True, Truer, Truest

Brian Weatherson

Brown University

November 2003

What the world needs now is another theory of vagueness. Not because the old theories are useless. Quite the contrary, the old theories provide many of the materials we need to construct the truest theory of vagueness ever seen. The theory shall be similar in motivation to supervaluationism, but more akin to many-valued theories in conceptualisation. What I take from the many-valued theories is the idea that some sentences can be *truer* than others. But I say very different things about the ordering over sentences this relation generates. I say it is not a linear ordering, so it cannot be represented by the real numbers. I also argue that since there is higher-order vagueness, any mapping between sentences and mathematical objects is bound to be inappropriate. This is no cause for regret; we can say all we want to say by using the comparative *truer than* without thinking of it as deriving from such a mapping. From supervaluationism I take the idea that we can keep classical logic without keeping the familiar bivalent semantics for classical logic. But my preservation of classical logic is more comprehensive than is normally permitted by supervaluationism, for I preserve the acceptability of classical inference rules as well as the validity of all classical sequents. And I do this without relying on the concept of acceptable precisifications as an unexplained explainer.

The world does not need another guide to varieties of theories of vagueness, especially since Timothy Williamson (1994) and Rosanna Keefe (2000) have already provided quite good guides. So I shall not go over existing theories in detail, though I need to say a little about their virtues and vices so as to highlight just how good my theory is.

1.   Many Valued Theories

The most familiar many-valued theory, (call it M), says that the traditional truth tables leave a few values out. There are continuum many truth values, and indeed they are structured like the continuum, so we can felicitously represent them by the interval $[0, 1]$. The four main logical connectives: *and*, *or*, *if* and *not* are truth-functional with respect to these expanded truth tables. The functions are as follows:

If V is the function from sentences to $[0, 1]$, $min(x, y)$ is the smaller of x and y (or x, if they are equal) and $max(x, y)$ is the larger of x and y (or x, if they are equal), then:

$$V(A \land B) = min(V(A), V(B))$$
$$V(A \lor B) = max(V(A), V(B))$$
$$V(A \rightarrow B) = min(1, 1 - V(A) + V(B))$$
$$V(\neg A) = 1 - V(A)$$

Adopting these rules for the connectives commits us to adopting the logic L_C. This logic is a little tricky to axiomatise; the following is one familiar axiomatisation (from Priest 2001).

$(A \rightarrow B) \rightarrow ((B \rightarrow C) \rightarrow (A \rightarrow C))$

$A \rightarrow (B \rightarrow A)$

$(A \rightarrow \neg B) \rightarrow (B \rightarrow \neg A)$

$((A \rightarrow B) \rightarrow B) \rightarrow ((B \rightarrow A) \rightarrow A))$

$((A \rightarrow B) \rightarrow B) \leftrightarrow (A \vee B)$

$(A \wedge B) \leftrightarrow \neg(\neg A \vee \neg B)$

Roughly then, *M* is the theory that the logic with that axiomatisation and semantic model is *the* logic of natural language. 'Roughly' because it turns out that there are some slightly perverse ways of interpreting those axioms and that semantic model in ways that alter the *philosophical* character of the resultant theory. But for now we will assume the axioms and semantics should be interpreted in the most natural way. This done, *M* has a couple of nice features. First, we get an account of what it is for there to be borderline cases of *F*-ness. *a* is a borderline *F* just in case the sentence *a is F* takes a truth value between 0 and 1 *exclusive.* Secondly, we get a nice account of the Sorites paradox. Any premise in a Sorites argument, say *If a man of height x nm is tall, then a man of height x-1 nm is tall*, will be *almost* true. Make the steps in the Sorites small enough, and the truth of each of the premises will be above 1-ε for arbitrarily small ε. We err in our evaluation of the Sorites argument because we mistake sentences that are almost true for sentences that are, simply, true. Or perhaps we err because we think that rules of inference that preserve truth, like *modus ponens*, preserve almost-truth. In either case, it is an understandable error, so we have a theory that makes it understandable why we might have gone wrong in evaluating Sorites arguments.

Those are the advantages of *M*, and I think they are substantial advantages, at least *prima facie*. The real problem with *M* is not that it has no nice features, but that it has so many negative features. I will focus on three.

1.1.   False Precision

Consider the following sentences, all of which I think are vague. (You don't need to know much about the cases except that they are all canonical borderline cases of the predicates in question.)

(1)   I am tall.

(2)   Edgar Martinez is a great baseball player.

(3)   Rosanna Keefe's *Theories of Vagueness* is a short book.

According to *M*, each of these is true to some degree or other. But which degree could it possibly be? What would make it that case that the truth value of (1) is 0.6 rather than 0.8, or that the truth value of (2) is 0.8 rather than 0.6? The intuition behind *M* is that there's no sharp border between the tall and the not-tall philosophers, or between the great and the less than great baseball players. But according to *M* there is a precise fact about how far from truth sentences like (1), (2) and (3) are.

This is not a new objection. It is, for instance, at the heart of the central objections Keefe (2000: Ch. 4) makes to *M*. Responses to Keefe have tended to stress that *M* is just meant to be a model for vague languages, that it is not meant to tell an exact story. (See, for example, Cook 2002.) These responses seem doubly misguided to me. Without a compelling argument for why we can't tell the exact story, for why we must suffer to be content with useful fictions like *M*, it seems this response shows at best that *M* is a useful waystation on the road to the truth. (Which, not coincidentally, is exactly what it is.)

And even in the fiction there is something quite misplaced. Grant that the particular allocation of numbers, say 0.6 to (1), 0.8 to (2), 0.3 to (3), is somewhat arbitrary. Still, it is either true in the model that (1) is closer to the truth than (2), or that (2) is closer to the truth than (1), or they are equally close to the truth. The last claim, that they are equally close to the truth, means that we can tell that they stand in the same relation to (3) (truer, less true or as true as) as each other, without knowing anything about (3). But even this seems too much determinacy, for in reality none of these is closer to the truth than any other. Note that the problem here does not concern the details of the model, the actual numbers, but the way they represent the world, which I think is quite telling against even the 'useful fiction' response.

1.2.   Higher Order Vagueness

As it stands, *M* has no way of dealing with higher-order vagueness. Intuitively, there might not only be borderline cases of great baseball players, but borderline cases of borderline cases of greatness with respect to baseball. If every sentence *NN is a great baseball player* either does or does not receive an integer truth value, then this intuitive possibility is ruled out. There is a quite deep problem here, one that we can't solve by, say, adding more truth values, somehow between 0 and (0, 1]. (Buying the latest theory of infinitesimals off the shelf will not save *M*.)

We cannot solve the problem simply by iterating the theory. (This is a point stressed by Williamson 1994: ch. 4.) We cannot say that it is true to degree 0.5 than (2) is true to degree 1, and true to degree 0.5 that it is true to degree 0.8. For then it is only true to degree 0.5 that (2) has some truth value or other. And the use of truth-tables to generate a logic presupposes that every sentence has some truth values or other. If this is not determinately true, *M* is not a complete theory.

Now, perhaps we can tell some *other* story about higher-order vagueness that will get us out of this difficulty. Such a move is fairly desperate, but not utterly hopeless. Perhaps we need to say different things about first-order and higher-order vagueness. Still, it would be better to have a unified theory.

1.3.   *Classical Logic*

One of the points of adopting *M* is to escape the tyrannical grasp of classical logic. But the rebellion goes too far. For one thing, it isn't too implausible to think that sentences of the form shown in (4) are logical truths, even if the substitutions for *A* and *B* are vague.

(4) $(A \,\&\, (A \to B)) \to B$

More importantly, *M* treats conjunctions very badly. Some contradictions can end up with a truth value as high as 0.5. Williamson (1994) twice suggests this is a very serious problem.

> The failure of excluded middle may seem natural enough in borderline cases. More disturbing is that the law of non-contradiction fails in the same way. $\sim(p \,\&\, \sim p)$ always has the same degree of truth as $p \lor \sim p$, and thus is perfectly true only when *p* is either perfectly true or perfectly false. When *p* is half-true, so are both $p \,\&\, \sim p$ and $\sim(p \,\&\, \sim p)$. (118)

> At some point [in waking up] 'He is awake' is supposed to be half-true, so 'He is not awake' will be half-true too. Then 'He is awake and he is not awake' will count as half-true. How can an explicit contradiction be true to any degree other than 0? (136)

I think this is a very telling objection, especially since, as we shall see, there are ways to avoid it while keeping the primary insight of *M*, that some vague sentences really have intermediate truth values.

2. Supervaluationism

Supervaluational theories do not have the first or third problem, but they still have the second problem, and two other problems as well. (I assume here a familiarity with how supervaluational theories work.)

