

True, Truer, Truest

Brian Weatherson

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 to 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 mapping it onto some mathematical objects. 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 classical inference rules as well as 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. I assume throughout familiarity with popular theories of vagueness.

1 Truer

The core of my theory is that some sentences involving vague terms are truer than others. I won't give an analysis of *truer*, instead I will argue that we already tacitly understand this relation. The main argument for this will turn on a consideration of two 'many-valued' theories of vagueness, one of which will play a central role (as the primary villain) in what follows.

The most familiar many-valued theory, call it M , says there are continuum many truth values, and they can be felicitously represented by the interval $[0, 1]$. The four main logical connectives: *and*, *or*, *if* and *not* are truth-functional. The functions are:

$$V(A \wedge B) = \min(V(A), V(B))$$

$$V(A \vee B) = \max(V(A), V(B))$$

$$V(A \rightarrow B) = \max(1, 1 - V(A) + V(B))$$

$$V(\neg A) = 1 - V(A)$$

[†] Penultimate draft only. Please cite published version if possible. Final version published in *Philosophical Studies* 123 (2005): 47-70. Thanks to audiences at the 2002 APA Central, Edinburgh and especially the 2003 BSPC for helpful comments. From the latter I'm especially grateful to Jonathan Bennett, Alex Byrne, Cian Dorr, Andy Egan, Elizabeth Harman, Robin Jeshion, Mike Nelson, Jonathan Schaffer, Ted Sider and Gabriel Uzquiano. I'm also grateful to two very helpful referee reports.

where V is the valuation function on sentences, $\min(x, y)$ is the smaller of x and y and $\max(x, y)$ is the larger of x and y .

Adopting these rules for the connectives commits us to adopting the logic \mathbb{L}_C . M is the theory that this semantic model, under its most natural interpretation, is appropriate for vague natural languages. (We'll discuss less natural interpretations presently.)

M tells a particularly nice story about the Sorites. A premise like *If she's rich, someone with just a little less money is also rich* will have a very high truth value. If we make the difference in money between the two subjects small enough, this conditional will have a truth value arbitrarily close to 1.

M also tells a nice story about borderline cases and determinateness. An object a is a borderline case of being an F just in case the sentence *a is F* has a truth value between 0 and 1 exclusive. Similarly, a is a determinate F just in case the truth value of *a is F* is 1. (It is worthwhile comparing how simple this analysis of determinateness is to the difficulties supervaluationists have in providing an analysis of determinateness. On this topic, see Williamson (1995), McGee and McLaughlin (1998) and Williamson (2004).)

But M tells a particularly implausible story about contradictions. Here is how Timothy Williamson (1994) makes this problem vivid.

More disturbing is that the law of non-contradiction fails ... $\neg(p \wedge \neg p)$ always has the same degree of truth as $p \vee \neg p$, and thus is perfectly true only when p is either perfectly true or perfectly false. When p is half-true, so are both $p \wedge \neg p$ and $\neg(p \wedge \neg p)$. (Williamson, 1994, 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? (Williamson, 1994, 136)

There is a way to keep the semantic engine behind M while avoiding this consequence. (The following few paragraphs are indebted pretty heavily to the criticisms of Strawson's theory of descriptions in Dummett (1959))

Consider an interpretation of the above semantics on which there are only two truth values: True and False. Any sentence that gets truth value 1 is true, all the others are false. The numbers in $[0, 1)$ represent different ways of being false. (As Tolstoy might have put it, all true sentences are alike, but every false sentence is false in its own unique way.) Which way a sentence is false can affect 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. For example, it agrees about what the axioms should be for a logic for natural language.

But it has several philosophical differences. In particular, it has none of the three characteristics of M we noted above.

It cannot tell as plausible story as M does about the Sorites. 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.

On the other hand, according to M_D , contradictions are always false. To be sure, a contradiction might be false in some obscure new way, but it is still false. Recall Williamson's objection that an explicit contradiction should be true to degree 0 and nothing more. This objection only works 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, since he intended it as an objection to M , but just to say that re-interpreting the semantic significance of the numbers in M makes a philosophical difference.

Despite M_D 's preferable treatment of contradictions, I think M is overall a better theory because it has a much better account of borderline cases. But for now I want to stress a simpler point: M and M_D are *different* theories 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 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.

