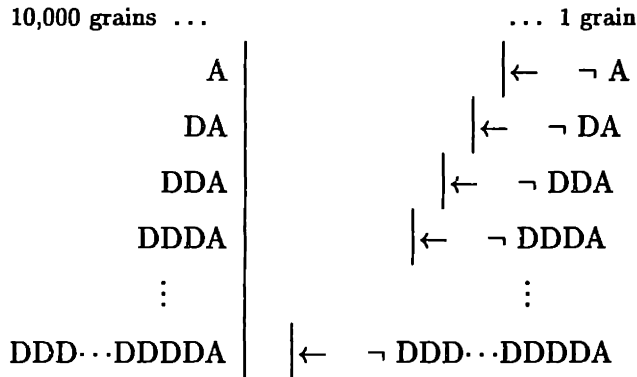
The class of things of which A is true, will be precisely the class of things of which DA is true (since DA is true whenever A is), which will be precisely the class of things of which $D\cdots DA$ is true, for any number of iterations of the definiteness operator. The class of things of which $\neg A$ is true may be a proper subclass of the class of things of which $\neg DA$ is true, which in turn may be a proper subclass of the class of things of which $\neg DDA$ is true. But since our series is finite, for some n the class of things of which $\neg D^n A$ is true must be precisely the class of things of which which $\neg DD^n A$ is true (where 'D^n' abbreviates n iterations of 'D'). But if higher-order vagueness never runs out, the gap may never close between the things of which $D^n A$ is true and those of which $\neg D^n A$ is, for any n. That is why we will not be able to locate a boundary straddled by things of the one sort on one side

and the things of the other sort on the other. But we have no explanation of why we cannot locate a boundary between the things of the first sort, and everything else. I provide below a simple model in which higher-order vagueness never runs out. It contains just these six points:

$$
\begin{array}{llll}
a = & 5, & \{5,6\}, & \{\{5,6\},\{6,7\}\}, & \ldots, & a_{i+1} = \{a_i, b_i\} \\
b = & 6, & \{6,7\}, & \{\{6,7\},\{7,8\}\}, & \ldots, & b_{i+1} = \{b_i, c_i\} \\
c = & 7, & \{7,8\}, & \{\{7,8\},\{8,9\}\}, & \ldots, & c_{i+1} = \{c_i, d_i\} \\
d = & 8, & \{8,9\}, & \{\{8,9\},\{9\}\}, & \ldots, & d_{i+1} = \{d_i, e_i\} \\
e = & 9, & \{9\}, & \{\{9\}\}, & \ldots, & e_{i+1} = \{e_i\} \\
f = & 10, & \{9,10\}, & \{\{9\},\{9,10\}\}, & \ldots, & f_{i+1} = \{e_i, f_i\}
\end{array}
$$

The accessibility relation, in addition to being reflexive, will be this:

$$ a \longrightarrow b \longrightarrow c \longrightarrow d \longrightarrow e \longleftarrow f $$

The class of things of which $D^n A$ is true is $\{n : n \geq 10\}$, for any n. The other classes are the following:

$$
\begin{array}{rcl}
\neg A & = & \{n : n \leq 4\} \\
\neg DA & = & \{n : n \leq 5\} \\
\neg DDA & = & \{n : n \leq 6\} \\
\neg DDDA & = & \{n : n \leq 7\} \\
\neg DDDDD^n A & = & \{n : n \leq 8\}, \text{ for any } n
\end{array}
$$

Even though higher-order vagueness never runs out, there is still a sharp boundary (between 10 and 9) between the things of which 'DA' is true and the things of which it isn't. Thus 'if 10 grains of sand can form a heap then 9 grains of sand can form a heap' would be that instance of the tolerance principle with a true antecedent but an untrue consequent. The commitment to such an instance has not been discharged, and no explanation is provided for why we cannot identify such an instance.

2.3 Speech Acts and the Sorites

In our discussion of supervaluations, it emerged that once it is taken for granted that the existence of borderline cases for vague predicates leads to truth-value gaps, we face a dilemma. The dilemma is that once we admit the gappiness of atomic sentences, we find

we have conflicting intuitions about the truth-conditions of complex sentences. On the one hand, we have the intuition that the connectives are truth-functional. We think that if a disjunction or existential generalization is true, then one of its disjuncts or instances must be true; we think that if a conjunction or universal generalization is false, then one of its conjuncts or instances must be false. On the other hand, we are inclined to regard some complex sentences as true, others as false, in virtue of what they mean, regardless of whether their constituents have a truth-value. If Jim is taller than Eric, then we are inclined to regard 'Jim is tall if Eric is' as *true*, even when both Jim and Eric are borderline cases. Following Kit Fine, Jamie Tappenden (1993) calls our intuition that such sentences are true the *penumbral intuition*.

It is easy to see that once we admit truth-value gaps (but only then!), our two intuitions about truth-conditions conflict. When Jim is taller than Eric, we still may not be inclined to regard 'Jim is *bald* if Eric is' even when both Jim and Eric are borderline cases of 'bald'. But this sentence has the same logical structure as the previous example. One way to resolve the conflict between the "truth-functional intuition" and the penumbral intuition would be to reject the existence of truth-value gaps. But they are being taken for granted.

A justification for supervaluational semantics is that it manages to accommodate the penumbral intuition, even if at the expense of the truth-functional intuition. Tappenden (1993), although he finds the truth-functional intuition to be fairly strong, objects to supervaluations on another front. Tappenden considers tolerance principles for vague predicates to be one of those "penumbral" sentences which we were inclined to think true, regardless of whether their constituents have a truth-value. He would take a tolerance principle for 'roughly three feet in length', for example, to express the semantic requirement that 'roughly three feet in length' not have sharp boundaries.[33] Since supervaluational semantics *falsifies* tolerance principles for vague predicates, it does not do the work it was invoked to do, namely, to verify the penumbral sentences.

---

[33] *Cf.* p. 567

But of course, no semantics we propose should *verify* tolerance principles for vague predicates. Thus as long as they are taken to be among the penumbral sentences, the penumbral intuition that the penumbral sentences are all true cannot be sustained. The strength of the truth-functional intuition, combined with the fact that the penumbral intuition must be off the mark, leads Tappenden to reject the identification of truth with super-truth. He adopts instead a truth-functional semantics.[34]

But even if the penumbral sentences cannot all be verified, the demand that they be in some way distinguished remains in force. Tappenden provides an account of how the context-dependence of vague predicates makes it possible to use the resources of the supervaluational framework to meet this demand. His idea is that vague predicates start out life as partial—as neither applying, nor failing to apply, to some things. In certain situations, though, we may have reason to use them "more precisely."

> ... the extension of a given vague predicate varies according to circumstances. We augment and relax the precision of vague predicates. Sometimes more things, and sometimes fewer things are counted as instances. This happens all the time: if a certain predicate is appropriate for some linguistic task, but insufficiently finely calibrated, we may lay down a more precise delineation by fiat. We might just say: "Let's count these among the heavy ones." (Said of some unclear cases.) The extent of the indeterminacy a predicate exhibits may be reduced or increased in a given context. (554)

A predicate is context-dependent, in the sense just indicated, if its extension varies as it is *stipulated* that it is going to apply to this or that thing. One would have hesitated, though, to think of such changes in extension as signaling what we normally think of as *context-dependence*. For the extension of any expression will vary if in a given context we stipulate that we are going to use it in an idiosyncratic way. If I stipulate that my coffee cup is going to count as a *shoe*, and I succeed in saying things about my "shoe," in getting refills of my "shoe," does that mean that *'shoe'* is context-dependent? If we construe the notion of a context-parameter so broadly as to include explicit stipulations, which after

---

[34]In particular, he accepts Kleene's strong three-valued tables for the connectives (see p. 27 above), and the natural extension of these tables to the the quantifiers.

all just seem arbitrary, does not the notion of context-dependence cease to be interesting? One might think. But not so, for not all stipulations are as arbitrary as in the 'shoe' case. Sometimes we can, and do, make stipulations that are not idiosyncratic, but are sanctioned by our linguistic practices as they stand. In the case of vague predicates in particular, Tappenden believes that we may "augment and relax" their precision, even by fiat, without a breach of semantic rules.

In supervaluational semantics, a precisification is an abstract mathematical structure. In Tappenden's account, a precisification is something we do; its result, a change in the propositions expressed by sentences of our language. But just as in supervaluational semantics, there are normative constraints. Some precisifications are admissible while other are not. I may say, "Let's count these among the heavy ones," while pointing to a bunch of 25lb. dumbbells, but if I say, "Let's count these among the yellow ones," while pointing to my blue-jeans, I've done something unacceptable. Since stipulations change the extension of an expression, they also succeed in changing the truth-values of sentences. One natural way to think of the normative constraints is as taking the form of a requirement that certain sentences not be falsified by a stipulation. The sentences that may not be falsified Tappenden calls *pre-analytic.*

To say that a sentence cannot admissibly be falsified is to say that it is not false in any admissible context—either in a context where no stipulations have been made, or in one where a predicate has been stipulatively sharpened. Sentences are evaluated with respect to a single context: the value of a sentence in a context is the value assigned it by the strong Kleene scheme. But a sentence, in addition to having the truth-value it has in a context, may be viewed on this account as also having one or more *modal* values: cannot be falsified, may be verified, etc.

Given the assumption that some tolerance principles for vague predicates are among the penumbral sentences, we know that our intuition that the penumbral sentences are all true cannot be accommodated, since from the assumption that a tolerance principle is true we can derive absurd consequences. We were left with the problem that tolerance principles

must nonetheless be in some way distinguished from other claims that cannot be true. But now, with Tappenden's account of the context-dependence of vague predicates, we have at our disposal the distinction between the truth-value a sentence has in a given context and the modal value it has relative to all contexts. With this distinction in hand, a revised version of the penumbral intuition may be accommodated without eschewing the truth-functional intuition. Instead of saying that the penumbral sentences must all be true, we may say that they cannot be *falsified*—that they are, that is, pre-analytic.

But what of the intuition that the sorites premise is *true*? Have we accounted for that intuition with the discovery that it cannot be falsified? Tappenden argues that the modal-value *cannot be falsified* supports a special kind of speech act. A sentence that cannot be falsified may on some occasions appropriately be uttered precisely because it has that value. The story might be told as follows: We left something out of our account of the context-dependence of vague predicates. We noted that some sharpenings of vague predicates are admissible while others are not. We identified the constraints on admissibility, but we did not explain how those constraints could be enforced. This part of the picture still needs to be filled in.