2.1.   *Higher-Order Vagueness (Again)*

As Rosanna Keefe notes, the following argument seems to be telling against the idea that supervaluationism can deal with higher-order vagueness:

> According to the theory, a sentence is true simpliciter iff it is true on all complete and admissible specifications [her term for precisifications]. But for any sentence, either it *is* true on all complete and admissible specifications (hence true simpliciter) or not (hence borderline or false). So there is no scope for avoiding sharp boundaries to the borderline cases or for accommodating borderline borderline cases. (202)

Keefe thinks this argument, though popular, is ultimately unsuccessful. I will argue here that there is something importantly right about it. Keefe's first response is to claim that the argument assumes, falsely, that there is "a precise and unique set of complete and admissible specifications." (202) But it is not clear where the argument does assume this. After all, the argument makes no mention of sets. Secondly, it is a little unclear just why this assumption should be false. Keefe argues that there is no such set because *complete and admissible specification* is vague, just as there is no precise and unique set of tall things because *tall* is vague. This is puzzling on several levels.

First, even though *complete and admissible specification* is vague, on every specification there is still a unique set of complete and admissible specifications. I am not sure what work *precise* is doing here, but it seems that on every specification that set is precise. Perhaps, then, *there is a unique and precise set of complete and admissible specifications* is one of those sentences that for some reason, we know not what, we should not supervaluate. (Lewis 1993 says that we should not supervaluate every sentence, and pragmatics tell us when we should apply supervaluationist principles. Perhaps this is right, though without some sense of just which pragmatic principle could possibly do the relevant work here, it's hard to make a final judgment.)

Secondly, even if this term *complete and admissible specification* is vague, it is not clear just how that should affect the argument. The idea was that a sentence is true iff it is true on all specifications (precisifications) that are complete and admissible, not iff it is true on all specifications satisfying the term *complete and admissible*. It is, I think, just a use/mention confusion to hold the latter view. But if the former is correct, the vagueness of any term is irrelevant to the above argument. Remember, supervaluationists hold, rightly, that vagueness is a semantic phenomenon. Properties and individuals are not vague, but some terms that purport to denote properties and individuals are. (As Trenton Merricks 2001 points out, it's a little hard to say all this consistently since it seems to imply that some properties, the semantic ones, are vague. In my (2003a), I argue that this is consistent. Once we clarify what kind of thing a semantic property is, and just what kinds of thing have them, it becomes clear that there can be semantic vagueness without vague semantic properties. But I think Merricks is onto a relatively deep issue here. Part of what is at issue between some supervaluationists and their opponents is that they don't believe that there are any semantic properties, at least in the 'sparse' sense of properties discussed in Lewis 1983. Space prevents anything like a full discussion of this point here, but I hope to return to it in future work.)

According to supervaluationism, a sentence is true iff it is true on all specifications having some property, one roughly captured by the term *complete and admissible precisification*. But it is the property that matters, not the words. And since properties are precise, there should be a precise and unique set of specifications having the crucial property, call this the *SV*-set. Now we can run the above argument replacing every occurrence of *complete and admissible specification* with *member of the SV-set*. It seems now that the argument is sound, showing that supervaluationism cannot accommodate higher-order vagueness.

2.2.   *Universalisability*

Supervaluationism cannot be consistently applied to all sentences. If it is, we at least seem to get the absurd result that *Every word is precise* is a true sentence, because on every precisification, every word is precise. (That's why they are *precis*ifications.) David Lewis (1993), in acknowledging this objection, says that we use pragmatic rules of interpretation to know when to supervaluate sentences, and when not to. If we had to rely on such rules of interpretation, we probably could (people are such good pragmatic interpreters, after all), but it would be nice to have a theory that did not rely on such machinery.

2.3.   **Inference Rules**

It is useful to distinguish two kinds of supervaluational theory. (This distinction is drawn most clearly in McGee and McLaughlin 1995.) One kind says that truth is supertruth - any sentence that is not supertrue is, thereby, not true. The other kind says that supertruth is sufficient, but not necessary, for truth. If a sentence is neither supertrue nor superfalse, then it is indeterminate whether that sentence is true. Intuitively, we say *S is true* is true according to a precisification iff *S* is true according to that precisification. *S is true* is just as determinate, or indeterminate, as *S*.

I much prefer the second option, because it avoids a few objections that are specific to the first. Keefe (2000: Ch. 7) responds to these objections, but not always successfully. One of these, which I don't want to rest too much weight on, is that on the first approach the T-schema is no longer universally valid. As Keefe says, we presumably have to do *something* to the T-schema to avoid the semantic paradoxes, so this may not be too large a cost.

A bigger problem is that, as Williamson (1994: Ch. 5) pointed out, this version of supervaluationism invalidates several inference rules that we normally take to be sound. The problem, in essence, is that *A* entails *It is supertrue that A* on this theory, and vice versa. However, it is not the case that *A* and *It is supertrue that A* are everywhere substitutable. This might be considered a rather bad problem already, but if anything worse is to come. The following is a fairly weak version of an $\vee$-Elimination, or Argument By Cases, rule.

$$\frac{A \vdash C \qquad\qquad B \vdash C \qquad\qquad \Gamma \vdash A \vee B}{\Gamma \vdash C}$$

Now this won't do for the supervaluationist. Let *p* be any indeterminate claim, let *A* be *p*, *B* be $\neg p$, *C* be *Determinately p or Determinately $\neg p$*, and Γ be $\{A \vee B\}$. Since on this version of supervaluationism, *p* entails *Determinately p*, all three of the entailments on the top line hold. But the derived entailment, $\Gamma \vdash C$ does not hold, since $A \vee B$ is a logical truth, and *C* is false. This looks fairly bad for this version of supervaluationism, since the rule is rather plausible. Keefe thinks that we take the rule to be plausible because we confuse it with the correct rule:

$$\frac{A \vdash C \qquad B \vdash C \qquad \Gamma \vdash \mathit{Det}\, A \lor \mathit{Det}\, B}{\Gamma \vdash C}$$

And supervaluationism does not yield any counterexamples to this rule. But it is hard to believe that this is the real $\lor$-Elimination rule. This rule is too weak to even allow us to prove $A \lor B \vdash B \lor A$, which even supervaluationists agree is valid. It is unclear whether we can provide *any* acceptable set of inference rules once we identify truth with supertruth. And that seems like a fairly striking problem.

2.4.   Identifying Determinate Truths

There is a simple way around the last problem. If we say that supertruth is sufficient, but not necessary, for truth, we can simply adopt classical inference rules. The problem with this approach is that it becomes difficult, if not impossible, to say just which truths are determinate, and hence to say which precisifications are acceptable. If truth is supertruth, then an acceptable precisification is just a maximal consistent extension of the set of true sentences. However, if we adopt the 'disquotational' approach, and say that *S is true* always takes the same truth value as *S*, then the set of true sentences is already maximal, so there is only one maximal consistent 'extension' of the set of true sentences, itself. So we need some other way of identifying the determinate truths. Space prevents me completely running through this debate again - see McGee and McLaughlin (1995) and Williamson (1995) for the most useful back and forth on the issue. Suffice to say no supervaluationist solution to this problem has yet been presented. (I assume, perhaps charitably, that this is not a problem for the other version of supervaluationism, because on that theory we can identify determinate truth with truth, and truth is well enough understood that this counts as an explanatory analysis.)

Even if these supervaluationists do provide an enlightening account of what determinate truth comes to on their theory, there remains a further problem, one hinted at by Fodor and Lepore (1996). As noted above, if there is more than one acceptable precisification, then some precisifications classify some false sentences as true, and vice versa. Given this, it is hard to know why we should care about precisifications. If precisifications are mismodels of the language, then why does the truth of a sentence on all precisifications imply its truth?

2.5.   Motivation

This leads to our final problem with supervaluationism - that it is in general undermotivated. Even committed supervaluationists like Kit Fine (1975) and Keefe (2000) have been forced to concede that supervaluationism has some counterintuitive features, such as the claim that all sentences of the form $p \lor \neg p$ are true, even when p is indeterminate. Supervaluationists rarely defend these consequences directly, but rather say that all things considered, supervaluationism gets closer than any of its rivals at capturing our pre-theoretic intuitions. I think the set of trade-offs that supervaluationism commits us to is not unreasonable - excluded middle is a little unintuitive but all the theories that reject it seem to have even less desirable consequences - but it would be good to have a

more positive argument than just that the rivals are worse. The theory I will offer has the resources to provide such an argument.

3. Epistemicism

The epistemic theory of vagueness, according to which there is an unknowable fact of the matter about the location of the boundary drawn by any vague predicate, has a quite different set of costs and benefits to the two theories sketched above. Although epistemicism is compatible in theory with any underlying logic, most of its adherents believe that classical logic is the right logic for ordinary language. One of the primary motivations for epistemicism is the view that the widespread intuitions that there could not be a metaphysically sharp boundary drawn by a vague predicate rest of a confusion epistemology and ontology. We see that we cannot, and perhaps in principle could not, find the border between being the great baseball players and the not-quite-great, and we conclude that there is no border to be found. If that were the argument for semantic indeterminacy it would rest on an implausible form of verificiationism. But I think the intuitions here do not, in fact, rest on any particular epistemic considerations. Rather, the two important intuitions are direct intuitions about the metaphysics of content. The first intuition is expressed in a famous passage by Jerry Fodor.