I have implicitly defined *truer* by noting its theoretical role. As David Lewis (1972) showed, terms can be implicitly defined by their theoretical role. There is one unfortunate twist here in that *truer* is defined by its role in a *false* theory, but that does not block the implicit definition story. We know what *phlogiston* and *ether* mean because of their role in some false theories. The meaning of *truer* can be extracted in the same way from the false theory M .

2 Further Reflections on *Truer*

As noted, I won't give a reductive analysis of *truer*. The hopes for doing that are no better than the hopes of giving a reductive analysis of *true*. But I will show that we pre-theoretically understand the concept.

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, 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. Not every predicate can be turned into a comparative. But most can, and our default assumption should be that *true* is like the majority.

I have heard two arguments against that assumption. First, it could be argued that most comparatives in English generate linear orderings, but *truer* 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, and they seem fairly indicative of large classes. Second, it could be argued that it's crucial to understanding comparatives that we understand the interaction of the underlying adjectives with comparison classes. Robin Jeshion and Mike Nelson made this objection in their comments on my paper at BSPC 2003. Again, the premise is not obviously true. 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.) *Straight* and *round* either don't have or don't need comparison classes, but they form comparatives. So *true*, which also does not take comparison classes, could also form a comparative.

Finally, if understanding the inferential role of a logical operator helps know its meaning, then it is notable that *truer* has a very clear inferential role. It is the same as a strict material implication $\Box(q \supset p)$ defined using a necessity operator whose logic is KT. Since many operators have just this logic, this doesn't individuate *truer*, but it helps with inferential role semantics aficionados.

I claim that the concept *truer*, and the associated concept *as true as*, are the only theoretical tools we need to provide a complete theory of vagueness. It is simplest to state the important features of my theory by contrasting it with *M*. I keep the following good features of *M*.

- G1** There are *intermediate* sentences, i.e. sentences that are truer than some sentences and less true than others. For definiteness, I will say *S* is intermediate iff *S* is truer than 0=1 and less true than 0=0.
- G2** *a* is a borderline *F* iff *a is F* is intermediate, and *a* is determinately *F* iff *a is F* and *a* is not a borderline *F*.

I won't repeat the arguments here, but I take G1 to be a large advantage of theories like *M* over epistemicist theories. (See Burgess (2001), Sider (2001) and Weatherson (2003a) for more detailed arguments to this effect.) And as noted G2 is a much simpler analysis of determinacy and borderline than supervaluationists have been able to offer.

I drop the following bad features of *M*.

B1 *Some contradictions are intermediate sentences.*

On my theory all contradictions are determinately false, and determinately determinately false, and so on. The argument for this has been given above.

B2 *Some classical tautologies are intermediate sentences.*

On my theory all classical tautologies are determinately true, and determinately determinately true, and so on. We will note three arguments for this being an improvement in the next section.

B3 Some classical inference rules are inadmissible.

On my theory all classical inference rules are admissible. As Williamson (1994) showed, the most prominent version of supervaluationism is like M in ruling some classical rules to be inadmissible, and this is clearly a cost of those theories.

B4 Sentences of the form S is intermediate are never intermediate

I will argue below this is a consequence of M , and it means it is impossible to provide a plausible theory of higher-order vagueness within M . In my theory we can say that there is higher-order vagueness by treating *truer* as an iterable operator, so we can say that S is *intermediate* is intermediate. If S is a is F , that's equivalent to saying that a is a borderline case of a borderline case of an F . Essentially we get our theory of higher-order vagueness by simply iterating our theory of first-order vagueness, which is what Williamson does in his justly celebrated treatment of higher-order vagueness. Note it's not just M that has troubles with higher-order vagueness. See Williamson (1994) and Weatherson (2003b) for the difficulties supervaluationists have with higher-order vagueness. The treatment of higher-order vagueness here is a substantial advantage of my theory over supervaluationism.

B5 Truer is a linear relation.

On my theory it need not be the case that S_1 is truer than S_2 , or S_2 is truer than S_1 , or they are as true as each other. In the last section I will argue that this is a substantial advantage of my theory. I claim that *truer* generates a Boolean lattice on possible sentences of the language. (For a familiar example of a Boolean lattice, think of the subsets of \mathbb{R} ordered by the subset relation.)