The suggestion as to how the constraints come to be enforced is that if someone proposes an inadmissible sharpening of a predicate, one that falsifies some pre-analytic sentence, we may induce them to retract their proposal by uttering the very sentence they have inadmissibly falsified. Suppose we are inspecting a sorites series of men, varying in height from 5'6 to 6', and you decide to propose a sharp boundary between the ones who are roughly 6' tall and the ones who are not roughly 6' tall. Well, in Tappenden's view, you have done something inadmissible and I may try to induce you to retract your proposal by uttering "But if a man is *roughly* 6' tall, then so is any man only one millimeter shorter than him." I had better not, however, have asserted the sentence. I better not, that is, have been trying to convey to you what I believe to be the case, since the sentence I have uttered is not true—cannot be true. But if the point of my utterance was just to enforce the constraint that you defied, then we need not say that I have asserted it. Rather, I've

performed some other speech act. In Tappenden's terms, I have *articulated* it.

The accounted is well-suited to provide an answer to the second on my list of questions: if tolerance principles for vague predicates are not true, why are we so attracted to them in the first place? The answer is that we mistake our good reasons for *uttering* tolerance principles on certain occasions as reason to believe them. We think they may be correctly asserted, because they may on occasion be correctly *articulated*. Because some tolerance principles may not admissibly be falsified, it will also be the case that sharp boundaries claims for vague predicates may not admissibly be verified. Sharp boundaries claims for vague predicates are not true in any context. Thus the answer to the third of my four questions is that despite the Sorites paradox, we are not required to accept such claims; in any context, a vague predicate will have at least some borderline cases, in the sense that there will be things (possible things if not actual things) to which they neither apply nor fail to apply.

One may wonder, though, whether the articulability of tolerance principles really does explain why are we attracted to them. A problem for the explanation, unacknowledged by Tappenden, is that a tolerance principle for a vague predicate may have exactly the same privileged status as its denial. It may be inadmissible to stipulatively use vague predicates in such a way that falsifies our tolerance principle for 'heap', for example. But it is also inadmissible to stipulatively use vague predicates in such a way that falsifies the claim that that for some n, n grains of sand can form a heap while $(n - 1)$ grains cannot. Given the semantics Tappenden accepts, the sharp boundaries claim could only be false if either every number of grains could form a heap, or no number of grains could. But surely no such stipulation is admissible, as in fact it is not by Tappenden's own lights. He requires that clear cases be preserved by stipulation—if a sentence has a truth value in absence of any stipulation, it must have that same truth-value on any admissible precisification. Thus if someone stipulates that 1,000 grains of sand is to count as being able to form a heap, and also they stipulate that if a number of grains is to count as being able to form a heap, then so is a number one less than it, then they have done something inadmissible. We may

induce them to retract their stipulation by uttering, "But look, there's got to be a cut-off point *somewhere.*" The proposed account of the context-dependence of vague predicates, the modal values a sentence has relative to those contexts, and the speech acts those modal values support, is unable to maintain a distinguished status for tolerance principles that it does not also ascribe to sharp boundaries claims. Since it turns out, on this account, that a sentence we are inclined to *reject* has the very same privileged status as sentences we are inclined to accept, the status can no longer be cited as explanation of our inclinations, and the account loses much of its appeal.

Another problem for the account is the limited extent of its application. It only covers a special class of vague predicates, those which Tappenden would, I gather, call "essentially vague." Examples are 'roughly three feet in length', 'shortish', 'about 12:00'. For these predicates there genuinely does seem something wrong with stipulating sharp boundaries. It does seem that someone does something impermissible by saying, "Show up for dinner about 7:00, by which I mean no earlier than 6:56 and no later than 7:12." But with other vague predicates, for example 'too heavy to be lifted safely', it seems perfectly admissible to stipulate sharp boundaries on occasion. Tappenden thinks so too. He writes:

> Someone drafting occupational health and safety regulations might specify precise boundaries by writing something like: "All objects too heavy to be lifted safely (i.e., over 90 kg.) must be moved with a forklift.

But clearly, the stipulation falsifies the sentence: if a weight is too heavy to be lifted safely, then so is a weight one milligram less than it. If the stipulation was admissible, then the sentence is not pre-analytic, hence not articulable. But still, in absence of stipulations like those above, aren't we tempted to think it true?

The only way to save the account is to change our minds and deem the above stipulation inadmissible after all. But what could be the justification for doing this? Certainly the stipulation *seems* admissible. Our intuitions about which stipulations are admissible is what we are taking to be a starting point. It is these intuitions that we are aiming to describe and explain. As it turned out, according to Tappenden anyway, the description

and explanation of the phenomenon that some stipulations are admissible and others not, provided us with a resource for explaining why sorites premises have such a grip on us. The option is not available to us to turn the account upsidedown and falsify the phenomena we were trying to describe in order to keep this account of the sorites.

2.4   Vagueness as Ignorance

All of the accounts we have so far considered reject bivalence for sentences containing vague predicates in order to account for the existence of borderline cases. None of the accounts we have so far considered has been able to provide a satisfactory answer to the first on my list of questions: Why are we unable to say which instance or instances of the tolerance principle for 'heap' (or any other vague predicate) are not true? One justification for the rejection of bivalence given our inability to decide in some cases whether or not a vague predicate applies, is the view that in such cases there no hidden facts; our knowledge of the meaning of vague predicates makes it always possible in principle for us to be in a position either to know that a predicate applies to an object if it does apply to that object, or to know that it doesn't if it doesn't. Our inability to say which instances of the tolerance principle are not true, however, should cast doubt on this idea. If we were always able to place a 'heap' sentence in the true category if it belongs in that category, and the "everything else" category when it doesn't, then we would be able to say just which instances of the tolerance principle are not true.

    This has led some philosophers to believe that grounds for rejecting bivalence in the face of borderline cases are not solid, while the costs of rejecting bivalence are too high. Given bivalence, the explanation of our inability to locate boundaries of any sort for vague predicates is that in some cases the facts about the applicability of a vague predicate are simply unknown to us. Timothy Williamson (1994) is the first proponent of this epistemic view of vagueness to provide an explanation of our ignorance in such cases.[35] The claim is that our ignorance is a consequence of the fact that we have inexact knowledge of which

---

[35]Other advocates of the epistemic view are Sorensen (1988), Cambell (1974), and Cargile (1969).

properties are expressed by vague predicates. The extensions of vague predicates are not determined by natural or stipulated boundaries, but rather in some mysterious way by the total pattern of their use. Our learning of vague predicates is essentially limited, and so the possibility that vague predicates might express any one of number of very similar properties is compatible with our experience and training in the use of them. It might just be that 'tall' (for a man) applies to a man just in case in is at least 6′1. But it is possible that we might have had just the same experiences we in fact had, but due to a slight difference in the overall use of the predicate, the cut-off point might have been just slightly lower instead. Thus we cannot know that it expresses the one property rather than the other.

But if 'tall' expresses a precise property, and we don't know which property that is, how can it be that we know that a 6'5 person is tall, or that a 5'5 person is not? Williamson argues that margin for error principles that govern inexact knowledge in general, also govern inexact knowledge of meaning in particular. Suppose you're at Fenway Park. You know that the seating capacity is about 34,000. You can see that there are very few empty seats. You know that there are more than 20,000 people in the stadium, and less than 50,000. But you do not know that there are not exactly 33,000. If there are exactly 33,000 people, then you will not know that there not exactly 32,999 people. The scene that would be before your eyes if there were 32,999 people would be too similar to the scene in fact before your eyes for you reliably to be able to judge the difference. A margin of error is required for your judgment to be a reliable one. The situation is similar in the case of our inexact knowledge of vague predicates. We know that 'tall' does not apply to a man of 5′5. We know that does apply to a man of 6′5. But if a man of 6′1 is the shortest man to which 'tall' applies, we cannot know this. Had it instead been the case instead that a man of $6'0\frac{9}{10}$ was the shortest man to which 'tall' applies, our training in the use of the predicate would not have been different enough, if different at all, for us reliably to be able to judge that 'tall' expressed the one property rather than the other. Our knowledge of the applicability of 'tall' gives out as we approach the boundary.

Given Williamson's acceptance of classical logic and semantics, the tolerance prin-

ciple for 'heap' (for example) will have exactly one untrue instance. That instance will have a true antecedent and a false consequent. Williamson is able to explain why we cannot discover which instance that is, where the accounts we have previously considered have not been able to. We cannot discover the false instance of the tolerance principle, since in order to know of the false instance that it is the false instance, we would have to know that its antecedent is true and that its consequent false—we would have to know where the boundary is between those numbers of grains of sand that can form a heap, and those numbers of grains that cannot. But in order to know this, we would have know that 'heap' did *not* express some very similar property. But this we cannot know, because our knowledge of the truth-conditions for sentences containing 'heap' is inexact. Williamson requires us to accept sharp boundaries claims for vague predicates but, in answer to my third question, argues that the existence of sharp boundaries is compatible with the existence of borderline cases. There is always a fact of the matter about whether a thing is heap. The essence of borderline cases is our inability to know the facts.

Williamson's solution of the Sorites paradox is incomplete, however. It provides no explanation of why we were so attracted to the tolerance principle in the first place. I have inexact knowledge of the number of people at the ballpark. But I am in no way inclined to believe that if there are n people, then there are also $(n-1)$ people. Nor does Williamson's provide any indication of whether we can recover, in some revised form, a version of the thought that adjacent members of a sorites series must have the same status. In the next chapter, I will propose answers to all of my four questions. I concur with Williamson's view, that in order to explain why we cannot locate the boundaries for vague predicates, some ignorance of facts must be postulated. I do not find, however, that our ignorance has just the same source that Williamson finds it to have.

Chapter 3

A Solution to the Sorites Nearly Everyone Could Accept

In this chapter I present a solution to the Sorites paradox that appeals to a certain kind of context-dependence. It is frequently acknowledged that vague predicates are context-dependent—that for example, an utterance of 'Mr. X is tall' might express a truth in most situations, and yet not express a truth, if Mr. X is a basketball player, during a conversation about the NBA draft. Our initial description of the phenomenon might be that in some contexts 'tall' means *tall for a man*, while in other contexts it may mean *tall for a basketball player*. The content of the utterance is not constant. It must seem to many philosophers that the context-dependence of vague predicates could be of no use in solving the Sorites, since even when contexts are fixed, the paradox still arises. We are as inclined to accept a tolerance principle for the predicate 'tall for a basketball player' as we are inclined to accept one for 'tall', and for that reason it must seem acceptable when dealing with the Sorites to treat vague predicates as if they were not context-dependent at all. I shall argue, however, that vague predicates are context-dependent in more than one way; that in particular they are context-dependent in a way that cannot be abstracted from when dealing with the Sorites, and which does provide resources for solving it.