> I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of *spin*, *charm*, and *charge* will perhaps appear upon their list. But *aboutness* surely won't; intentionality simply doesn't go that deep. It's hard to see, in face of this consideration, how one can be a Realist about intentionality without also being, to some extent or other, a Reductionist. (Fodor 1987: 97)

So the first intuition implies that that if *wealthy* denotes the property of having more than \$850,000, rather than the property of having more than \$840,000, this cannot be a primitive fact, it must rest on some other facts. The alternative would be primitive meaning facts, and these seem monumentally implausible.

The second intuition is that there are really only three kinds of facts on which meaning facts rest. That is, if a particular thing, a brain state perhaps, or a string of chalk markings, have a meaning, this meaning must be determined by the interaction of three kinds of facts. First, there is the connection between entities of this type and other entities with meanings. This cannot give us primitive intentionality, but it is obviously crucial in practice. The second are correlations, either causal or merely statistical, between meaningful things and things in the world. (Because of the first clause, this might include the correlation between entities of different types to this one with things in the world.) And, finally, facts about which properties are the 'ultimate and irreducible properties of things' can determine meanings. In David Lewis's terminology, sometimes a predicate denotes F rather than G simply because F is more 'natural' than G. The important intuition is that no facts beyond these can determine content - if these three kinds of facts do not determine the precise content of an entity on an occasion, then it does

not have a precise content. (The argument here owes quite a bit to the arguments for indeterminacy in Sider 2001 and Burgess 2001.)

Timothy Williamson has made a sustained effort to respond to this argument, but has never really been successful. Some of the arguments have been, in effect, to deny one or other of the intuitions here. The better arguments have tried to show that the intuitions do not undermine epistemicism, since there are enough facts, especially of my first or second kind, to go around. John Burgess (2001) has demonstrated that none of these moves work. (Burgess's argument has a small lacunae, but I've shown how to patch that. See Weatherson 2003b.)

If there is a fact of the matter about where the boundary lies between those that instantiate *wealthy* and those who do not, then there is a fact of the matter about whether the predicate applies or not to each of the infinitude of possible cases scattered around its boundary. It is implausible that there are any natural concepts around here for our predicate to lock onto, and it is implausible that the pattern of occurrences of the term *wealthy* or the concept WEALTHY, or any other term or concept whose meaning may be relevant to the content of *wealthy*, is detailed enough to generate an answer for each of these cases. Even taken collectively, users of the language have not been confronted with enough cases that there reactions could be sharp enough to generate the number of answers needed. This argument has gone by rather quickly, but I hope I've sketched enough reasons here to justify continuing working on theories that allow for semantic indeterminacy, despite the problems with the two best-known such theories, *M* and supervaluationism.

4.   *M* and Contradictions

It is crucial to the philosophical understanding of *M* that we take the truth values between 0 and 1 to be genuinely *intermediate* truth values. Consider what happens when we do not take them that way. (The following few paragraphs are indebted pretty heavily to the criticisms of Strawson's theory of descriptions in Dummett 1959.)

One could have a theory of vagueness based on the logic L_C, that's the logic given by the above axiomatisation of *M*, with indeed the same semantics as *M*, without the assumption that the numbers that lie at the heart of that semantics represent truth values. Rather, on this interpretation, there are only two truth values (True and False!), and the numbers represent *ways* of taking truth values. Any sentence that gets value 1 is true, the others are all false, though many of them are false in different ways. (As Tolstoy might have put it, all true sentences are alike, but every false sentence is false in its own unique way.) The ways in which sentences are false might matter to the truth value of compounds containing that sentence. In particular, if *A* and *B* are false, then the truth values of *Not A* and *If A then B* will depend on the ways *A* and *B* take their truth values. If $V(A) = 0$ and $V(B) = 0.3$, then *Not A* and *If A then B* will be true, but if $V(A)$ becomes 0.6, and remember this is just another way of being false, both *Not A* and *If A then B* will be false.

The new theory we get, one I'll call M_D, is similar to *M* in some respects. Most prominently, it agrees about what the axioms should be for a logic for natural language. But it has enough philosophical differences that it seems worthwhile to treat it as a different theory. For one thing, it cannot explain what is going on in the Sorites with the same panache as *M*. If any sentence with truth value below 1 is false, then many of the premises in a

Sorites argument are *false*. This is terrible – it was bad enough to be told that one of the premises were false, but now we find many thousands of them are false. I doubt that being told they are false in a distinctive way will improve our estimation of the theory. Similarly, it is hard to see just how the new theory has anything interesting to say about the concept of a borderline case.

Against that, M_D does not have all the costs of M: on the new theory a contradiction is always false. To be sure, it might be false in some obscure new way, but we should be grateful for its falsity. Compare again Williamson's objection to how M handles contradictions.

> At some point [in waking up] 'He is awake' is supposed to be half-true, so 'He is not awake' will be half-true too. Then 'He is awake and he is not awake' will count as half-true. How can an explicit contradiction be true to any degree other than 0? (136)

This objection only goes through if being true to degree 0.5 is meant to be semantically significant. If being 'true to degree 0.5' is just another way of being false, then there is presumably nothing wrong with contradictions are true to degree 0.5. This is not to say Williamson's objection is no good. In fact it is a perfectly sound objection to M. What this reveals is that re-interpreting the numbers in $(0, 1)$ as ways of being false, rather than as new truth values, makes a difference to the philosophical plausibility of the theory.

I don't want to judge whether M_D is a better or worse theory of vagueness than M. All I want to stress here is that it is a *different* theory of vagueness, and that we grasp the difference between these theories. One crucial difference between the two theories is that in M, but not M_D, S_1 is truer than S_2 if $v(S_1)$ is greater than $v(S_2)$. In M_D, if S_1 is truer than S_2, $v(S_1)$ must be one and $v(S_2)$ less than one. And that, I think, is the only difference between the two theories. So if we understand this difference, we must grasp this concept *truer than*. Indeed, it is in virtue of grasping this concept that we understand why saying each of the Sorites conditionals is almost true is a *prima facie* plausible response to the Sorites, and why having a theory that implies contradictions are truer than many other sentences is a rather embarrassing thing. So I intend to develop my theory on the basis of this concept.

Note that I have implicitly defined what I mean by *truer* by noting its theoretical role. As David Lewis (1972) showed by building on work by Ramsey and Carnap, terms can be implicitly defined by their theoretical role. There is one unfortunate twist here in that the term *truer* is defined by its role in a *false* theory, but presumably that does not prevent the implicit definition story going through. We all know what the words *phlogiston* and *ether* mean by extracting their role in a particular theory; that is how we know there isn't any of either (at least nearby). The meaning of *truer* can be extracted in the same way. Note that most theorists should not object to there *being* such a concept – even a hard core epistemicist can accept the concept exists. On her theory A is truer than B just in case A is true and B false. So the existence of the concept is not particularly radical.

5. Truer

The concept *truer than* will play a crucial role in what follows, so I should say a few more words about its nature. I mean not to give a reductive analysis of *truer*. The hopes for doing that are no better than the hopes of giving a reductive analysis of *true*. What I do aim to show is that we intuitively understand the concept well enough that it can be used in an informative philosophical theory of vagueness. Indeed, I hope the following range of considerations will convince you that I'm latching onto a concept you already possess, and these considerations will help isolate the concept, if not fully explicate it a la *Meno*.

My primary argument for this has already been given. Intuitively we do understand the difference between M and its M_D, and this is only explicable by our understanding *truer*. Hence we understand *truer*.

Second, *truer* occurs fairly often in ordinary language, and it's reasonable to conclude that we competent speakers know the meaning of commonly used terms in natural language. (I just found 237,000 hits in Google for *truer*, although to be fair at least a handful of those are to my uses of the term, so perhaps only 236,000 or so are genuinely data points.) Perhaps not coincidentally, the word also recurs frequently in Shakespeare. Many of the uses, both in Shakespeare and in real life, do not match up with the usage we have here. When we say that x is a truer friend than y, Shakespeare's most common usage, we certainly do not mean the same thing. This of course is because the meaning of *truer* there is generated by applying the *-er* suffix to *true* as it appears in *true friend*, not to *true* as it appears in *true sentence*. But sometimes we do use *truer* just as I mean here. In *Midsummer Night's Dream*, for instance, Shakespeare means by *truer* just what I mean by it in this paper.

> **LYSANDER** Ay, by my life;
>   And never did desire to see thee more.
>   Therefore be out of hope, of question, of doubt;
>   Be certain, nothing truer; 'tis no jest
>   That I do hate thee and love Helena (Act III, Scene 2)

I think it's a reasonable philosophical principle that we can appeal to any concept in Shakespeare without analysis without anyone accusing us of using concepts that are *too* obscure.

Third, it's noteworthy that *truer* is morphologically complex. If we understand *true*, and understand the modifier *-er*, then we know enough in principle to know how they combine. To be sure, *-er* cannot be appended to every predicate, and perhaps some predicates cannot be turned into comparatives by any means. (The two clauses are distinct, because some predicates cannot be combined with *-er* but can be turned into comparatives using the *more…than* construction. *Intelligent* is an example of this.) But I think it needs argument to show why *true* should be unlike the majority of predicates in this respect.