I also provide a very different, and much more general, treatment of the Sorites than is available within M . The biggest technical difference between my theory and M concerns the relationship between the semantics and the logic. In M the logic falls out from the truth-tables. Since I do not have the concept of an intermediate truth value in my theory, I could not provide anything like a truth-table. Instead I posit several constraints on the interaction of *truer* with familiar connectives, posit an analysis of validity in terms of *truer*, and note that those two posits imply that all and only classically admissible inference rules are admissible.

3 Constraints on Truer and Classical Logic

The following ten constraints on *truer* seem intuitively compelling. I've listed here both the philosophically important informal claim, and the formal interpretation of that claim. (I use $A \geq_T B$ as shorthand for A is at least as true as B . Note all the quantifiers over sentences here are *possibilist* quantifiers, we quantify over all possible sentences in the language.)

(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 $C \geq_T B$
- (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
 $A \wedge \neg A =_T S_F$
- (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 is a 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 $A \geq_T \phi n$ and $\exists x(\phi x) \geq_T A$ 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
1. $B \rightarrow A \geq_T S_T$ iff $A \geq_T B$
 2. $(A \wedge B) \rightarrow C =_T A \rightarrow (B \rightarrow C)$

Apart from (A10) these are fairly straightforward. We can't argue for (A10) by saying English *if...then* is a material implication, because that leads directly to the paradoxes of material implication. Assuming that $\neg A \vee B$ is a material implication is equivalent to assuming (inter alia) that $A \vee \neg A$ is perfectly true. I believe this, but since it is denied by many I want that to be a conclusion, not a premise. So the argument for (A10) must be a little indirect. In particular, we will appeal to the behaviour of quantifiers. We can formally represent *All Fs are Gs* in two ways: using restricted or unrestricted quantifiers. In the first case the formal representation will look like:

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

with some connective in place of '?'. But it seems clear that whatever connective goes in there must be a material implication. In the second case, the formal representation will look like:

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

In that case, we can define a connective ∇ that satisfies the definition of a material implication:

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

This is equivalent to the odd (but intelligible) sentence *Everything such that A is such that B*. Again, considerations about what should be logical truths involving quantifiers suggests that ∇ must be a material implication. So either way there should be a material implication present in the language, as (A10) says.

Given (A1) to (A10) it follows that this material implication is equivalent to $\neg A \vee B$, and hence $A \vee \neg A$ is a logical truth. This is a surprising *conclusion*, since intuitively vagueness poses problems for excluded middle, but I think it is more plausible that vague instances of excluded middle are problematic for *pragmatic* reasons than that any of (A1) to (A10) are false.

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, i.e. iff it is impossible for the conjunction of the premises to be truer than the conclusion. Given that, we get:

$$\forall A_1, \dots, A_n, B: A_1, \dots, A_n \vdash_T B \text{ iff, according to classical logic, } A_1, \dots, A_n \vdash B$$

(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. Indeed, Boolean lattices are models for classical logic in the strong sense that classical inference rules, such as conditional proof and *reductio*, are admissible in logics defined on them, so we also get the admissibility of classical inference rules in this theory. (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 both directions.)

There are three reasons for wanting to keep classical logic in a theory of vagueness. First, as Williamson has stressed, classical logic is at the heart of many successful research programs. Second, non-classical theories of vagueness tend to abandon too much of classical logic. For instance, *M* abandons the very plausible schema $(A \wedge A \rightarrow B) \rightarrow B$. The third reason is the one given here - these ten independently plausible constraints on *truer* entail that the logic for a language containing *truer* should be classical. These three arguments add up to a powerful case that non-classical theories like *M* are mistaken, and we should prefer a theory that preserved classical logic.

4 Semantics and Proof Theory

In this section I will describe a semantics and proof theory for a language containing *truer* as an iterable operator. This is important for the theory of higher-order vagueness. I say that a is a borderline borderline F just in case the sentence a is a borderline F is intermediate, where ‘borderline’ is analysed as in section 2. It might not be obvious that it is consistent with (A1) to (A10) that any sentence a is a borderline F could be consistent. One virtue of the model theory outlined here is that it shows this is consistent.

For comparison, note that M as it stands has no way of dealing with higher-order vagueness, i.e. with borderline cases of borderline cases of F -ness. If every sentence a is a borderline F either does or does not receive an integer truth value, then this intuitive possibility is ruled out. We cannot solve the problem simply by iterating M . (This is a point stressed by (Williamson, 1994, Ch. 4).) We cannot say that it is true to degree 0.5 that (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. So the model theory will show that our theory is substantially better than M in this respect.