3.1 Comparison Classes and Standards of "Precision"

In order to get a feel for the ways in which vague predicates can be context-dependent, I would like to begin with a brief comment on David Lewis's (1979) discussion of vagueness in 'Scorekeeping in a Language Game'.[1] Lewis describes the context-dependence of vague predicates as a variation in the "standards of precision" that are "in force" in a conversation:

> Austin's "France is hexagonal" is a good example .... Under low standards of precision it is acceptable. Raise the standards of precision and it loses its acceptability (245).

In Lewis's sense, to increase our standards of precision for a vague term is to count fewer things among its positive instances. This is to be contrasted with increases in precision in Tappenden's sense, where to increase our standards of precision for a vague term is to count *more* things among its *definite* instances—to narrow the gap between its extension and anti-extension.

It is misleading, however, to think of the differing standards, in Lewis's sense, as being standards of *precision*. The reason it may seem appropriate to describe them that way, is that 'hexagonal', like the other vague predicates Lewis discusses—'bald' and 'flat', have precise uses. In the limit case, 'bald' may mean *has absolutely no scalp hair*, and 'flat' may mean *has absolutely no bumps*. The more hair or bumps we permit a thing to have while still counting it as bald or flat, the less precisely (we might say) we are using the expressions. The use may be called less precise because it is less like the precise use. But any departure from precision is really as imprecise as any other.

Moreover, vague predicates like 'tall' or 'rich' which do not have limit cases, and hence cannot have precise uses, may still be used with varying "standards of precision" in Lewis's sense, because on different occasions more or fewer things may be counted as positive instances. There is no precise use of 'tall' or 'rich', however, to which these different uses can be compared. It is probably better, then, to think of the differing standards in the use of vague predicates as being stricter or more lax, and not as more or less precise.

---

[1] All page references will be to the reprint of the article in (Lewis 1983).

Using vague predicates with different standards in different circumstances is a fairly widespread practice, I think, and not an occasional indulgence. The ends of communication are so varied that is very economical to have a language that licenses variations in the standards of use of its predicates. We speak to explain why things happened ("He voted for the proposal because he's rich"), to get people to do things ("Please hand me the red one"), to caution them ("That's slippery"). If we did not have some discretion to vary our standards as it suits our purposes, we would either have to have a very large vocabulary or speak very long sentences in order to achieve our communicative ends. If the book I want is sitting amongst a pile of green books, why should I be required to say, "Please hand me the one that's sort of a cross between brick red and burgundy but on the brownish side," when a mere 'red' would do.

One question to ask is whether the variation in our standards carries the truth along with it. If I say that the book is red on one occasion and not red on another, can it be that I have spoken truly both times? It has always been a guiding principle in the philosophy of language to take our disposition to utter a sentence in a situation as indication that the sentence is true in that situation. I find it difficult to see that the principle needs any defense; the burden of proof lies with those who deny it. It is not so much that we want to pursue a strategy of verifying as much of what people say as possible. People make mistakes all the time—they speak falsely all the time. Any account of the truth-conditions of our utterances that verified as much of what people say as possible would have it, for example, that 'The world is flat' has changed in meaning, and for that reason would surely be incorrect. Rather, we only apply the principle when it is safe to assume (or better, possible to ensure) that a speaker is *not* misinformed, misperceiving, misunderstanding, miscalculating or misremembering, and so on. Once it is understood that the principle has only this limited application, it should seem like a truism. Accepting it should seem no worse than accepting the principle that if a sentence would be uttered in a situation by an honest and omniscient speaker of our language—one who has communicative ends like ours, and also an interest like ours in economy of speech—if a sentence would be uttered by such

a speaker in a situation, then we are to take that as indication that the utterance is true in that situation.

An important shift in the philosophy of language was to reject a corresponding principle that if a sentence would be *refrained* from being uttered in a situation by an honest and informed speaker, then we are to take that as indication that the utterance is *not* true in that situation. Our resistance to uttering certain sentences in certain situations can often be explained without appeal to either lack of truth on the part of the utterance or lack of information on the part of the utterer, when conversation is understood as a form of rational, cooperative behavior.[2] Once the corresponding principle is rejected, however, one might wonder whether the original principle should be rejected on similar grounds: if we often refrain from uttering true sentences when so doing would not contribute effectively to the cooperative enterprise, then perhaps we also often utter *untrue* sentences just because doing so *would* contribute effectively to the cooperative enterprise. I cannot undertake a full-scale defense of my position here, but my position is to accept the original principle— that is, I will take the sincere, informed utterance of a sentence in a situation to be an indication of the truth of the utterance in that situation; hence I assume that the variation in our standards for, *e.g.*, 'red' does carry the truth along with it. Accepting the principle proves fruitful, and anyway, not accepting the principle would leave us with little else to go on.

Another question we should ask is whether every variation in the standards of use of vague predicates should be accounted for by a variation in comparison class. We opened the chapter by noting that sometimes 'tall' may be used to mean *tall for a basketball player*, while other times it may be used to mean just *tall for a man*. Different comparison classes for 'tall' may be salient in different contexts. When we say that France is hexagonal, we may mean something like *hexagonal for a country*. But when I say, "*LBM* is the red one" when it is sitting amongst a pile of dark-green books, but "*LBM* is the brown one" when it is sitting amongst a pile of fire-engine red books, should we also attribute this to a difference in the

---

[2]*Cf.*, for example, Grice's *Logic and Conversation* in Grice (1989) and Stalnaker (1973, 1974, 1975).

salience of one or another comparison class? It is difficult to answer this question without a theory in hand of exactly *how* the salience of comparison classes affects the content of an utterance. But I would like to give some reasons why, in the abstract, we should suspect that not every variation in standards can be made to fit into the comparison class mold.

If the variation in our standards for 'red', as displayed in the *LBM* example, is to be accounted for by a variation in comparison class, then the comparison class that is salient in each case must be something like: the class of books *LBM* is piled with. Now assume that '*F* for class *C*' means 'meets some minimal standard for *F*-ness, and is also *F*-er than the mean average for class *C*'. Then 'tall for a basketball player' would be true of X just in case X meets some minimal standard for tallness, and X's height is taller than the mean height of basketball players. Also assume that it is possible somehow to average the color of a group of books, and that given the mean color of a pile containing *LBM* and dark-green books, the color of *LBM* is indeed redder than that mean, but not redder than the mean color of a pile containing *LBM* and fire-engine red books. If we make these assumptions, then it looks as if the variation in our standards for 'red' in this case can be accounted for as a variation in the salience of one or another comparison class. Nevertheless, I do not think the first assumption can be justified. The truth conditions for 'X is *F* for class *C*' cannot be a function just of the *mean F*-ness of members of *C*. Nor will choosing some measure other than the mean—such as the mode, or the mean average of the *F*-est 50%—do the trick. We can imagine the NBA, fifteen years from now, being composed *entirely* of tall players. It is not merely that we could *now* say without contradicting ourselves, "I wonder whether in fifteen years the NBA *will be* composed entirely of tall players," using the class of present players as our comparison class. We can imagine being able to say truly, fifteen years from now, "The NBA *is* composed entirely of tall players." The problem, in a nutshell, is that the satisfaction-conditions for predicates of the form 'is *F* for class *C*' require class *C* to form a *kind*, for they depend on some conception of what a typical member of the kind as a whole is like, and not just what is typical for its presently existing instances.

The problem is compounded when we consider variations in standards which if they were to be understood as standards relative to some comparison class, the comparison class would have to be a class containing only one member. Imagine that I and my friend Linda are two casting agents. I'm auditioning actors to play Mikhail Gorbachev in a movie. Linda is auditioning actors to play the part of Yul Brenner. We can imagine that Linda and I turn actors away all day, citing their lack of baldness as the reason. At the end of the day, at last, a perfect Gorbachev shows up, and I say, "Finally! Someone bald!" My perfect Gorbachev, however, is someone that Linda earlier turned away for the part of Yul, due to his lack of baldness. Linda and I have been using different standards. If the difference in standards is to be accounted for as a difference in comparison class, then we have only two choices for what that comparison class might be. One option is to take the comparison classes to be singletons in each case—the class containing Yul and the class containing Mikhail. The option is ruled out, I think, because we have no conception of what a *typical* member of a class is like when the class has only one member.

The other option would be to take the comparison classes to consist of people whose hair coverage is *like* Yul's or *like* Mikhail's. I think the strategy should not be pursued. If we try to make the variation in the standards of our use of 'bald' fit the comparison class mold, we will not be able to account for an important semantic difference between the two sorts of context-dependence.

To illustrate the difference, we will introduce another example. Imagine now that Linda and I are doing the casting for a documentary about tall pets. We spend the day auditioning pets, in hopes of finding tall cats, tall dogs, tall hamsters, tall canaries, and so on. But we have no luck. At the end of the day, when I arrive home and am asked how my day was, I complain, "It was horrible. None of the animals we auditioned were *tall*." Evidently, this is an appropriate way of conveying that none of the mice we auditioned were tall for a mouse, and none of the iguanas we auditioned were tall for an iguana, and so on. The logical form of the sentence I uttered is not $\neg \exists x (x$ was tall for $C)$, but rather something more like: $\neg \exists x (x$ was tall for x's type$)$. There is not a unique, most salient

comparison class that is relevant for evaluating the utterance. The comparison class seems to occupy a position that can be bound by the existential quantifier.[3]

Now return to the example where Linda and I are holding auditions for the parts of Yul Brenner and Gorbachev. But this time suppose that my perfect Gorbachev goes for a drink after being rejected for the part of Yul, and never returns to try out for the part he was made for. Now when I return home and am asked how my day was, I cannot claim, "It was horrible. None of the actors we auditioned were *bald.*" I may say that none of the actors we auditioned were bald *enough*, but I cannot say simply that none of them were bald. Why? That would imply that we wanted someone bald to play Gorby, and that by the very same standard none of the Yul auditioners were bald. But that is not what happened. Unlike the comparison classes for 'tall' in the pet case, the different standards for 'bald' used by Linda and me do not seem to occupy a position that can be bound by the existential quantifier. We should conclude that two different sorts of context-dependent are at work in the two examples. This squares well with the intuitive appeal of saying that by 'bald' Linda and I both meant *bald for a man.*

The above example shows, I think, that keeping comparison classes fixed from context to context will not be sufficient to render vague predicates context-invariant. The predicates 'bald for a man' and 'tall for a basketball player', just like 'bald' and 'tall', are context-dependent since they may be used with more or less strict standards on different occasions. Once this is recognized, the context-dependence of vague predicates cannot be so easily ignored when discussing the Sorites paradox, since it is not the case that for each of the different standards of use of 'tall', for example, there is some context-invariant complex predicate in the language which can be argued to be equally vague. Thus it cannot be assumed that the Sorites paradox still arises once all contextual elements relevant to determining the truth-conditions of sentences containing vague predicates are fixed. My view is that not only can this not be assumed, it is precisely not the case.