I have heard two arguments to just that conclusion. First, it could be argued that most comparatives in English generate linear orderings, but *truer* I say generates a non-linear ordering. I reject the premise of this

argument. *Cuter*, *Smarter*, *Smellier*, and *Tougher* all generate non-linear orderings over their respective domains. Second, it could be argued that it's crucial to our understanding of familiar comparatives that we understand the behaviour of comparison classes with respect to the underlying adjectives. Robin Jeshion and Mike Nelson made this objection in their comments on my paper at BSPC 2003. If comparison classes are crucial to the semantics for comparatives, as proposed by Ewan Klein (1980), this objection has quite a bit of force. However, it is not clear that all comparatives require that the underlying predicate interact with a comparison class in natural ways. We can talk about some objects being *straighter* or *rounder* despite the fact that it's hard to understand *round for an office building* or *straight for a line drive*. (Jonathan Bennett made this point in discussion at BSPC.) Arguably, *straight* and *round* either don't have or don't need comparison classes, but they can form comparatives. So perhaps *true*, which also does not take comparison classes, is not blocked from forming a comparative for this reason.

Finally, if one thinks that understanding the inferential role of a logical operator takes one a long way towards knowing its meaning, it is worth noting that by the end of the paper I'll have specified quite precisely the inferential role of *truer*. It is the same as a strict material implication *Necessarily* $(p \supset q)$ defined using a necessity operator whose logic is KT. Since many operators have just this logic, this hardly serves to individuate *truer*, but it might be worth something to those with an affinity for inferential role semantics.

6. Truer and Borderline Cases

So everyone should agree that there is an operator *truer* that either could be added to English, or, as I think, already appears in English. Even an epistemicist can think that such an operator is defined with *p is truer than q* just meaning *p* is true and *q* is false. I do not think this is its only application. In particular, I think that there are 'intermediate' sentences with respect to *truer*. A sentence *p* is intermediate iff there are sentences *q* and *r* such that *q* is truer than *p*, and *p* is truer than *r*. In symbols

p is intermediate $=_{df} \exists q, r: q >_T p$ and $p >_T r$

I will use $>_T$ as short for *truer than* in what follows. The sentences here are sentence tokens, and the quantification is over *possible* sentence tokens. All quantifiers over words and sentences in what follows are possibilist quantifiers. My analysis of a borderline case is that *x* is a borderline case of an *F* iff for some *a*, *a* refers to *x* and *a is F* is intermediate. This already, I think, gives us some of the benefits of *M*. As the denotation of *a* gets wealthier and wealthier, *a is F* gets truer and truer, until it becomes perfectly true. Conversely, my analysis of determinacy is that *x* is a determinate case of an *F* iff for some *a*, *a* refers to *x* and *a is F* is as true as a logical truth. For definiteness, I will take 0=0 to be my paradigm of a logical truth. For completeness, note that *x* is a determinate non-*F* if *a is determinately F* is as true as 0=1, assuming again that *a* names *x*.

One might worry that this will lead to a theory that cannot handle higher order vagueness. As long as we treat *truer than* as an non-trivially iterable operator on sentences, this worry is misguided. There are borderline

12

cases of *being determinately F* as long as there are intermediate sentences of the form *a is determinately F*. Formally, this will be the case if the following is ever true.

$$0{=}0 >_T (a\ is\ F =_T 0{=}0) >_T 0{=}1$$

As we shall see in the formal model below, nothing in the concept *truer* rules out the possibility of this. As you may have guessed, I use $=_T$ as shorthand for *as true as*. It will be very useful to have a shorthand for *truer than or as true as*, and I will use $\geq_T$ for this operator.

One of the objections to *M* was that it implied that *truer than* is linear. That is, it implies:

$$\forall p, q: (p >_T q) \vee (p =_T q) \vee (q >_T p),\ \text{or, equivalently}$$
$$\forall p, q: (p \geq_T q) \vee (q \geq_T p)$$

Ideally, our theory would not have this result. Again, nothing in the concept *truer* seems to require this conclusion. It is simply an unfortunate assumption embedded in *M*. To see this, assume that we could somehow solve the problems associated with supervaluationism, and define a set of precisifications. Then we could provide the following definition of *truer*.

$p >_T q$ iff the set of acceptable precisifications on which p is true is a proper superset of the set of acceptable precisifications on which q.

$p =_T q$ iff p and q are true on exactly the same acceptable precisifications.

Now this won't do as a definition, because as noted above it is impossible to independently define what is an acceptable precisification. But it is a useful model. Note that given this definition, it is possible to have p and q such that neither $(p >_T q)$ nor $(p =_T q)$ nor $(q >_T p)$ is true. For instance, (1) and (2) are true on quite different precisifications, so none of the three disjuncts would be true if we substituted them for p and q. So it is possible to provide formal models for *truer* that don't commit us to it being a linear relation. That already puts us in a better position to provide a realistic treatment of borderline cases than we would be if we endorsed *M*.

Note that on this model, all classical tautologies are as true as $0{=}0$, since they are all true on all precisifications. This is an important general point. That there are intermediate sentences does not entail that classical tautologies are amongst them. Indeed, we can accept that there are intermediate sentences without in any way qualifying our commitment to classical logic. We do, I think, thereby qualify our commitment to what is normally taken to be classical semantics, since that semantics is bivalent, and the existence of intermediate sentences is incompatible with bivalence. (Though it is not, it turns, incompatible with some familiar statements of bivalence. If we adopt a disquotational theory of truth, then *S is true or S is not true* is equivalent to *S or not S*,

and as we shall see, in the preferred analysis of truer, this is always perfectly true.) The important technical point is that there is more than one semantic model for classical logic, so we can give up the truth tables without giving up classical logic.

(For accounting purposes, it is worth noting that I have introduced a new concept in this section: *as true as*. Since *truer* is not linear, this cannot be simply analysed as follows: *A* is as true as *B* iff neither *A* is truer than *B* nor *B* is truer than *A*. This biconditional will fail when *A* and *B* are incomparable. The right hand side will be true, and the left hand side false. We have two options here. One option is to take *as true as* to be a new concept, identified by the same considerations as *truer* was in the previous section. A second is to use a different analysis. The following analysis gets the intension of *as true as* right, though I'm not sure it's philosophically very enlightening: *As true as* denotes the largest relation that is an equivalence relation with respect to *truer than*.)

7.   Constraints on Truer and Classical Logic

One core claim of my theory is that we know rather little about truer. We do not know enough, for instance, to use it to generate anything like a valuation function on sentences, as the proponent of M assumes it does. Since that assumption leads directly to the false precision objection to M, this is a good deviation from M. But we do know a few things about M, including the following ten claims. (I've listed here both the informal claim, which is what is philosophically important, and the formal interpretation of that claim.)

(A1) $\geq_T$ is a weak ordering (i.e. reflexive and transitive)

   If $A \geq_T B$ and $B \geq_T C$ then $A \geq_T C$

   $A \geq_T A$

(A2) $\wedge$ is a greatest lower bound with respect to $\geq_T$

   $A \wedge B \geq_T C$ iff $A \geq_T C$ and $B \geq_T C$

   $C \geq_T A \wedge B$ iff for all S such that $A \geq_T S$ and $B \geq_T S$ it is also the case that $C \geq_T S$

(A3) $\vee$ is a least upper bound with respect to $\geq_T$

   $A \vee B \geq_T C$ iff for all S such that $S \geq_T A$ and $S \geq_T B$, it is also the case that $S \geq_T C$

   $C \geq_T A \vee B$ iff $C \geq_T A$ and $B \geq_T C$

(A4) $\neg$ is ordering inverting with respect to $\geq_T$

   $A \geq_T B$ iff $\neg B \geq_T \neg A$

(A5) Double negation is redundant

   $\neg\neg A =_T A$

(A6) There is an absolutely false sentence S_F and an absolutely true sentence S_T

   There are sentences S_F and S_T such that $S_F =_T \neg S_T$ and $\neg S_F =_T S_T$ and for all S: $S_T \geq_T S \geq_T S_F$

(A7) Contradictions are absolutely false

   $A \wedge \neg A =_T S_F$

14

(A8) $\forall$ is a greatest lower bound with respect to $\geq_T$

   $A \geq_T \forall x(\phi x)$ iff for all S such that for all o, if n is a name of o then $\phi n \geq_T S$, it is the case that $A \geq_T S$

   $\forall x(\phi x) \geq_T A$ iff for all o, if n is a name of o then $\phi n \geq_T A$

(A9) $\exists$ iss least upper bound with respect to $\geq_T$

   $A \geq_T \exists x(\phi x)$ iff for all o, if n is a name of o then $A \geq_T \phi n$

   $\exists x(\phi x) \geq_T A$ iff for all S such that for all o, if n is a name of o then $S \geq_T \phi n$, $S \geq_T A$

(A10) A material implication with respect to $\geq_T$ can be defined

   There is an operative $\rightarrow$ such that

   (a) $B \rightarrow A \geq_T S_T$ iff $A \geq_T B$

   (b) $(A \wedge B) \rightarrow C =_T A \rightarrow (B \rightarrow C)$

Most of these are fairly straightforward, but let me say a little bit about (A10). The argument for (A10) is not that English *if…then* is a material implication in this sense. Saying that would be to commit us to the paradoxes of material implication. Nor am I assuming that $\neg A \vee B$ is a material implication. I believe that is true, but I do not think it is obviously true. The reason that it is not obviously true is that we can derive immediately from that definition and clause (a) that $\neg A \vee A$ is as true as 0=0 by just substituting $\neg A$ for B. And it is not obvious that we should keep excluded middle once we acknowledge that there are intermediate sentences. I am *not* saying here it is obvious we should scrap excluded middle, but I think it is not obvious that we should keep it. The argument for (A10) is a little more indirect than this.