Consider the following (minor) variant on KT. Vary the syntax so $\Box A$ is only well-formed if A is of the form $B \rightarrow C$. Call the resulting logic KT_R , with the R indicating the syntactic restriction. The restriction makes very little difference. Since A is equivalent to $(A \rightarrow A) \rightarrow A$, even if $\Box A$ is not well-formed in KT_R , the KT-equivalent sentence $\Box((A \rightarrow A) \rightarrow A)$ will be well-formed. 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. (There is no restriction on the accessibility relation other than reflexivity.)

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 an axiomatic proof theory. 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 S of sentences with the following three properties:

- All theorems of KT_R are in S .

- For all A , either A is in S or $\neg A$ is in S .
- S 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 $\langle W, R, V \rangle$ where

- W is the set of maximal consistent sets for KT_R
- R is a subset of $W \times W$ 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 contains \geq_T but not \Box . Just replace $\Box(B \rightarrow A)$ with $A \geq_T B$ wherever \Box occurs including inside sentences. (We appeal here and here alone to the restriction in KT_R .) 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 If $A \rightarrow B$ is a theorem and A is a theorem, then B is a theorem.

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

A Kripke model for this logic is just a Kripke model for KT , except we say $B \geq_T A$ is true at a point iff B is true at all accessible points at which A is true. This leads to a semantic definition of validity. An argument is valid iff it preserves truth at any point in all such models.

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 omitted. The maximal consistent sets for \geq_T , i.e. the points in the canonical model, just are the results of applying the translation rule $\Box(B \rightarrow A) \Rightarrow A \geq_T B$ to the (sentences in the) maximal consistent sets for KT_R .

That's important because the points in the canonical model for \geq_T are useful for understanding the relationship between *truer* and *true*, and for understanding what languages are. 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. For many purposes it is useful to think of the theory based on *truer* as a variant on M . But considering the canonical model for \geq_T 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 my theory and supervaluationism. I identify languages with a single point rather than with a set of points, which leads to the smoother treatment of higher-order vagueness on my account. Also, I don't start with a set of acceptable points/precisifications. The canonical model contains all the points that are formally consistent, and I identify particular languages, like English, by vaguely saying that the point that represents English is (roughly) there. (Imagine my vaguely pointing at some part of 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, but they all favour my theory over supervaluationism.

To return to the point about higher order vagueness, note that all of the following sentences are consistent in KT, and hence their 'equivalents' using $>_T$ are also consistent.

- (1) $\neg\Box p \wedge \neg\Box\neg p$
 $0=0 >_T p >_T 0=1$
- (2) $\neg\Box(\neg\Box p \wedge \neg\Box\neg p) \wedge \neg\Box\neg(\neg\Box p \wedge \neg\Box\neg p)$
 $0=0 >_T (0=0 >_T p >_T 0=1) >_T 0=1$
- (3) $\neg\Box(\neg\Box(\neg\Box p \wedge \neg\Box\neg p) \wedge \neg\Box\neg(\neg\Box p \wedge \neg\Box\neg p))$
 $\wedge \neg\Box\neg(\neg\Box(\neg\Box p \wedge \neg\Box\neg p) \wedge \neg\Box\neg(\neg\Box p \wedge \neg\Box\neg p))$
 $0=0 >_T (0=0 >_T (0=0 >_T p >_T 0=1) >_T 0=1) >_T 0=1$

And obviously this pattern can be extended indefinitely. In general, any claim of the form that a is an n -th order borderline case of an F is consistent in this theory, as can be seen by comparison with KT.

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 proves my earlier claim that I can keep all inference rules of classical logic. Second, it helps justify (A1) to (A10). Most rules correspond directly to one of the constraints. For that reason I've set all the rules, even though you've probably seen most of them before.

- (\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) $\Gamma \vdash A \geq_T B \Rightarrow \Gamma \vdash B \rightarrow A$

(Thanks to Gabriel Uzquiano for several probing questions that led to this section being written.)

5 Sexy Sorites

A good theory of vagueness should tell us two things about the Sorites. The easy part is to say what is wrong with Sorites arguments: not all premises are perfectly true. The hard part is to say why the premises looked plausible to start with. The M theorist has the beginnings of a story, though not the end of a story. The beginning is that all the premises in a typical Sorites argument are nearly true, and they look plausible 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.