---

[3]For discussion of binding hidden variables, and whether they should be syntactically represented, see Partee (1989).

3.2 Constraints on Standards and the Sorites Paradox

We have a lot of leeway in the standards we use for vague predicates, but still, we cannot use them any old way we like. What are the constraints? Well first, there are what we may call *normative* constraints. For each predicate, there will be a limited range of cases which it will be permissible to count as positive instances. We can never use the word 'green' in such as way as to apply to the color of the sun. For each predicate there will also be a class of things which it will be mandatory to count as positive instances. No matter what standard is in place for 'blue', the predicate applies to the color of a clear afternoon sky. There will also be *relational* constraints for some predicates: Whatever standard is in place for 'tall', anything *the same height as or taller than* something that meets the standard itself meets the standard.[4] A further sort of constraint will *coordinate* the standards in use for related predicates: Whatever standards are in use for 'rich' and 'poor', nothing can meet both, and it must be possible for something to meet neither.

Are these the only sorts of constraints? I think not. The standards in use for a predicate may be constrained by the nature of the immediate surroundings. Here's an example. You and I are waiting in the airport, and I want to make some remark about a man in the vicinity. I can't point, because he will see me. So while looking away ever so casually, I whisper to you, "See that guy over there by the water fountain, the one with mustache?" But by now, there are three men by the water fountain, two of which have a mustache. "Which one?" you ask, "Is he tall?" I have a choice to make. If my man is appreciably taller than the other man with a mustache, then I may answer "yes." Alternatively, I could answer, "He's not *tall*, but he's around 5'11." My response sets the standard.

Now suppose there are two men talking to each other, whom I wish to point out to you. You ask, "Are they tall?" If the two men are pretty much the same height, I may not choose a standard that one meets but the other doesn't, even if one is just noticeably taller.

---

[4] A qualification may be required here. If basketball players are the salient comparison class, then a sky-craper may not meet the standard in place for 'tall'. Something cannot be tall for a basketball player unless it is a basketball player—I don't *think*.

The option is not available to me to set a standard that divides the two.

On the other hand, we may use standards for vague predicates that do divide similar pairs, as long as the similarity of the pair can for some reason be set aside, or ignored. Imagine an eccentric art collector who keeps all and only her paintings containing just red pigments in one room, and all and only her paintings containing just orange pigments in another. One day she is presented as a gift a painted color spectrum ranging from primary red on one end, to orange on the other. She resolves to cut the canvas in half. If she cuts without thinking, perhaps in a state of mad excitement because she is so eccentric, she will most likely cut in just the right place, and once the halves are re-framed and hung, she may still proclaim, with pride, that all and only her paintings containing just red pigments are in one room, and all and only her paintings containing just orange pigments are in another. If the decision about where to cut is labored, in contrast, she will likely find herself unable to locate the boundary between the red and the orange, the pigments on either side of any proposed cut being too obviously similar for one to go in the red room and the other in the orange.

The proposed account of what is going on in the preceding examples, is that vague predicates are context-dependent in a special sort of way. If the similarity of two objects is not relevant, then one may be in the extension of a vague predicate while the other is not. If their similarity is made relevant, however, the extension changes and they are either both in or both out.

What answers does the account provide to the list of questions I earlier specified as being *the* questions (p. 23) to answer when solving the Sorites? The first question I posed was: why can't we say which instances of the tolerance principle are not true? Our answer is that in considering any particular instance, say 'If 486 grains of sand can form a heap then so can 485', we invoke a context in which that instance is true, since by assessing it, we make the similarity of 486 and 485 relevant. On a related front, to answer the second question, we are attracted to the tolerance principle because it is very natural to think that a universal is true, given that each of its instances is true when we assess that instance. Our

mistake is that we don't recognize that the instances are not all true evaluated with respect to a single context. Though there may be a sharp boundary that divides the heaps from the non-heaps, we could not inspect that boundary without it ceasing to be a boundary, for by inspecting the boundary we change the context relative to which 'heap' is evaluated. If there are sharp boundaries, we cannot find them, because they are not stable.

We will put off answering the third question for the moment, and move on to the fourth: Can some revised version of the tolerance principle be recovered? When the question was initially posed, I suggested two forms that such a revision might take. The first suggestion was that the revision take the form of a qualification: all adjacent pairs of stages in our series that meet condition C always have the same status. The second suggested form was this: every stage bears some special relation R to adjacent stages, which it does not bear to more distant stages—where R should be in some way *like* 'has the same status as'. The degree-theorist proposed a revision taking the second form, by letting R be *has a status very similar to*. I propose a revision taking the first form by letting C be the condition that the similarity of the pair is relevant. It is not surprising that we would not have noticed the requirement that the condition obtain, because we can't think of an adjacent pair in the series that doesn't meet the requirement (while we're thinking of it, anyway).

Now what about the third question? Can we somehow, despite the Sorites paradox, resist accepting the sharp boundaries claim, and in some way maintain the idea that vague predicates have borderline cases, and if so how? I want to stress that nothing in my account, as so far presented, commits us one way or the other to a verdict on the existence of truth-value gaps, nor to a verdict on the truth of the sharp boundaries claim. I have so far not said anything about whether the standards in use of a vague predicate must cover all cases, in the sense that once a standard is in place, everything either meets the standard or it doesn't.[5] Nor have I said anything about whether once a standard is in place, different things can meet the standard to varying degrees. As far as I am concerned, my account can

---

[5]I did actually suggest at one point that the standards do cover all cases. But that was just for ease of exposition, and nothing in my answers to questions 1-3 depends on it.

and should be adopted by all parties.

Without an account of higher-order vagueness, the degree-theorist seemed committed to the existence of a sharp boundary between the things that satisfy a predicate to the highest degree, and those things that satisfy it to a degree less than one. Thus he seemed unable to explain why we cannot say which instances of the tolerance principle are (completely) true, and which instances are not. If the degree-theorist combines the account presented here with his own, then he can explain why we cannot find the boundary between the things that do and the things that don't satisfy a predicate to the highest degree. And he can do this without an account of higher-order vagueness, in fact without even accepting the existence of higher-order vagueness.

Just to give a quick sketch: the degree-theorist could admit that vague predicates are used with different standards on different occasions, and could adopt the constraints on standards proposed here, only in a modified form that accords with his understanding of truth. For example, the degree-theorist's modified normative constraint would be: for each predicate there will be some cases such that it is *permissible* to adopt a standard for the predicate which the cases meet to degree 1 (0); there will also be some cases such that it is *mandatory* to adopt a standard for the predicate which those cases meet to degree 1 (0). Our relevant similarity constraint would be put in the following way: if it is relevant in a context that a and b are similar (in the respect relevant for F), then if a satisfies F to degree 1 (0) in that context, so does b. The reason, then, that we cannot find that instance of the tolerance principle which has a perfectly true antecedent but not a perfectly true consequent, is that in a essing any instance, we invoke a context in which the standard for 'heap' is met to degree 1 by both or neither.

The supervaluation-theorist was in a worse-off position than the degree-theorist with respect to our list of questions. The supervaluation-theorist could not provide any explanation of why, if the tolerance principle is (super-) false, we are nonetheless so inclined to accept it. Nor was the supervaluation-theorist able to recover an acceptable revision of the tolerance principle. The supervaluationist's account of higher-order vagueness, moreover,

provided no explanation for our inability to say which instances of the tolerance principle are untrue.

The supervaluation-theorist can, however, also supplement his account of vagueness with ours. The two accounts, in combination, would then provide what I would regard as a complete solution to the Sorites. What would the combined account look like? In each context, the standard in use for a predicate has an extension (the things that meet the standard) and an anti-extension (the things that do not meet the standard), which are disjoint but not required to jointly exhaust the domain. The supervaluationist's definition of truth as super-truth may then be relativized to contexts.[6] In adding to his picture the account proposed here of how the context constrains the available choice of standards, the supervaluation-theorist loses nothing, but gains the resources to provide answers to all of questions 1-3, without being required to account for higher-order vagueness.

The epistemic-theorist was in somewhat of a better-off position than the supervaluationist, insofar as his view proved well-suited to explain why we cannot know which instance of the tolerance principle is untrue. But still the account was lacking. I think it should be obvious by now that there is no formal bar to combining the view advanced by Williamson with the account presented here. Instead of holding that a sentence containing a vague predicate expresses just the same bivalent proposition on every occasion of its use, and that we have inexact knowledge of what proposition that is due to our limited exposure to uses of the predicate, the combined view would hold that that a sentence containing a vague predicate may express different bivalent propositions on different occasions, and that on each occasion we have inexact knowledge of which proposition that is. It is this combined view that I accept.

On the view I affirm, then, no matter what standard is in use for a vague predicate, there will be things that meet the standard although we cannot know that they meet the standard, and there will be things that do not meet the standard although we cannot know

---

[6]Stalnaker (1987) and Lewis (1970, 1979) relativize supervaluations to contexts in this way. Lewis adds, in addition, non-degree-functional degrees of truth, where the degree of truth of a sentence is determined by a measure function on the set of admissible precisifications.

that they do not meet the standard. The severity of this position is rather softened though, I think, once one recognizes that it is compatible with the common sense view that when we know a person's height, for example, we know enough to correctly assess whether or not that person is tall. How could the position I take possibly be compatible with the common sense view? How could it be that we are always able to correctly assess whether a person is tall, when there are things that are unknowably tall and things that are unknowably not tall? In order to demonstrate the possibility of compatibility, I must provide the proper restatement of the common sense view: when we know a person's height, we know enough to correctly assess whether or not that person is tall, in the sense that *when* we do assess whether or not that person is tall, the standard in use for 'tall' will be such that we can know whether the person meets it. When we are not assessing whether that person is tall, however, when for whatever reason we are able to conveniently ignore the fact that his height is very similar to the height of people just slightly shorter or taller than him—when that height is out of our view, so to speak. The standard in use for 'tall' at that time may just be such that he is the shortest person to meet it. The picture is this: we cannot locate the boundaries for vague predicates because the boundaries will never be where we are looking, but the flip side is that wherever we are looking, we may know which side of the boundary we are on.

An initial worry about my proposal will be that it seems to require that there is always at least one pair of heights the similarity of which is not relevant—there must always be at least one place that the boundary between the tall and the not tall could be. But when we are *confronted* with a sorites series of men, ranging from five-foot on one end to seven-foot on the other, each differing in height from his neighbor by only a millimeter, then the similarity of every adjacent pair will be extremely relevant, and so given my constraint that boundaries cannot separate relevantly similar pairs, there will be no place for the boundary to be. But surely it cannot be that *none* of the men in the series are tall, and surely it must be that at least some of them are.