   Consideration of how quantified propositions must be composed provides a good reason to believe in such a material implication. There are two ways we might try to formulate a quantified sentence, say *All Fs are Gs*. First, we may formulate it using unrestricted quantifiers. In that case the sentence will presumably look something like this:

   $\forall x(Fx\ ?\ Gx)$

where some connective goes in place of '?'. It seems very plausible that whatever goes in there should be a material implication. Alternatively, we might formulate it using restricted quantifiers, so the sentence will look something like this:

   $[\forall x: Fx]\, Gx$

In that case, we can define a connective ∇ that looks like it satisfies the two clauses in the definition of a material implication:

15

$A \, \nabla \, B =_{df} [\forall x: A \wedge x=x] \, (B \wedge x=x)$

Essentially here we are using null quantifiers; if it were well-formed we would say something like *Everything such that A is such that B*, where *A* and *B* are closed sentences. Consideration of the logic for quantifiers, particularly the fact that *All Fs are Fs* must be a logical truth, suggests that ∇ might well be a material implication. So I suggest that getting the logic of the quantifiers right will imply that there must be a material implication, as in (A10).

It is provable, by the way, that if (A1) to (A10) are all satisfied, then this material implication must be equivalent to $\neg A \vee B$. This is a somewhat surprising result, since it guarantees that excluded middle is to be preserved. But I do not think it is such an implausible *conclusion* that drawing it should invalidate the whole theory, even if I think it is contentious enough that it should not be used as a *premise* in arguing about the logic appropriate for vague natural languages.

What is interesting about these ten constraints is that they suffice for classical logic, with just one more supposition. I assume that an argument is *valid* iff it is impossible for the premises taken collectively to be truer than the conclusion. That is, an argument is valid iff it is impossible for the conjunction of the premises to be truer than the conclusion. Given that, we can get the following result:

$\forall A_1, \ldots, A_n, B: A_1, \ldots, A_n \vDash_T B$ iff, according to classical logic, $A_1, \ldots, A_n \vdash B$

(I use $\Gamma \vDash_T A$ to mean that in all models for $\geq_T$ that satisfy the constraints, here (A1) to (A10), the conclusion is at least as true as the greatest lower bound of the premises.) I won't prove this result, but the idea is that (A1) to (A10) imply that $\geq_T$ defines a Boolean lattice over equivalence classes of sentences with respect to $=_T$. And all Boolean lattices are models for classical logic, from which our result follows. (Note that this result only holds in the right-to-left direction for languages that contain the $\geq_T$ operator. Once this operator is added, some arguments that are not classically valid, such as $B \geq_T A, A \vDash_T B$, will valid. But the addition of this operator is conservative: if we look at the $\geq_T$-free fragment of such languages, the above result still holds. In the next section we will look at languages with such an operator.)

I think (A1) to (A10) are the only constraints on *truer*. It is worth noting one special class of models for those axioms, because these models illustrate how weak the analytic constraints on *truer* are. If the constraints were so tight as to prevent $0=0 >_T (A =_T 0=0) >_T 0=1$ from ever being true, then we would have ruled out higher-order vagueness. That would be a disaster, so it is worth noting how this possibility is left alive.

Let $<W, R, V>$ be a familiar Kripke model for KT. So W is a set, intuitively of worlds, R is a reflexive relation on W, and V a function from atomic sentences to subsets of W, intuitively the set of worlds at which they are true. By familiar methods then, V is extended to a function on all sentences. We can now define $\geq_T$ on the model. $A \geq_T B$ iff $\Box(B \supset A)$ is true at w. If $\Box A$ is true at some, but not all, of the worlds accessibly from w, then

$0=0 >_T (A =_T 0=0) >_T 0=1$ will be true in the model, as required. So adopting (A1) to (A10) does not rule out higher-order vagueness.

8. Semantics and Proof Theory

The last idea leads to a natural picture for how to generate a semantics and proof theory for a logic containing *truer* as an iterable operator. Consider the following (minor) variant on KT. In the syntax for the language, we do not introduce $\Box$ as an operator that maps any sentence onto another sentence. Rather, $\Box A$ is only well-formed if A is of the form $B \rightarrow C$. Call the resulting logic KT_R, with the R indicating that we are dealing with a restricted logic. The restriction makes very little difference. Since A is equivalent to $(A \rightarrow A) \rightarrow A$ in classical logic (or in any other logic worth considering as a logic for vague languages), we can always express what $\Box A$ says in KT by $\Box(A \rightarrow A) \rightarrow A$ in KT_R. The Kripke models for KT_R are quite natural. $\Box(B \rightarrow C)$ is true at a point iff all accessible points at which B is true are points at which C is true. (Naturally, the accessibility relation must be reflexive, and need not satisfy any other formal constraints. It is harder to deal with the behaviour of $\Diamond$ in KT_R than the behaviour of $\Box$, since if we take $\Diamond$ to be defined as $\neg\Box\neg$, then it is impossible for both $\Diamond A$ and $\Box A$ to be well-formed sentences of KT_R. We will only consider the role of $\Box$ in KT_R in what follows.)

Since KT_R is so similar to KT, we can derive most of its formal properties by looking at the derivations of similar properties for KT. (The next few paragraphs owe a lot to Goldblatt 1992: Chs. 1-3.) Let's start with a proof theory. We'll present an axiomatic proof system first, with a natural deduction system to follow below. The axioms for KT_R are:

- All classical tautologies
- All well-formed instances of K: $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$
- All well-formed instances of T: $\Box A \rightarrow A$

The rules for KT_R are

- Modus Ponens: If $A \rightarrow B$ is a theorem and A is a theorem, then B is a theorem
- Restricted Necessitation: If A is a theorem and $\Box A$ is well-formed, then $\Box A$ is a theorem.

Given these, we can now define a maximal consistent set for KT_R. It is a set of sentences with the following three properties:

- All theorems of KT_R are in the set.
- For any well-formed sentence A, either A is in the set or $\neg A$ is in the set.
- The set is closed under modus ponens.

The existence of Kripke models for KT_R show that some maximal consistent sets exist: the set of truths at any point will be a maximal consistent set. The canonical model for KT_R is $<W, R, V>$ where

- W is the set of maximal consistent sets for KT_R
- R is the relation such that $w_1 R w_2$ iff for all A such that $\Box A \in w_1$, $A \in w_2$
- V is the valuation such that $V(A) = \{w : A \in w\}$

Since all instances of T are theorems, it can be easily shown that R is reflexive, and hence that this is a frame for KT_R and hence that KT_R is canonically complete.

We can translate all sentences of KT_R into a language that does not contain $\square$, but does contain $\geq_T$. The only move needed is to replace $\square(B \rightarrow A)$ with $A \geq_T B$ wherever $\square$ occurs. Since $\square$ always occurs in front of a conditional, this translation procedure will remove all instances of $\square$. Translating the axioms for KT_R, we get the following axioms for the logic of $\geq_T$.

- All classical tautologies
- All instances of: $(B \geq_T A) \rightarrow ((A \geq_T (A \rightarrow A)) \rightarrow (B \geq_T (B \rightarrow B)))$
- All instances of: $(B \geq_T A) \rightarrow (A \rightarrow B)$

The rules are

- Modus ponens
- Determination: If $A \rightarrow B$ is a theorem, then $B \geq_T A$ is a theorem

We can simplify somewhat by replacing the second axiom schema with

- All instances of: $A \geq_T B \rightarrow (B \geq_T C \rightarrow A \geq_T C)$

The Kripke models for $\geq_T$ are relatively natural. We require that the relation R be reflexive, and that $B \geq_T A$ is true at a point iff all accessible points at which A is true are points at which B is true. Given that we can provide a semantic definition of validity with respect to $\geq_T$. An argument is valid iff there is no point in any such Kripke model where the premises are true and the conclusion not true.