The story the M theorist tells 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. According to M , every premise like (SI) is almost true.

(SI) If a is tall, then a' is tall.

But we could have built a Sorites argument with premises like (SA).

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

And premises of this form are not, in general, almost true. Indeed, some will have a truth value not much about 0.5. So M has no explanation for why premises like (SA) look persuasive. This is quite bad, because (SA) is *more* plausible than (SI) as I'll now show. 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), how do you persuade them? My first instinct is to appeal to something like (SA). If that doesn't work, I appeal to theoretical considerations about how our use of *tall* couldn't possibly pick a boundary

between a and a' . I think I find (SI) plausible *because* I find (SA) plausible, and I would try to get the students to feel likewise. There's an asymmetry here. I wouldn't defend (SA) by appealing to (SI), and I don't find (SA) plausible because it follows 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.)

Sadly, many theories have little to say 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. The supervaluationist story about (SA) is no better.

Here's a surprising fact about the Sorites that puts an unexpected constraint on explanations of why (SA) is plausible. In the history of debates about it, I don't think anyone has put forward a Sorites argument where the major premises are like (SO).

(SO) Either a is not tall, or a' is tall.

(This point is also noticed in Braun and Sider (2007).) 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 hardly any theory of vagueness (certainly not M or supervaluationism or epistemicism) provide grounds for distinguishing (SA) from (SO), since most theories of vagueness endorse 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. Fara (2000); Sorensen (2001); Eklund (2002)) seem to do any better at distinguishing (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.

6 Linearity Intuitions

The assumption that *truer* is a non-linear relation is the basis for most of the distinctive features of my theory, so it should be defended. There are two reasons to believe it.

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.

Linearity (or at least determinate linearity) also makes it difficult to tell a plausible story about higher order vagueness. 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 $A =_T B$. If *truer* is determinately linear, that disjunction is determinately true. 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. That is, A is either determinately intermediate or determinately determinate. But intuitively neither of these need be true, for A might be in the 'penumbra' between the determinately intermediate and the determinately determinate. This argument is only a problem if we assume determinate linearity, but it's hard to see the theoretical motivation for believing in linearity but not determinate linearity.

Still, it is very easy to believe in linearity. Even for comparatives that are clearly non-linear, like *more intelligent than*, there is a strong temptation to treat them as linear. (Numerical measurements of intelligence are obviously inappropriate given that *more intelligent than* is non-linear, but there's a large industry involved in producing such measurements.) And this temptation leads to some prima facie plausible objections to my theory. (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 my theory plus the plausible assumption that *If S then S is true* is axiomatic. 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 added this as inference rule, we would rule out all intermediate sentences. To prove this assume, for *reductio*, that A is intermediate. 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. Then the new inference rule implies $A \geq_T \neg A$, hence $A \wedge \neg A \geq_T \neg A$, since $\neg A \geq_T \neg A$, and hence $0=1 \geq_T \neg A$, since $0=1 \geq_T A \wedge \neg A$, and \geq_T is transitive. So A is determinately

true, not intermediate. A converse proof shows that if $\neg A$, then A is determinately false, not intermediate. So by (\vee -Out) it follows that A is not intermediate, but since A was arbitrary, there are no intermediate truths. So this rule is unacceptable, despite its plausibility.

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*. 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 $B \geq_T \neg B$, hence B is determinately true (by the reasoning of the last paragraph). Since $\neg B$ is not determinately false, it follows that $B \wedge \neg B$ is not determinately false, contradicting (A7).

Since I accept (A7) I must reject the initial assumption 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)

One intuitive way to understand the behaviour of *truer* is that A is truer than B iff A is true on every admissible precisification on which B is true and the converse does not hold. This can't be an analysis of *truer*, since it assumes we can independently define what is an admissible precisification, and this seems impossible. But it's a useful heuristic. And reflecting on it brings up a surprising consequence of my theory. 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*.

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. So the arguments against linearity are ipso facto arguments that we should simply drop the intuitions Dorr, Schaffer and Harman are relying upon.

To conclude, it's worth noting that a very similar inferential rule to the rule Dorr suggests is admissible. From the fact that A is determinately true, and B is determinately false, it follows that A is truer than B . If we assume, as seems reasonable,

that we're only in a position to say that A is true when it is determinately true, then whenever we're in a position to say A is true and B is false, it will be true that A is truer than B . This line of defence is obviously similar to the explanation I gave in the previous section of why Sorites premises look plausible, and to the argument Rosanna Keefe gives that the failure of classical inference rules is no difficulty for supervenience because it admits very similar inference rules (Keefe, 2000).