One option available to me would be to argue that although it is relevant that each man in the series is similar in height to his neighbor, it is not the case that each man in the series is such that it is relevant that *he* is similar in height to his neighbor. The claim would be that 'it is relevant that' does not commute inward across the universal quantifier. What is at issue is whether it can be argued that relevance is not closed under consequence— whether it may be relevant that a universal holds, without each of its instances being such that it is relevant that they do. Nevertheless, it just seems to me wrong-headed to deny that when confronted with the sorites series, (it needn't be in person), each man *is* relevantly similar to his neighbor. Perhaps we are unable, because of our limited capacities, to actively hold before our minds the similarity of each adjacent pair at any one given time. I can see no reason, though, why the active consideration of a fact should be a necessary condition of its relevance.

What then are we to do? If the similarity of every pair is relevant then the standard of use for 'tall' in that context must be one such that either every man meets the standard or every man fails to meet the standard. But we also have what I called the normative constraint on standards, which in this case rules out a standard that the shortest man meets, but also rules out a standard that the tallest man does not meet. In this context, every potential standard of use for 'tall' is ruled out by some constraint. The conclusion we must draw, then, is that in this context sentences containing 'tall' express no proposition at all. This may be very difficult to swallow. Isn't it *true* that the last man in the series is tall? And mustn't the claim that he is tall, then, say that something is the case? Well if you can say truly that the last man in the series is tall, then you must somehow be managing to ignore the similarity of at least one pair in the series. Otherwise, your inclination to regard the sentence as true must be accounted for by the strong pull of the normative constraints, the feeling of obligation that it places on you to adopt a standard which the last man meets. I do not think it should be an entirely unwelcome consequence of my view that when confronted with a sorites series, sentences containing vague predicates express no propositions. Didn't you feel that thinking about the Sorites paradox somehow screws

things up?

Another worry about the proposal will be that it does not really provide an explanation, in the way suggested, of our inability to locate the boundaries for vague predicates. *Perhaps* it is the case, the objection would run, that as long as we are *unaware* of the way in which our standards of use for vague predicates are constrained by context—perhaps as long as we are unaware of this, we will be unable to locate the boundaries of those standards, as a consequence of our lack of awareness that as we shift the focus of our attention, the standards in use of our predicates may change. But if we acknowledge that our standards change as we shift the focus of our attention, then we should have no reason to believe, just because the boundary between the tall and the not tall is not *here*, given my *present* standards, that it *was* not here, given my *previous* standards. Shouldn't we, moreover, while in a context c, be able to introduce a context-invariant predicate 'tall$_c$' that expresses just the same property that 'tall' does in c? I may say to myself, "I shall always use 'tall$_{now}$' with just the same standards with which I use 'tall' *now*." As I proceed to focus my attention on what was then the boundary for 'tall', I will, by stipulation, be focusing my attention on what is still the boundary for 'tall$_{now}$'. What explanation is there for my inability to recognize this boundary as the boundary for 'tall$_{now}$', and hence as the boundary for 'tall' as I used it then?

Our response is that our only means of insight into what the standards are for 'tall$_{now}$', is reflection upon what our standards were for 'tall' when 'tall$_{now}$' was introduced. Suppose, for *reductio*, that the cut-off height for 'tall$_{now}$' is at $6'1$. In order, therefore, for me to know that the cut-off height for 'tall$_{now}$' is at $6'1$, I would have to have known then that $6'1$ was the cut-off height for 'tall'. But if I knew it, I believed it. And if I believed that the cut-off height for 'tall' was at $6'1$, then the similarity of heights of $6'1$ and $6'1 - \varepsilon$ would have been relevant. So by the relevant similarity constraint for 'tall', $6'1$ was not the cut-off point. Contradiction. Note: I do not claim that everything we believe, or even know, at a time is relevant at that time. Quite the opposite, in fact. In order for our use

of vague terms to be successful, that is, to express properties, we are required to ignore—put out of our minds—at least some of what we know. Knowledge is a necessary but not sufficient condition for relevance. The principle appealed to in the *reductio*, was just the more restricted claim that *certain* beliefs are sufficient for the relevance of *certain* facts, specifically, that a belief of two similar heights that they straddle the boundary for 'tall', would be sufficient to make the similarity of those heights relevant.

3.3   The Pragmatics of Indeterminacy

So far I have left open the question of whether borderline cases lead to truth-value gaps. Nothing in my answers to questions one, two and four depended on it. But still, it seems that implicit in my account of the context-dependence of vague predicates, is a conception of what it is for an object a to be a borderline case of a predicate F that is compatible with either or the truth or falsity of $\ulcorner a$ is $F \urcorner$. In fact, there seems to be more than one such conception implicit in my account. On the one hand, I agreed with Williamson that that we are essentially ignorant of the extent of the applicability of vague predicates. In every context in which 'tall', for example expresses a property at all, there will be a least height that is a tall height,[7] and we cannot know which height that is. We cannot believe a height h to be the least tall height without it becoming relevant that a just slightly shorter height is too similar to h for it to fall on the other side of the boundary. So perhaps I should say, with Williamson, that what seemed to be indeterminacy is really a form of ignorance. On the other hand, I could adopt an idea of Crispin Wright's (1992), that what what it is for something to be a borderline case of a predicate is for it to be permissible but not mandatory to say that the predicate applies, and hence that two people may disagree about whether the predicate applies without either being guilty of a "cognitive shortcoming." If we have intuitions about whether it makes sense to think of something's being a borderline case *in* a context, or whether the notion of a borderline case is something that requires quantification over contexts, then we might be able to decide in favor of one or the other

---

[7]Or a greatest height that is a not-tall height.

conception.

When trying to answer the question what is it for something to be a borderline case of a vague predicate, I think it important to begin by saying just what phenomenon it is we are trying to account for. Some philosophers, such as Fine and Tappenden, just define borderline cases as those things of which a predicate is neither true nor false. Timothy Williamson regards it as better to proceed by giving examples. The first approach will not do since then it remains an open question whether vague predicates do have borderline cases, and so if vagueness is then defined as the possibility of having borderline cases, it would remain an open question whether any predicates are vague. Williamson's approach will not do either, since it seems impossible to find examples about which people can agree.

We are prompted to regard a thing as a borderline case of a predicate when it elicits in us one of a variety of related verbal behaviors. When asked, for example, whether a particular man is nice, we may give what may be called a hedging response. Hedging responses include: "He's nice-*ish*," "Well, it depends on how you look at it," "I wouldn't say he's nice, I wouldn't say he's not nice," "It could go either way," "He's kind of in between," "It's not that cut and dry," and even, "He's a borderline case." If it is demanded that a 'yes' or 'no' response is required, we may feel that neither answer would be correct, or that we are just unable to decide. In other cases, it may be that one person says that the man is nice, another says he is not, while we as third-party onlookers neither person is wrong, and yet it can't be that both are right.

In asking what it is for something to be a borderline case of a predicate, I think we should ask what might prompt one of this array of responses. There is no justification for assuming at the outset that it is always the same thing in every case. In fact, I think hedging responses may have a variety of causes.

Suppose you have been mugged, and are trying to provide a description of your assailant to the police. You are asked whether the man is bald. If the man had had no hair, you would most certainly respond 'yes', while if he had had a full head of hair you would most certainly respond 'no'. But the man was somewhere in the middle, so instead

you try to give as detailed a description of his hair cover as would be helpful. But now suppose you are on the witness stand, and are asked by the defense attorney whether your assailant was bald. You are permitted only to answer 'yes' or 'no', on pain of being in contempt of court. (The judge is strict to a fault.) You may find it difficult to answer one way the other, since either answer would be misleading. Given Grice's (1989) *maximum of quantity*: "Make your contribution as informative as is required (for the current purposes of the exchange)," a simple yes answer would lead your audience to believe that 'bald' is a sufficiently informative description of his hair cover, and hence they would be led to believe that your assailant is a prototypical bald man. Similar remarks may be made for a simple 'no' answer. Despite the odd rules governing interrogation of witnesses, conversational maxims remain in force. Thus in this case, neither 'yes' or 'no' would be a correct answer, not because neither would be correct, but because each would have a false implicature. With a more reasonable judge presiding, one who required 'yes' or 'no' answers of witnesses, but who also permitted qualifications or elaborations, you might find that a 'no' answer would come much more easily since you would be permitted to *cancel* the implicature of a simple 'yes' or 'no': "Well, no he's not *bald*, but he does have a rather large bald spot in back."

I do not claim that every hedging response is a result of 'yes' or 'no' answers having false implicatures. But this does seem to me to be the most common cause of hedging responses that occur in the course of normal conversations. I should confess, though, that I must be far more opinionated than most philosophers who write about vagueness, because I find borderline cases very hard to come by. Nearly always I feel, when presented with a putative borderline case of a vague predicate, that if really pressed to give a simple 'yes' or 'no' answer, I would not have too much trouble deciding which one was correct. If pressed, I would say that the cover of *The Logical Easis of Metaphysics* is red.

If presented with a sorites series of men varying in heights by small increments from the short to the tall, and if asked, in order, of each man whether he is short, I can imagine (I can only imagine since of course I have never been presented with such a series) that when my 'yes' answers finally give out, they would not be immediately followed by 'no' answers.

Instead, at some point, I would just feel it impossible to say either way. My feeling of discomfort at having to decide the question would not be prompted by the fear of false implicature—I can imagine that my audience is very good at ignoring implicatures. What then is the cause of it? We have really already provided an answer to this question. My discomfort is the result of being in an environment in which no standard of use of 'short' could satisfy every constraint.

In the situation in which two people disagree about whether a vague predicate applies to a given object, and it seems to a third party, adjudicating "from above," that there is no fact of the matter about who is right, we have yet another explanation of the cause of the hedging. It may be that the participants to the dispute are each correct, because there may be license in the choice of standards to be adopted. Each has his own standard, and is correct in applying that standard. The onlooker is prompted to hedge, since there is no way of expressing this without going meta-linguistic.

Chapter 4

Phenomenal Continua and the Sorites

I would like to conclude with a discussion of a version of the paradox to which I would not wish to extend my solution. Imagine we have thirty color patches, each of which looks homogeneously colored, arranged in a row such that each patch looks the same as its neighbor(s), and yet the first patch looks red and the last patch does not look red. (Suppose it looks orange.) By hypothesis, then, the first premise of the argument below is true, the conclusion is false, and the antecedent of each instance of the second premise is true whenever the two patches are adjacent. Yet the reasoning which takes us from the premises to the conclusion is impeccable.