Maximal consistent sets with respect to $\geq_T$ and a canonical model for $\geq_T$ can be easily constructed by parallel with the maximal consistent sets and canonical models for KT_R. These constructions show that if A is a theorem of the logic for $\geq_T$, then it is true at all points in all models. More generally, they can be used to show that this logic is canonically complete, though the details of the proof are left as an exercise for the reader. There will be a one-one correspondence between the points in the canonical model for KT_R and the canonical model for $\geq_T$. The maximal consistent sets for $\geq_T$, i.e. the points in the canonical model, just are the results of applying the translation rule $\square(B \rightarrow A) \Rightarrow A \geq_T B$ to the (sentences in the) maximal consistent sets for KT_R.

This last point is important because the points in the canonical model for $\geq_T$ are useful for understanding what truth is, and for understanding what languages are. (One of the frequent, and justified, criticisms of the version of this paper presented at BSPC was that I hadn't said enough about the connection between *truer* and *true*. The following is intended to address that criticism.) The set of true sentences in English is one of the points in the canonical model for $\geq_T$. For semantic purposes, languages just are points in this canonical model. It is indeterminate just which such point English is, but it is one of them. Earlier I set out the logic for $\geq_T$ by highlighting its similarity with M. For instance, I noted that we could, to a first approximation, regard the truth values as being subsets of [0, 1]. This way of presenting things makes it unclear where truth fits into the picture, for none of these truth values represents truth. The way of putting things in this section obviously highlights the similarities with supervaluationism rather than the similarities with M, for the points in the canonical model look a

lot like precisifications. It is, however, worth noting the many differences between what I say here and what supervaluationists typically say. One difference is that I identify languages with a single point rather than with a set of points. I think this is the fundamental reason the treatment of higher-order vagueness is much smoother on my account than supervaluationist accounts. Another difference is that I don't start with anything like a set of acceptable points/precisifications. I just start with all the points that are formally consistent and a vague statement that the right semantic model for English is somewhere in there. (Imagine, if you like, my vaguely pointing to a region in the model when saying this.) The most important difference is that I take the points, with the truer than relation already defined, to be primitive, and the accessibility/acceptability relation to be defined in terms of them. This reflects the fact that I take the *truer* relation to be primitive, and determinacy to be defined in terms of it, whereas typically supervaluationists do things the other way around. None of these differences are huge. They are really just differences of emphasis, but I think my way of highlighting things is less vulnerable to criticisms than the supervaluationist approach. But the main benefit of thinking about the canonical model is not that it brings out the similarities and differences between my account and supervaluationism. The main benefit is that it gives us a formal model in which both *true in a language* and *truer in a language* are defined (given the identification of languages with points), and this provides a nice representation of the relationship between *true* and *truer*.

To close this section, I will note that we can also provide a fairly straightforward natural deduction system for the logic of $\geq_T$. There are two philosophical benefits to doing this. First, it makes it explicit that in my theory, unlike in typical supervaluational theories, we can keep all the inference rules of classical logic. Second, it provides some justification for the constraints on $\geq_T$ that I endorsed in section 7. Given the relationship between $\geq_T$ and $\vdash$ as set out in the three rules concerning $\geq_T$ at the end, it is easy to see the correspondence between the natural deduction rules for $\wedge$, $\vee$, $\rightarrow$ and $\neg$ and the constraints on their relationship with $\geq_T$ in section 7. For that reason I've set out those rules here, even though they'll probably be very familiar to most readers.

($\wedge$ In) $\Gamma \vdash A, \Delta \vdash B \Rightarrow \Gamma \cup \Delta \vdash A \wedge B$

($\wedge$ Out-left) $\Gamma \vdash A \wedge B \Rightarrow \Gamma \vdash A$

($\wedge$ Out-right) $\Gamma \vdash A \wedge B \Rightarrow \Gamma \vdash B$

($\vee$ In-left) $\Gamma \vdash B \Rightarrow \Gamma \vdash A \vee B$

($\vee$ In-right) $\Gamma \vdash A \Rightarrow \Gamma \vdash A \vee B$

($\vee$ Out) $\Gamma \cup \{A\} \vdash C, \Delta \cup \{B\} \vdash C, \Lambda \vdash A \vee B \Rightarrow \Gamma \cup \Delta \cup \Lambda \vdash C$

($\rightarrow$ In) $\Gamma \cup \{A\} \vdash B \Rightarrow \Gamma \vdash A \rightarrow B$

($\rightarrow$ Out) $\Gamma \vdash A \rightarrow B, \Delta \vdash A \Rightarrow \Gamma \cup \Delta \vdash B$

($\neg$ In) $\Gamma \cup \{A\} \vdash B \wedge \neg B \Rightarrow \Gamma \vdash \neg A$

($\neg$ Out) $\Gamma \vdash \neg\neg A \Rightarrow \Gamma \vdash A$

($\geq_T$ In) $\Gamma \vdash A \Rightarrow \{B \geq_T C : B \in \Gamma\} \vdash A \geq_T C$

($\geq_T$ Convert) $\Gamma \vdash A \geq_T B \Rightarrow \Gamma \vdash (B \rightarrow A) \geq_T C$

($\geq_T$ Out) $\Gamma \vdash A \geq_T B \Rightarrow \Gamma \vdash B \to A$

It is unfortunate that we have to have two 'out' rules for $\geq_T$, but there's no simple rule that captures both the point that $A \geq_T B$ entails $(B \to A) \geq_T C$ and that it entails $B \to A$. In any case, moving from classical logic to KT_R involves three additions (two axioms, K and T, and a rule, necessitation) so the need for three rules here is not unexpected. (Thanks to Gabriel Uzquiano for several probing questions that led to this section being written.)

9. Sexy Sorites

One of the benefits of M was that it gave us a story about what was going on in the Sorites. A good theory of vagueness should tell us two things about the Sorites. First, it should tell us what is wrong with Sorites arguments. This part is relatively easy. Just about every theory of vagueness on the market agrees that what is wrong with Sorites arguments is that not all of their premises are perfectly true. (The subvaluationist theory defended by Dominic Hyde 1997 is the most prominent exception.) Second, it should tell us why the premises looked persuasive in the first place. This is something that is intrinsically nice to know - it's always refreshing to have our mistakes documented - and it's also a reasonable constraint on answers to the first puzzle that they not imply there was no serious puzzle here to begin with. Someone who simply denies the truth of the premises in a Sorites without a story as to why we may have thought they were true is not much better off, I think, than a person who responds to the Argument from Evil by simply denying that God is obliged to bring about the best of all possible worlds, but without a story as to why it might have seemed reasonable to think that an all powerful all knowing God had that obligation. It would be mean, and *perhaps* unfair, to compare them to people who simply assert this is the best of all possible worlds without any supporting argument. (On the constraint that solutions to the paradox must explain why there *seemed* to be a paradox, see Kaplan and Montague 1960. Thanks to Chris Barker for drawing my attention to the connection with this paper.)

The M theorist has the beginnings of a story, though as we'll see not the end of a story. The story starts by saying that all the premises in a typical Sorites argument are nearly true, and we think they are perfectly true because we confuse near truth for truth. Can I say the same thing, since my theory is like M? No, for two reasons. First, since my theory explicitly gets rid of numerical representations of intermediate truth values, I don't have any way to analyse *almost true*. Second, since I say that one of the Sorites premises is false, I'd be committed to the odd view that some false sentence is almost perfectly true. (Thanks to Cian Dorr for pointing out this consequence.)

So I can't tell the story the M theorist tells. Is this a bad thing? No, because that story does not generalise. The problem is that not all Sorites arguments involve conditionals. A typical Sorites situation involves a chain from a definite F to a definite not-F. Let ´ denote the successor relation in this sequence, so if F is *is tall* and a is 178cm tall, then $a´$ will be 177.99cm tall, assuming the sequence progresses 0.1mm at a time. The M theorist notes that every premise of the following form is almost perfectly true.

(SI)   If *a* is tall, then *a´* is tall.

But we need not use such premises in a Sorites argument. We could, instead, have used negated conjunctions in the statement of the argument. We could replace (SI) with (SA).

(SA)   It is not the case that *a* is tall and *a´* is not tall.

And it is not true that every premise of the form (SA) is almost true. Indeed, premises of this form are arbitrarily close in truth value to 0.5. There's no explanation in *M* for why all premises of the form of (SA) look persuasive. This is quite bad, because if anything (SA) is *more* plausible than (SI). Consider the following thought experiment. You are trying to get a group of (typically non-responsive) undergraduates to appreciate the force of the Sorites paradox. If they don't feel the force of (SI), what do you use to persuade them? My first instinct is to appeal to something like (SA). And if that doesn't work, I start appealing to theoretical considerations about how our use of *tall* couldn't possibly pick a boundary between *a* and *a´*. As far as I can tell, I find (SI) plausible *because* I find (SA) plausible, and my first reaction when teaching would be to assume my students would feel likewise. There's an asymmetry here. I can't imagine trying to defend (SA) to an undergraduate by appeal to something like (SI), nor can I imagine that the reason I find (SA) intuitively plausible is that I'm tacitly inferring it from (SI). (This is not to endorse universally quantified versions of either (SA) or (SI). They are like Axiom V - claims that remain intuitively plausible even when we know they are false.)