References

- Braun, David and Sider, Theodore, (2007). "Vague, So Untrue." *Noûs* 41: 133-156, doi:10.1111/j.1468-0068.2007.00641.x. (12)
- Burgess, John, (2001). "Vagueness, Epistemicism and Response-Dependence." *Australasian Journal of Philosophy* 79: 507-524, doi:10.1080/713659306. (4)
- Burgess, J. A. and Humberstone, I. L., (1987). "Natural Deduction Rules for a Logic of Vagueness." *Erkenntnis* 27: 197-229, doi:10.1007/bf00175369. (12)
- Dummett, Michael, (1959). "Truth." *Proceedings of the Aristotelian Society* 59: 141-162, doi:10.1093/aristotelian/59.1.141. (2)
- Eklund, Matti, (2002). "Inconsistent Languages." *Philosophy and Phenomenological Research* 64: 251-275, doi:10.1111/j.1933-1592.2002.tb00001.x. (12)
- Fara, Delia Graff, (2000). "Shifting Sands: An Interest-Relative Theory of Vagueness." *Philosophical Topics* 28: 45-81., doi:10.5840/philtopics20002816. This paper was published under the name 'Delia Graff'. (12)
- Fine, Kit, (1975). "Vagueness, Truth and Logic." *Synthese* 30: 265-300, doi:10.1007/bf00485047. (12)
- Goldblatt, Robert, (1992). *Logics of Time and Computation*. Palo Alto: CSLI. (8)
- Greenough, Patrick, (2003). "Vagueness: A Minimal Theory." *Mind* 112: 235-281, doi:10.1093/mind/112.446.235. (13)
- Grice, H. Paul, (1989). *Studies in the Way of Words*. Cambridge, MA.: Harvard University Press. (13)
- Keefe, Rosanna, (2000). *Theories of Vagueness*. Cambridge: Cambridge University Press. (1, 13, 16)
- Levinson, Stephen, (2000). *Presumptive Meanings*. Cambridge, MA: MIT Press. (13)
- Lewis, David, (1972). "Psychophysical and Theoretical Identifications." *Australasian Journal of Philosophy* 50: 249-258, doi:10.1080/00048407212341301. Reprinted in *Papers in Metaphysics and Epistemology*, pp. 248-261. (3)
- McGee, Vann and McLaughlin, Brian, (1998). "Review of Timothy Williamson's *Vagueness*." *Linguistics and Philosophy* 21: 221-231. (2)

- Sedivy, Julie, Tanenhaus, Michael., Chambers, Craig, and Carlson, Gregory, (1999). "Achieving incremental semantic interpretation through contextual representation." *Cognition* 71: 109-147, doi:10.1016/s0010-0277(99)00025-6. (13)
- Sider, Theodore, (2001). "Maximality and Intrinsic Properties." *Philosophy and Phenomenological Research* 63: 357-364, doi:10.1111/j.1933-1592.2001.tb00109.x. (4)
- Sorensen, Roy, (2001). *Vagueness and Contradiction*. Oxford: Oxford University Press. (12)
- Tappenden, Jamie, (1993). "The Liar and Sorites Paradoxes: Toward a Unified Treatment." *Journal of Philosophy* 90: 551-577, doi:10.2307/2940846. (12)
- Weatherson, Brian, (2003). "Epistemicism, Parasites, and Vague Names." *Australasian Journal of Philosophy* 81: 276-279, doi:10.1093/ajp/jag209. (4)
- , (2003). "Review of Rosanna Keefe, *Theories of Vagueness*." *Philosophy and Phenomenological Research* 67: 491-494. (5)
- Williamson, Timothy, (1994). *Vagueness*. Routledge. (1, 2, 5, 8)
- , (1995). "Definiteness and Knowability." *Southern Journal of Philosophy* 33 (Supp): 171-191, doi:10.1111/j.2041-6962.1995.tb00769.x. (2)
- , (2004). "Reply to McGee and McLaughlin." *Linguistics and Philosophy* 27: 113-122, doi:10.1023/b:ling.0000010847.78827.d0. (2)