> Patch #1 looks red
> If two patches look the same in color, then if one looks red so does the other
> Patch #1 looks the same as patch #2
> Patch #2 looks the same as patch #3
>   ⋮
> Patch #29 looks the same as patch #30
> ———————————————————————
> Patch #30 looks red

The paradoxical nature of the argument need not turn on the vagueness of 'red'.

The argument might just as easily have been run with 'looks square' instead. One thing disinctive about this argument is its use of the observational copula 'looks'. But second, and more importantly, the conditional premise of the argument is not really a *tolerance* principle at all. It does not say that the applicability of 'looks red' will tolerate *small* changes in apparent color; but rather that the applicability of 'looks red' will not vary when there is *no* change in apparent color.

This second version of the paradox is in a way more paradoxical than the first, because its second premise—its sorites premise—is even harder to give up. If one patch looks red, and another doesn't, then they do not look the same. End of story. The tolerance principle for 'heap' may seem true, but the sorites premise for 'looks red' seems like a truism. Someone who sincerely claimed that two patches looked the same and yet that one looked red and the other not, would not merely seem to be plainly mistaken, but also to be in a state of confusion.

There are, though, different senses of 'looks', and it may be that the extent to which the sorites premise for 'looks red' sounds like a truism varies accordingly. The sense of 'looks' I have in mind here is that generally used for making observation reports. It is the sense according to which I may truly say of a friend gently illuminated by a neon green sign, in a dimly lit bar, that she looks green, whether or not I believe the ill-appearance of her complexion to be due only to the sign. It is the sense according to which a thing may change in look quite rapidly, while undergoing nothing we would naturally think of as a change in it, as when I say, placing a paint sample against different backgrounds, "Now it looks red ... Now it looks pink." It is the sense of 'looks' which may be explicitly relativized to an observer, and according to which when one person says, "It looks very dark blue" and another says, of the same thing, "It looks black to me," they do not contradict each other.

Following Frank Jackson, I will call this the "phenomenal" sense of looks.[1] Note

---

[1] See Jackson (1977, ch. 2). I think we have the same sense in mind, but that does not mean that I am co..mitted to agreeing with everything he says about it. For example, Jackson (pp. 34, 36) considers the case of someone who is red color-blind—and describes the person as someone to whom red things look a particular shade (or range of shades) of grey. (Whether anyone has ever had such a condition, or whether such a condition is even physically possible should have no bearing on the point.) According to Jackson,

that the sorites argument for 'looks red' (p. 87, above) is only paradoxical if just a single observer is involved. There is no problem if patch #1 looks red to Amy, neighboring patches look the same to Sue, and patch #30 looks orange to Tim, since one cannot infer anything about the way a patch looks to Tim from the way other patches (or even the same patch) look to Amy and Sue. If different observers were involved in the above argument, we would simply reject the reasoning as invalid. This could be made explicit by using subscripts. So, for example, 'Patch #2 looks$_{Amy}$ red' cannot be inferred from 'Patch #1 looks$_{Amy}$ red' and 'Patch #1 looks$_{Sue}$ the same as patch #2'.[2]

But if we suppose that there is just one observer to whom patch #1 looks red, neighboring patches look the same, and patch #30 looks orange (and hence not the same as patch #1), then we really do have a problem. The problem is that it follows from this supposition that 'looks the same as' is non-transitive. If an object which looks red and an object which looks orange can be "connected" by a chain of objects, each of which looks the same as its immediate neighbors, then it had better not be the case that whenever an object a bears the 'looks the same as' relation to an object b, then if a looks red then b does, since the conditional *is* transitive. Rather than reject the truism, I will, in what follows, argue against the non-transitivity supposition. I will argue, that is, that objects which look different (in our case a red patch and an orange patch) cannot be connected by a looks-the-same-as chain.

Reconsider the supposition which entailed non-transitivity and thus led to paradox,

---

nothing looks red, in the phenomenal sense, to such a person. But it seems that such a person could truly say to a friend in a dimly lit bar, under the glow of a red neon sign, that she looks red. Although he could also truly say, in just the same situation, perhaps if describing his disability, that she looks grey. It might be tempting to conclude that there are *two* different senses of 'looks' that are used for making observation reports. But though this may be so, it is the wrong conclusion to draw in this case. The right thing to say, it seems to me, is that only a single sense of 'looks' is involved here, and that whether it is true to say that things look red or only that they look grey, in this one sense, to the color-blind person is something that varies with the context of utterance, in particular with whether or not his disability is relevant to the discussion. Furthermore, I think there is room, within the view Jackson presents in *Perception* for accepting such a position. At any rate, the issue will not be one that is pertinent to the present discussion.

[2]I assume that when 'looks' is used in the phenomenal sense, it must be relativized to an observer, implicitly if not explicitly, which is to say that every utterance of a sentence 'a looks F' has the same truth-conditions as some explicitly relativized sentence ' a looks F to b'.

namely, that the following hold:

1. Patch #1 looks red
2. Patch #n looks the same as patch #$n + 1$ (for each positive $n < 30$)
3. Patch #30 does not look red, it looks orange

Remember, we are assuming, so as not to make the sorites argument for 'looks red' just trivially fallacious, that only a single observer, call her 'Olivia', is involved. One scenario which would support 1-3 is the following: Olivia is shown patch #1, and it looks red to her; then she is shown patch #30 and it looks orange; then she is shown each of the twenty-nine pairs of adjacent patches in succession (not necessarily in order), and each pair looks the same. But if this is the scenario which is taken to support 1-3, then it *cannot* be concluded that 'looks the same as' is not transitive. Concluding, from the scenario, that 'looks that same as' is not transitive, would be like concluding that 'is taller than' is not transitive from the following: Michael is taller than Graham (in 1990); Graham is taller than Dan (in 1985); and Dan is taller than Michael (in 1995). Just as we have no reason to presume that Dan's height when compared to Graham's in 1985 is the same as his height when compared to Michael's in 1995, we have no reason to presume, if Olivia is shown the pairs in succession, that patch #15, say, looks the same when it is presented with patch #14 as it does when presented with patch #16. In particular, we have no reason to suppose that if patch #15 looks red when it is presented with patch #14, then it also looks red when presented with patch #16—there is no license to carry over the "middle term."

This in essence is Jackson's response to the claim that 'looks the same as' is not transitive (1977, p. 113).[3] But, as Jackson goes on quite rightly to point out, if there is a *single* time at which 1-3 above hold, say as Olivia surveys the entire series of patches at once, then we must after all conclude that 'looks the same as' is not transitive.[4] To this Jackson replies that such a situation is *logically impossible*. He has been, however, virtually alone in his opinion.

---

[3]Although he puts the point in terms of sense data, and to a different end from mine here

[4]Just as if there were a single time at which Michael were taller than Graham, Graham taller than Dan, and Michael not taller than Dan, we would have to conclude that 'is taller than' is not transitive.

Indeed, Crispin Wright (1975, pp. 345-47) has purported to *prove* that there could not be apparently continuous change—phenomenal continua—if 'looks the same as' were transitive. But even without a proof, it is fairly easy to convince people that 'looks the same as' is not transitive. Just tell this story: "Suppose you have a color spectrum, ranging from red on one end, and looking to change continuously to orange on the other. Now you could cut the spectrum up into small enough strips, so tha' each strip looked homogeneous in color. Each strip will then look the same as its immediate neighbor(s), but the end strips will not look the same. Whether or not the strips are viewed in succession, or all at once, has no effect on the situation, so 'looks the same as' must be non-transitive. In fact, the strips don't really need to be thought of as cut-outs, but just as regions of the original spectrum." Convincing as this story may be, I shall provide reasons for siding with Jackson against Wright.

Let us examine the above story. It consists of three claims amounting to the following, the last of which is supposed to be the conclusion of an argument:

1. There are (or could be) phenomenal continua—changes in color across a spectrum, for example, which look like *continuous* changes.

2. Small enough regions of phenomenal continua look homogeneous.

3. 'Looks the same as' is not transitive with respect to regions of phenomenal continua: adjacent regions look the same, but the end regions do not look the same.

The argument for the conclusion can be construed in two ways. One construal is this: what is required for a change in color to look continuous is that narrow enough regions of the spectrum look homogeneous in color; and that if narrow enough regions look homogeneous in color, then adjacent regions must look the same. The thought here is that (1) may be assumed; it entails (2); and together (1) and (2) entail (3). I will call this *the simple thought*. On the other hand, the argument could be construed in the following way: *given* that when change in color on a spectrum looks continuous, narrow enough regions of the spectrum look homogeneous in color, it must be that narrow enough adjacent regions on the spectrum look the same in color, even though the end regions do not look the same.

The idea here is that (1) and (2) may each be assumed, and together they entail (3) I will call this *the speculative thought.*

The simple thought is mistaken—it reflects a misunderstanding of the nature of continua. To see this, it is perhaps easier to consider the analogous case of continuous motion. Say we have an object moving in one direction along a straight line.[5] In order for the motion to *be continuous*, it is not required that for adjacent intervals of time I and J, even very small ones, the object be in the same place during I that it is in during J. Indeed, it is not required that the object remain in a single place during any time-span longer than an instant.[6] And if this is not required to be the case in order for the object to be moving continuously, why should it be required to *look* the case in order for the object to *look* to be moving continuously?

What is required for continuous motion is that for any positive distance ε, no matter how small, there is a positive amount of time δ which is small enough so that during any time-span of length δ, the object moves less than ε in that time-span.[7] What is required for motion to *look* continuous, then, is presumably that for any distance ε, no matter how small, there is a positive amount of time δ which is small enough so that during any time-span of length δ, the object *looks* to move less than ε in that time-span—in other words, given any distance you like, no matter how small, there's some short enough time-span in which the object always looks to cover less than that distance. Analogously, for change in color to look continuous, it is required only that given any positive amount of change in color, there is a narrow enough width such that in any region of that width on the spectrum, the color looks to change less than that amount in that region. Crucially, it is not required that any region, however narrow, look either homogeneous in color, or the same as its immediate

---

[5] Just as we may think of the motion of the object as a function which, given a time as argument, yields the location of the object at that time as value, we may think of change in color along the spectrum as a function which, given a point on the spectrum as argument, yields the color of the spectrum at that point as value. The continuity of either motion or change in color can then be identified with what it is for these respective functions to be continuous.

[6] And in order for the intervals to be *adjacent*, it is required that at least one of them be longer than an instant.

[7] We are here assuming that the object is in *some* location or other at each moment—that it never ceases to exist.

neighbors.

The speculative thought, in contrast, is not so obviously mistaken. If, as a matter of fact of human perception, small enough changes in color, or in location, are perceived by us as being no change at all, then it does seem right to conclude that adjacent regions of the spectrum, when they are small enough, must look the same in order for change in color to look continuous; and that in adjacent intervals of time, when they are small enough, a moving object must look the same in respect of position.