This is not to say that *anyone* has a decent story about why (SA) seems true. The official epistemicist story is that speakers only accept sentences that are determinately, i.e. knowably, true. But some instances of (SA) are actually *false*, and many many more are not knowably true. It's hard to see how the supervaluationist account is better in this respect. There's another problem facing every party here. In the history of debate about the Sorites, I don't think anyone has put forward a Sorites argument where the major premises have the form of (SO).

(SO)   Either *a* is not tall, or *a´* is tall.

(This point is also noticed in Braun and Sider forthcoming.) There's a good reason for this: (SO) is *not* intuitively true, unless perhaps one sees it as a roundabout way of saying (SA). In this respect it conflicts quite sharply with (SA), which *is* intuitively true. But none of the three theories discussed provide grounds for distinguishing (SA) from (SO), since they all accept DeMorgan's laws. Further, none of the many and varied recent solutions to the Sorites that do not rely on varying the underlying logic (e.g. Soames 1999, Graff 2000, Sorensen 2001, Eklund 2002) seem to do any better at distinguishing (SA) from (SO). (This is a fairly sweeping claim, and ideally it would be backed up with a *detailed* exegesis of these views to demonstrate that they don't say anything that would distinguish (SA) from (SO). But space prevents a comprehensive proof of this negative. Soames does

endorse a three-valued non-classical logic for vagueness, but this is not at the core of his solution to the *Sorites*, which is his contextualism. The three-valued logic does some work - but it does not help distinguish (SA) from (SO).) As far as I can tell none of these theories *could*, given their current conceptual resources, tell a story about why (SA) is intuitively plausible that does not falsely predict (SO) is intuitively plausible. That is, none of these theories could solve the Sorites paradox with their current resources.

There is, however, a simple theory that does predict that (SA) will look plausible while (SO) will not. Kit Fine (1975) noted that if we assume that speakers systematically confuse *p* for *Determinately p*, even when *p* occurs as a constituent of larger sentences rather than as a standalone sentence, then we can explain why speakers may accept vague instances of the law of non-contradiction, but not vague instances of the law of excluded middle. (That speakers do have these differing reactions to the two laws has been noted in a few places, most prominently Burgess and Humberstone 1987 and Tappenden 1993.) It's actually rather remarkable how many true predictions one can make using Fine's hypothesis. It correctly predicts that (5) should sound acceptable.

(5)   It is not the case that I am tall, but nor is it the case that I am not tall.

Now (5) is a contradiction, so both the fact that it sounds acceptable if I am a borderline case of vagueness, and the fact that some theory predicts this, are quite remarkable. This is about as good as it gets in terms of evidence for a philosophical claim.

(We might wonder just why Fine's hypothesis is true. One idea is that there really isn't any difference in truth value between *p* and *Determinately p*. This leads to the absurd position that some contradictions, like (5), are literally true. I prefer the following two-part explanation. The first part is that when one utters a simple subject-predicate sentence, one implicates that the subject *determinately* satisfies the predicate. This is a much stronger implicature than conversational implicature, since it is not cancellable. And it does not seem to be a conventional implicature. Rather, it falls into the category of nonconventional nonconversational implicatures Grice suggests exists on pg. 41 of his 1989. The second part is that some implicatures, including determinacy implicatures, are computed locally and the results of the computations passed up to whatever system computes the intuitive content of the whole sentence. This implies that constituents of sentences can have implicatures. This theme has been studied quite a bit recently; see Levinson 2000 for a survey of the linguistic data and Sedivy et al 1999 for some empirical evidence supporting up this claim. Just which, if any, implicatures are computed locally is a major research question, but there is *some* evidence that Fine's hypothesis is the consequence of a relatively deep fact about linguistic processing. This isn't essential to the current project - really all that matters is that Fine's hypothesis is true - but it does suggest some interesting further lines of research and connections to ongoing research projects.)

If Fine's hypothesis is true, then we have a simple explanation for the attractiveness of (SA). Speakers regularly confuse (SA) for (6), which is true, while they confuse (SO) for (7), which is false.

(6)   It is not the case that a is determinately tall and $a´$ is determinately not tall.

(7)   Either a is determinately not tall, or $a´$ is determinately tall.

This explanation cannot *directly* explain why speakers find (SI) attractive. My explanation for this, however, has already been given. The intuitive force behind (SI) comes from the fact that it follows, or at least appears to follow, from (SA), which looks practically undeniable.

So Fine's hypothesis gives us an explanation of what's going on in Sorites arguments that is available in principle to a wide variety of theorists. Fine proposed it in part to defend a supervaluationist theory, and Keefe (2000) adopts it for a similar purpose. Patrick Greenough (2003) has recently adopted a similar looking proposal to provide an epistemicist explanation of similar data. (Nothing in the explanation of the attractiveness of Sorites premises turns on any *analysis* of determinacy, so the story can be told by epistemicists and supervaluationists alike.) And the story can be added to the theory of *truer* sketched here. It might be regretted that we don't have a *distinctive* story about the Sorites in terms of *truer*. But the hypothesis that some sentences are truer than others is basically a *semantic* hypothesis, and if the reason Sorites premises look attractive is anything like the reason (5) looks *prima facie* attractive, then that attractiveness should receive a *pragmatic* explanation. What is really important is that there be some story about the Sorites we can tell.

10. Linearity Intuitions

The following three objections all look to me to stem from the same source: the intuition that *truer* is a linear relation. (All of these objections arose in the discussion of the paper at BSPC.)

True and Truer (due to Cian Dorr)

Here's an odd consequence of the theory I have sketched above. (Or at least the theory I've sketched above plus the conditional *If S then S is true*, which I think always holds around here, whatever qualifications we have to put on it to deal with the semantic paradoxes.) We can't infer from A is true and B is false that A is truer than B. But this looks like a reasonably plausible inference.

If we simply added this as another inference rule, we would be back with epistemicism (or something like it.) Using this inference rule we can prove that there are no sentences that take intermediate truth values. Assume, for *reductio*, that A is such a sentence. Since we keep classical logic, we know $A \vee \neg A$ is true. If A, then A is true, and hence $\neg A$ is false. By the inference rule under consideration, it follows that $A \geq_T \neg A$. From this it follows that $0{=}1 \geq_T \neg A$, since $0{=}1 \geq_T A \wedge \neg A$, and whenever $B \geq_T C$, $B \wedge C \geq_T C$. And from $0{=}1 \geq_T \neg A$ it follows that $A \geq_T 0{=}0$, i.e. that A is determinately true. A converse proof shows that if $\neg A$, then $\neg A$ is determinately true and A is determinately false. Since either disjunct implies A is not indeterminate, and we have ($\vee$-Out) it follows that there are no indeterminate truths. So we do not have the option of simply adding this rule.

Comparing Negative and Positive (due to Jonathan Schaffer)

Let *a* be a regular borderline case of genius, somewhere near the middle of the penumbra. Let *b* be someone who is not a determinate case of genius, but is very close. Let *A* be *a is a genius* and *B* be *b is a genius*. Given the facts, it seems plausible that $A \geq_T \neg B$, since *a* is right around the middle of the borderline cases of genius, but *b* is only a smidgen short of clear genius. But since *b* is closer to being a genius than *a*, we definitely have $B \geq_T A$. By transitivity, it follows that $B \geq_T \neg B$, and hence $B \wedge \neg B \geq_T \neg B$. But since $\neg B$ is not determinately false, it follows that $B \wedge \neg B$ is not determinately false, contradicting (A7).

Since I accept (A7) it follows that I must reject the assumptions of this line of reasoning. In particular, I reject the initial premise that $A \geq_T \neg B$. But it's worth noting that this case is quite general. Similar reasoning could be used to show that for any indeterminate propositions of the form *x is a genius* and *y is not a genius*, the first is not truer than the second. This seems odd, since intuitively these could both be indeterminate while the first is very nearly true and the second very nearly false.

Comparing Different Predicates (due to Elizabeth Harman)

Earlier I said a good way to understand *truer* is that *A* is truer than *B* iff *A* is true on every precisification on which *B* is true and the converse does not hold. If we assume that precisifications of predicates from different subject areas (e.g. *hexagonal* and *honest*) are independent, it follows that subject-predicate sentences involving those predicates and indeterminate instances of them are incomparable with respect to truth. But this seems implausible. If France is a borderline case of being hexagonal that is close to the lower bound, and George Washington is a borderline case of being honest who is close to the upper bound, then we should think *George Washington is honest* is truer than *France is hexagonal.*

Now at first this objection might look less pressing than the previous two objections. We cannot consistently add the principle *If A is true and B is false then A is truer than B*, nor the principle *For some indeterminate sentences a is F and b is not F, a is F is truer than b is not F*, without contradicting some principles previously adopted. On the other hand, there's nothing inconsistent about dropping the connection with supervaluationism and saying that *George Washington is honest* is truer than *France is hexagonal.* But this move has at least one significant cost. If *George Washington is honest* is truer than *France is hexagonal*, then *If France is hexagonal then George Washington is honest* is determinately true. This would be an odd penumbral connection, but perhaps one we can live with. It's noteworthy that when the predicates are somehow connected such conditionals seem rather natural, as in *If France is hexagonal then Brazil is triangular.* But this is not to minimise the counterintuitive aspects of the theory around here. It just doesn't seem true that intuitions about whether *A* is truer than *B* fit together nicely with intuitions about whether *If B then A* is determinately true.

All three of these objections seem to me to turn on an underlying intuition that *truer* should be a linear relation. If we are given this, then the inference principle Dorr suggests looks unimpeachable, and the comparisons Schaffer

and Harman suggested look right. But once we drop the idea that truer is linear, I think the plausibility of these claims falls away. And there are two good reasons to drop linearity.

One is that we can't simultaneously accept all of the following five principles.

- *Truer* is a linear relation.
- (A2), that conjunction is a greatest lower bound.
- (A4), that negation is order inverting.
- (A7), that contradictions are determinately false.
- There are indeterminate sentences.

I think by far the least plausible of these is the first, so it must go.

Another reason is that linearity makes it very difficult to tell a plausible story about higher order vagueness. (Here we see that the false precision problem for M was not due entirely to its misguided use of numerical truth values. The structure of M already leads to something like the false precision problem before the numbers are introduced.) Linearity is the claim that for any two sentences A and B, the following disjunction holds. Either $A >_T B$, or $B >_T A$, or $A =_T B$. Presumably if we are saying that *truer* is linear, then we are committed to that being determinately true. But if *truer* is determinately linear, then that disjunction must be determinately true for all A, B. And if *truer* is linear, and if that disjunction is determinately true, then one of its disjuncts must be determinately true, for linearity rules out the possibility of a determinately true disjunction with no determinately true disjunct. Now take a special case of that disjunction, where B is 0=0. In that case we can rule out $A >_T B$. So the only options are $B >_T A$ or $A =_T B$. We have concluded that given linearity, one of these disjuncts must be determinately true. But if A is determinately true, but not determinately determinately true, as it seems plausible that A could be, then neither of these should be determinately true. So linearity has led to a contradiction. This suggests that combining (determinate) linearity with the most natural account of how higher-order vagueness is to be expressed in terms of *truer* leads to a contradiction. Possibly a determined linearity theorist could find some other account of higher-order vagueness that didn't have this difficulty, but I think the prospects are grim.

So I think we should say the following thing about these three objections. All of the objections really do rest on intuitions that have a degree of plausibility. But their plausibility flows from their foundation in an untenable theory: the theory that *truer* is a linear relation. Once we see that theory is mistaken, as these arguments show, we should simply resist the intuitions that the theory underwrites.

Here's one move we shouldn't make, as attractive as it might be here. There are really two concepts *truer than*, a non-linear one governed by (A1) to (A10), and a linear one that validates the claims here. If we believe these two concepts exist, then we have no reason to think that the methods listed in section five for isolating the concept *truer* really isolate the non-linear one I want to focus on. If we want to be confident that those methods do isolate the preferred concept, it is much better to say that there is one concept here, and I disagree with the proponent of M about whether it picks out a linear relation. This is the most natural thing to say in any case. There

is no reason to think that if two people are disagreeing about whether *more intelligent than* is a linear relation, the most natural thing to say is that they disagree about the logical properties of a particular relation, not that they are simply picking out different concepts by their common words. That disagreement is a useful one to focus on because it's probably an important disagreement, since if *more intelligent than* is non-linear, that puts an upper bound on our ability to accurately rank people for intelligence through the use of standardised tests and the like. And that in turn reveals something about our tendency to see comparatives as linear orderings. *More intelligent than* is probably the most plausible candidate out there for a non-linear comparative, yet there is a huge industry of intelligence testing that presupposes it really is linear, and that presupposition is rarely questioned even by critics of that industry. It requires quite a change to our regular world-view to not presuppose that comparatives generate a linear order, and it seems to me that when we make this change the intuitions behind these objections start to look less plausible.

11. Benefits of the Truer Theory

The new theory I want to propose is now before you. I argue that there is a *truer* operator, and that indeterminate sentences of English are less true than perfectly true sentences, and more true than perfectly false sentences. If x is a borderline case of F-ness, then as x gets more and more F-ish, *a is F*, where a is a name for x, becomes truer and truer. All classical tautologies are perfectly true, and all classical inference rules are admissible, given a natural definition of entailment in terms of *truer*. Higher-order vagueness is allowed for by the model, and our theory of higher-order vagueness is produced simply by iterating our theory of first-order vagueness. To close, let me just list the advantages the truer theory has over the other theories.

- Since we do not appeal to truth values, but only to a comparative truer than, we do not have the problem of false precision that bothers M.
- Unlike M, we do not assume that any two sentences are comparable with respect to how true they are.
- Unlike M, and supervaluationism, our theory of higher-order vagueness can be generated by just iterating our theory of first-order vagueness.
- Unlike M, contradictions are always perfectly false in our theory.
- The concept *truer*, which does much of the theoretical work here, is much less mysterious, and much more intuitive, than the concept of an acceptable precisification which does the equivalent work in supervaluational theory.
- Unlike in supervaluational theory, we do not have to 'turn the theory off' when talking about vague sentences.
- Unlike the dominant version of supervaluationism, we get to keep all classical inference rules.
- Unlike epistemicism, we get to acknowledge that there is such a thing as semantic indeterminacy.

References

Braun, David and Theodore Sider (ms.) "Vague, so Untrue".

Burgess, John (2001) "Vagueness, Epistemicism and Response-Dependence" *Australasian Journal of Philosophy* 79: 507-24.

Burgess, John and I. L. Humberstone, (1987) "Natural Deduction Rules for a Logic of Vagueness" *Erkenntnis* 27: 197-229.

Cook, Roy (2002) "Vagueness and Mathematical Precision" *Mind* 111: 225-48.

Dummett, Michael (1959) "Truth" *Proceedings of the Aristotelian Society. New Series* 59: 141-62.

Eklund, Matti (2002) "Inconsistent Languages" *Philosophy and Phenomenological Research* 64: 251-75.

Fodor, Jerry (1987) *Psychosemantics*. Cambridge, MA: MIT Press.

Fodor, Jerry and Ernest Lepore (1996) "What Cannot be Valuated Cannot be Valuated, and it Cannot be Supervaluated Either" *Journal of Philosophy* 94: 516-35.

Graff, Delia (2000) "Shifting Sands: An Interest-Relative Theory of Vagueness" *Philosophical Topics* 28: 45-81.

Greenough, Patrick (2003) "Vagueness: A Minimal Account" *Mind* 112: 235-81.

Hyde, Dominic (1997) "From Heaps and Gaps to Heaps of Gluts" *Mind* 106: 641-60.

Kaplan, David and Richard Montague (1960) "A Paradox Regained" *Notre Dame Journal of Formal Logic* 1: 79-90.

Keefe, Rosanna (2000) *Theories of Vagueness*. Cambridge: Cambridge University Press.

Klein, Ewan (1980) "A Semantics for Positive and Comparative Adjectives" *Linguistics and Philosophy* 4:1-45.

Levinson, Stephen (2000) *Presumptive Meanings*. Cambridge, MA: MIT Press.

Lewis, David (1972) "Psychophysical and theoretical identifications" *Australasian Journal of Philosophy* 50: 249-58

Lewis, David (1983) "New Work for a Theory of Universals" *Australasian Journal of Philosophy* 61: 343-77.

Lewis, David (1993) "Many, but Almost One" in *Ontology, Causality and Mind: Essays in Honour of D M Armstrong*, Bacon, John (ed) New York: Cambridge University Press.

McGee, Vann and Brian McLaughlin (1995) "Distinctions Without a Difference" *Southern Journal of Philosophy* 33 (Supp): 203–51.

Merricks, Trenton (2001) "Varieties of Vagueness" *Philosophy and Phenomenological Research* 62: 145-57.

Priest, Graham (2001) *Introduction to Non-Classical Logics*. Oxford: Oxford University Press.

Sedivy, J., M. Tanenhaus, C. Chambers and G. Carlson (1999) "Achieving incremental semantic interpretation through contextual representation" *Cognition*, 71: 109-47.

Sider, Theodore (2001) "Personal Identity and the Limits of Conceptual Analysis" *Philosophical Perspectives* 15: 189-209.

Soames, Scott (1999) *Understanding Truth*. New York: Oxford University Press.

Sorensen, Roy (2001) *Vagueness and Contradiction*. Oxford: Oxford University Press.

Tappenden, Jamie (1993) "The Liar and Sorites Paradoxes; Toward a Unified Treatment" *Journal of Philosophy* 90: 551-77.

Weatherson, Brian (2003a) "Many Many Problems" *Philosophical Quarterly* 53: 481-501.

Weatherson, Brian (2003b) "Epistemicism, Parasites and Vague Names" *Australasian Journal of Philosophy* 81: 276-9.

Williamson, Timothy (1994) *Vagueness*. London: Routledge.

Williamson, Timothy (1995) "Definiteness and Knowability" *Southern Journal of Philosophy* 33 (Supp) 171-91.