It is tempting to argue as follows: If on the color spectrum, as a matter of fact of human perception, small enough regions of the spectrum look homogeneous, then adjacent such regions must look the same or we would see an abrupt change in color somewhere. Anyone who is tempted by this argument, though, betrays a proclivity for transitivity. The defender of non-transitivity must deny that all adjacent, homogeneous-looking patches look the same. Here's why: if for some narrow width w, all regions narrower than w look homogeneous, then there is a greatest such width—call it w_{max}. Now consider some width w' that is less than w_{max}, but greater than $w_{max}/2$. By hypothesis, all regions of width w' look homogeneous. But if each pair of adjacent regions of width w' looked the same as each other, then all regions of width $2w'$ would look homogeneous. This contradicts our original assumption since $2w'$ is greater than w_{max}. So even if, as a matter of fact of human perception, small enough regions of the spectrum look homogeneous, it cannot be that all adjacent such regions look the same.

A stronger argument for the validity of the speculative thought goes like this: given that for some narrow width, all regions less than that width look homogeneous, then if we *divide* any such region into *two*, the two will look the same. For if they do not look the same then the original region cannot look homogeneous. Given also that distant enough regions do not look the same, non-transitivity clearly follows.

I accept this stronger argument for the validity of the speculative thought, and instead wish to question its soundness. I will begin by highlighting a certain absurdity which comes with accepting its conclusion that 'looks the same as' is not transitive. I will

then proceed by showing that the grounds typically offered, or assumed, in support of its premises are not conclusive ones.

We already have reason to be wary of the claim that 'looks the same as' is not transitive, namely, that if we accept it, we cannot consistently accept what seems to be a truism: If two color patches look the same, then if one looks red so does the other. But even if we set aside this unwelcome consequence, a little reflection will reveal the claim to be implausible. I consider first the case of 'looks the same as' in respect of position. If this relation were genuinely non-transitive, then it should be possible for us to place three stationary objects in such a way that the first looked in the same position as the second, the second in the same position as the third, with the first and the third not looking to be in the same positions. But we cannot, or so I will presently argue.

In *Wang's Paradox*, Michael Dummett discusses a very coarse observer. This is "someone with a vision so coarse that it can directly discriminate only four distinct positions in the visual field ... : that is, it is not possible to arrange more than four objects, big enough for this person to see, so that he can distinguish between their position" (Dummett 1975, p. 267). Adding some detail to the example, we will suppose that this coarse observer cannot distinguish hand-positions on a clock that are separated by an angle of less than 90°, but can distinguish positions that are separated by an angle of greater than 90°. By hypothesis, then 'looks in the same position as' is non-transitive for this observer, since a hand at twelve o'clock will look in the same position as a hand at two o'clock, which will in turn look in the same position as a hand at four o'clock. But the outer hands will not look in the same position.[8]

---

[8]Dummett introduces the example in order to argue that, although the identity criterion for phenomenal qualities proposed by Nelson Goodman is perfectly well defined (see Goodman (1977, ch. IX, §2)), it is not an identity criterion for anything we would normally think of as a phenomenal quality. The criterion as applied to the present case will yield the intuitively wrong result that even for the very coarse observer, there are continuum-many phenomenal positions. Given Goodman's criterion, two hands on the clock, however close, will be phenomenally distinct for the very coarse observer since a third hand placed less than 90° from one but more than 90° from the other will look in the same position as the one but not as the other.

I claim that no such coarse observer is possible. According to those who maintain that for us, 'looks in the same position as' is non-transitive, the only difference between us and the very coarse observer is but a difference of degree. For the coarse observer, the minimally discriminable angle between clock hands is 90°, while for us it is supposed to be some much lesser angle. So if the coarse observer cannot possibly exist, then since we do exist, we are not coarse observers.

Suppose we place three hands on the clock—one at twelve o'clock, one at two o'clock, and one at four o'clock—and ask the very coarse observer to tell us what he sees. By hypothesis, the middle hand looks to him to be in the same position as both of the outer hands, which in turn look to be in different positions. Suppose also that the hands are each the same size, and completely opaque. Then two hands will look in the same position just in case one looks to completely obstruct the other. Since the two outer hands look in different positions, neither looks to completely obstruct the other. Will the observer see the middle hand, or not? Well, if he does see it, then it must look to completely obstruct both of the other hands, since by hypothesis it looks to be in the same position as both. But if the middle hand looks to obstruct both of the other hands, then he will not be able to see *them*, and hence they cannot look to be in different positions, contrary to the supposition. If the coarse observer does *not* see the middle hand, it is not because he has blindspots— were we to place a hand at two o'clock, and one at five o'clock, he would see both since by hypothesis they would look to be in different positions. If he does not see the third hand, then it is because it looks completely obstructed by at least one of the other two. If the observer knows there is a third hand, then he may say that it looks in the same position as *one* of the other two. But he cannot say which, because it is obstructed. What he cannot say is that it looks to be completely obstructed by *both* of the other two hands, since by hypothesis these look to him to be in different positions.

Suppose now that the hands are not opaque but translucent, so that all three may be seen even if one is obstructed. Let the hand at twelve o'clock be translucent red, the one at two o'clock be translucent yellow, and the one at four o'clock be translucent blue.

Now when one hand obstructs another, both will still be visible. When the red and yellow hands look in the same position, together they will look like a single orange hand.[9] Now what will the coarse observer see? Well, the red hand looks to him in one position, call it X, and the yellow hand looks also to be in X, so at X he will see a single orange hand. The blue hand looks in a different position from the red one, call it Y, but the yellow hand also looks in the same position as the blue, so it looks to be in Y, so at Y he will see a single green hand. Thus, if we ask him what position the yellow hand looks to be in, he will say it looks to be in two places at once! But if it is only in virtue looking to be in *two* positions that the yellow hand can look to be in the same position as the other two, then all bets are off—since if the red hand (or the blue hand) does not look to be in two positions at once, then it does not look the *same* in respect of position as the yellow hand.

It cannot be objected, against my considerations, that we have no reason to presume to be able to *understand* the very coarse observer—no reason to presume to be able to make *sense* of what he says he sees. If the proponent of non-transitivity were right, we should have just the same experiences as the very coarse observer (only in cases where the hands were much closer together), and hence be able to make perfect sense of what he says he sees. Given the foregoing considerations, we must reject the possibility of a very coarse observer. We are to reject, that is, the possibility that any being (us included), can have a visual experience as of three objects, one of which looks in the same position as both of the other two, when the other two do not look in the same position.

I now propose that the non-transitivity of 'looks the same color as' stands or falls with the non-transitivity of 'looks in the same position as'. If the latter is implausible, so is the former. Suppose we have a series of color patches arranged horizontally, each patch set in a vertical track which allows us to slide it individually up and down. Now certainly it should always be possible to carry out the following instruction: Place the patches along their tracks so that two patches look in the same vertical position if and only if they look

---

[9]This can be made more plausible by supposing that the clock is a two-dimensional one, say on a computer screen.

the same in respect of color. If 'looks the same as' in respect of position is not transitive, then we cannot place three patches so that the first looks in the same vertical position as the second, the second in the same position as the third, but the first and the third not in the same position. But then if we have carried out our instruction, there cannot be three patches for which 'looks the same as' in respect of color is not transitive either.

The question to which we now turn is: must we accept the premises on which the speculative thought rests? Must we, despite the implausibility of the non-transitivity of 'looks the same as', accept the two claims which entail it? Let us address the second premise (p. 91, above) first. Is it true, in the case of apparently continuous change in color, that narrow enough regions of the spectrum will always look homogeneous? And is it true, in the case of apparently continuous change in position, that in small enough intervals of time, a moving object will look to remain in a single place? Wright, well aware that this assumption is required in the argument for the non-transitivity of 'looks the same as', describes it as a "very natural presupposition." To suppose otherwise, he writes, would be to suppose that "we have infinite powers of discrimination ..., that we can always directly discern some distinction more minute than any discerned so far" (Wright 1975, p. 346).

In order to address the present question, we need to examine what justification there might be for taking as given that small enough changes in color or position are perceived by us as no changes at all. In the case of color, the justification seems to be that we can actually cut up the spectrum into regions which look homogeneous; or that we can cover up all but an inch of the spectrum, and be left with a region that looks homogeneous; or even that we can leave the entire spectrum exposed and intact, and still judge of narrow regions of it, looking from one to another, that they each look homogeneous. Unfortunately, this justification is not very good—it provides us only with a 'can' from which is concluded a 'must'. It is certainly true that regions of the spectrum may sometimes look homogeneous, especially when attention is focused on them in some way. But that is no reason to suppose that there is any width such that regions of the spectrum less than that width always looks

homogeneous. If you take a foot-long spectrum ranging from red on the left to orange on the right, place it about two feet ahead of you, and stare at the red edge of it for a few seconds, soon nearly the entire spectrum will look homogeneously red. But just because a two-thirds region of the spectrum *can* look homogeneously red, does not mean that it *must*, as one can verify by focusing one's attention somewhere nearer the middle.

In the case of position, the justification for taking as given that small enough changes are perceived by us as no changes at all seems to be that sometimes objects move so slowly that they it seems they look still—take the moon, or the hour-hand on a clock. The thought is that if a moving object seems to look still during an interval, then it must be because we cannot visually distinguish any of the positions it is in during that interval—the *reason* it seems to look still is that our ability to discriminate does not extend to positions that are too close together.

Though this may be a plausible explanation of why when we look at the hour-hand on a clock, it seems to look still, it is not the only explanation. Another explanation, consistent with transitivity, is that when we look at the hour-hand on a clock, although it does in fact look to change in position, the change in appearance is too slight, and too slow, for us to *notice* it. It should not be objected that the distinction introduced here is an *ad hoc* one, tailored to the present purpose, since it is a distinction naturally invoked in other cases. Suppose a friend whom I see daily comes up to me and says, expectantly, "Do I look different?" I may be unable to discover any change in her appearance. Still, once she goes on to tell me that she has lightened her hair (say), I may be able to tell that, yes, her hair color does look different than it did the day before. Yet it seems wrong to say that her hair color looked different to me only *after* I was informed of the change. Instead, it looked different all along, but only after being informed did I notice it. My inability to discover the change, prior to being informed, was not a visual failing but a cognitive one.

So at present we have two competing explanations of what is going on when the hour-hand of a clock seems to look still. The first explanation is that when the hour-hand looks still, say for a period of a minute, it looks in the same position at the end of the

minute as at the start. The alternative explanation is that when the hour-hand looks still, although it does not look in the same position at the end of the minute as at the start, we do not notice this. *Noticing* the change in apparent position requires not only that there be an apparent change, but also that we believe there to be one. A reason for deciding in favor of the latter explanation is this: If we look at the hour-hand for enough time, say five minutes, we will realize that it looks to be in a different position from the one in which it started. But the suddenness with which we realize this is much more plausibly accounted for as the sudden formation of a belief, rather than a sudden change in look.

I have now given reasons for suspecting that we do not have conclusive grounds for premise (2) of the speculative thought. On the one hand this should be welcome, since together with (1), (2) entails a claim we should regard as implausible. But on the other hand, don't we *have to* suppose (2) to be true? As Wright said, to suppose otherwise would be to suppose that we have infinite powers of discrimination.

Assuming there are phenomenal continua, Wright is correct on this score. For let us suppose that we do have only finite powers of discrimination—that, for example, as an object moves from the left of our visual field to the right, it passes through only finitely many visibly distinct positions. Does it follow, as Wright claims it follows, that for some small distance, whenever the object traverses that distance, it looks as if it does not change position at all? Suppose the visibly distinct positions in our visual field number only four— call them A, B, C, and D. Now by hypothesis, whenever an object moves *within* any of these regions it looks still. It might nonetheless be that when an object moves from one region to another, no matter how small the distance it traverses, it does look to change position. But then the motion could not look *continuous*. It would instead look like there were discrete changes in the position of the object—like a cursor on a computer screen, which looks to "jump" from one character position to the next. So if we do have finite powers of discrimination, premise (2) of the speculative thought is true.

But can we really be sure that motion does in fact ever look continuous to us? Can we be sure, that is, that premise (1) of the speculative thought is true? After all, if there

really are only finitely many visibly distinct positions in our visual field, there are very many more than four of them. Motion certainly *strikes* us as continuous-looking. But can we be sure, when an object passes before us, that it does not really look as if it is moving *discontinuously*, looking to take very tiny discrete jumps? And can we be sure, when looking at a color spectrum, that it does not really look to change discontinuously from red on the left to orange on the right? Can we be sure that the spectrum does not really look, say, like one million very narrow, homogeneously-colored vertical stripes, each looking not the same as, but just slightly different from its immediate neighbors? I do not see that we can be at all sure of these things.

Imagine we could have both of the following two visual experiences: the first, of a cursor on a computer screen looking to move discontinuously from one character position to the next; the second, of a cursor on a computer screen looking to move continuously, but jerkily, from one character position to the next—looking to move very quickly *between* character positions, and looking to pause momentarily in each character position. Clearly, the motion of the cursor cannot look the same in the two cases, since looking to move discontinuously is incompatible with looking to move continuously. But as I watch the cursor on my computer screen moving before me now, I cannot be sure that I am having one of these experiences and not the other. Although continuous-but-very-jerky-looking motion is different from discontinuous-and-very-jerky-looking motion, it would seem that they would not really *strike* us as being different. Both would strike us as discontinuous-looking.

Similarly, imagine we could have the following two experiences: the first, of a cursor on a computer screen looking to move discontinuously from one *pixel* to the next (suppose also that pixels are incredibly small); the second of the cursor looking to move continuously, but jerkily from one pixel to the next. Again, the two experiences, although different, would not strike us as being different. Both would strike us as experiences of continuous-looking motion.

What position should we now take with respect to the argument I called *the speculative thought*? Its conclusion, that 'looks the same as' is not transitive, I argued to be implausible: on the one hand because were it true, we could not accept that if two things look the same, then if one looks red (say) so does the other; on the other hand, because to accept it would be to accept that there could be very coarse observers of the kind Dummett discusses, and this it seems there cannot be. We should therefore reject at least one of the premises on which the speculative thought rests. I have argued that the grounds typically offered for each are not conclusive. Nevertheless, if we do not have infinite powers of discrimination, then we must, as Wright says, accept premise (2), in which case (1) should be rejected. This seems like a sensible position to take anyway: if we really have only finite powers of discrimination, how *could* there be phenomenal continua? If, in contrast, our powers of discrimination are infinite, then we are free to reject premise (2), since what I called *the simple thought* is wrong; (2) is not entailed by (1).

Let us now return to the version of the paradox considered at the outset of this section.

> Patch #1 looks red
> If two patches look the same, then if one looks red so does the other
> Patch #1 looks the same as patch #2
> Patch #2 looks the same as patch #3
> $\vdots$
> Patch #29 looks the same as patch #30
> _____
> Patch #30 looks red

I submit that the argument is not after all paradoxical. It only seems paradoxical when it is assumed that there can be a *sorites series* of the required sort for 'looks red', a series the first and last members of which look different, while adjacent members look the same. Although the existence of such a series is not, as Jackson thought, *logically impossible*, we do not have sufficient reason to believe that such a series could exist, and hence may with peace of mind accept the claim that if two patches look the same, then if one looks

red so does the other.

Bibliography

Bolc, L. & Borowik, P. (1992), *Many-Valued Logics I: Theoretical Foundations*, Springer-Verlag, Berlin.

Boolos, G. (1991), 'Zooming down the slippery slope', *NOÛS* **25**: 695–706.

Burns, L. (1991), *Vagueness: An Investigation into Natural Languages and the Sorites Paradox*, Kluwer, Dordrecht.

Cambell, R. (1974), 'The Sorites paradox', *Philosophical Studies* **26**: 175–191.

Cargile, J. (1969), 'The Sorites paradox', *British Journal for the Philosophy of Science* **26**: 175–91.

Dummett, M. (1975), 'Wang's paradox', *Synthese* **30**: 301–24. Page references are to reprint in Dummett (1978).

Dummett, M. (1978), *Truth and Other Enigmas*, Harvard University Press, Cambridge, MA.

Dummett, M. (1991), *The Logical Basis of Metaphysics*, Harvard University Press, Cambridege, MA.

Edgington, D. (1992), 'Validity, uncertainty and vagueness', *Analysis* **52**(4): 193–204.

Fine, K. (1975), 'Vagueness, truth and logic', *Synthese* pp. 265–300.

Goguen, J. A. (1967), 'The logic of inexact concepts', *Synthese* **19**: 325–373.

Goodman, N. (1977), *The Structure of Appearance*, third edition, D. Reidel, Dordrecht.

Grice, P. (1989), *Studies in the Ways of Words*, Harvard University Press, Cambridge, MA.

Horgan, T., ed. (1995), *Spindel Conference 1994: Vagueness*, Vol. 33, supplement, *The Southern Journal of Philosophy*.

Jackson, F. (1977), *Perception: A Representative Theory*, Cambridge University Press, Cambridge, England.

Jackson, F., ed. (1991), *Conditionals*, Oxford University Press, Oxford.

James, E., Slater, J. et al., eds. (1983–), *The Collected Papers of Bertrand Russell*, Allen & Unwin/Unwin Hyman, London.

Kamp, J. A. W. (1975), 'Two theories about adjectives', in E. L. Keenan, ed., *Formal Semantics of Natural Language*, Cambridge University Press, Cambridge, England.

Kamp, J. A. W. (1981), 'The paradox of the heap', in U. Mönnich, ed., *Aspects of Philosophical Logic*, D. Reidel, Dordrecht.

Kleene, S. C. (1952), *Introduction to Metamathematics*, Van Nostrand, New York.

Kripke, S. (1972), 'Naming and necessity', in D. Davidson & G. Harman, eds., *Semantics of Natural Language*, Reidel, Dordrecht, pp. 253–355. Reprinted as Kripke (1980).

Kripke, S. (1975), 'Outline of a theory of truth', *Journal of Philosophy* **72**: 690–716.

Kripke, S. (1980), *Naming and Necessity*, Harvard University Press, Cambridge, MA.

Lakoff, G. (1973), 'Hedges: A study in meaning criteria and the logic of fuzzy concepts', *Journal of Philosophical Logic* **2**: 458–508.

Lewis, D. (1970), 'General semantics', *Synthese* **22**: 18–67. Page references are to reprint in Lewis (1983).

Lewis, D. (1979), 'Scorekeeping in a language game', *Journal of Philosophical Logic* **8**: 339–59. Reprinted in Lewis (1983).

Lewis, D. (1983), *Philosophical Papers*, Vol. 1, Oxford University Press, Oxford.

Lukasiewicz, J. & Tarski, A. (1930), 'Untersuchungen über den aussagenkalkül', *Comptes Rendus des séances de la Société des Sciences et des Lettres de Varsovie* **23**(iii): 30–50. Reprinted in Tarski (1983). Translated by J. H. Woodger.

Machina, K. F. (1976), 'Truth, belief, and vagueness', *Journal of Philosophical Logic* **5**: 47–78.

McGee, V. (1985), 'A counterexample to Modus Ponens', *Journal of Philosophy* **82**(9): 462–471.

McGee, V. & McLaughlin, B. (1995), 'Distinctions without a difference', in T. Horgan, ed. (1995), pp. 203–251.

Partee, B. (1989), 'Binding implicit variables in quantified contexts', *CLS* **25**.

Rolf, B. (1984), 'Sorites', *Synthese* **58**: 219–50.

Russell, B. (1923), 'Vagueness', *Australasian Journal of Philosophy and Psychology* 1: 84–92. Page references are to reprint in E. James, J. Slater et al., eds. (1983–).

Sainsbury, R. M. (1988), *Paradoxes*, first edition, Cambridge University Press.

Sorensen, R. (1988), *Blindspots*, Clarendon Press, Oxford.

Stalnaker, R. (1973), 'Presuppositions', *Journal of Philosophical Logic* 2: 447–57.

Stalnaker, R. (1974), 'Pragmatic presuppositions', in M. Munitz & P. Unger, eds., *Semantics of Natural Language*, New York University Press, New York.

Stalnaker, R. (1975), 'Indicative conditionals', *Philosophia* 5: 269–86. Reprinted in F. Jackson, ed. (1991).

Stalnaker, R. (1987), *Inquiry*, MIT Press, Cambridge, MA.

Tappenden, J. (1993), 'The Liar and Sorites paradoxes: Toward a unified treatment', *Journal of Philosophy* 90(11): 551–77.

Tarski, A. (1983), *Logic, Semantics, Metamathematics*, second edition, Hackett. Translated by J. H. Woodger.

Unger, P. (1979), 'There are no ordinary things', *Synthese* 41: 117–54.

Williamson, T. (1994), *Vagueness*, Routledge, London.

Williamson, T. (1996), 'Cognitive homelessness', *Journal of Philosophy* pp. 554–573.

Wright, C. (1975), 'On the coherence of vague predicates', *Synthese* 30: 325–365.

Wright, C. (1987), 'Further reflections on the Sorites paradox', *Philosophical Topics* 15: 227–290.

Wright, C. (1992), *Truth and Objectivity*, Harvard University Press, Cambridge, MA.