Supervaluationism

Cian Dorr

September 21, 2005

1 The law of the excluded middle

Intuitive counterexamples to the rule that a conjunction of borderline conjuncts is always borderline (rather than definitely false): 'The patch is both red and pink'; 'Nothing is both a bachelor and married'; 'This item is both a heap and a non-heap'.

Assume we want to hold onto De Morgan's Laws and Double Negation Elimination. Then if we accept these sentences as true (and not borderline), we'll also have to accept counterexamples to the rule that a disjunction of borderline disjuncts is always borderline (rather than definitely true): 'The patch is either not red or not pink'; 'Everything is either a non-bachelor or unmarried'; 'This item is either a heap or a non-heap'.

By one of De Morgan's Laws and Double Negation Elimination, every instance of the Law of the Excluded Middle ('Either *P* or not-*P*') is equivalent to an instance of the Law of Non-Contradiction ('Not (*P* and not-*P*'). So if all instances of the latter law are true, so are all instances of the former.

- If 'either the patch is red or the patch isn't red' is literally true, we need some other explanation of why this seems like such a bad thing to assert (when the patch in question is borderline red). An explanation in terms of *pragmatics* rather than *semantics*.
- My idea: this is a bad sentence to assert because it so naturally suggests the unanswerable question 'Which is it?'—unanswerable in the sense that neither the answer 'red' nor the answer 'not red' would be (definitely) true.

LEM is the single bit of classical logic that seems most threatened specifically by vagueness. So if we can reconcile ourselves to LEM, it isn't a big further step to the claim that the set of (definitely) true sentences is closed under all classically valid inferences.

2 Excluded Middle and the sorites

Given Excluded Middle, we can mount an almost irresistible argument for the claim that there is a 'cutoff' n such that n grains can't make a heap and $n+1$ grains can make a heap. Here is a simplified version to give the idea ('Hn' means 'n grains can make a heap').

(1)	$\sim H0$	Premise
(2)	$\sim H1 \vee H1$	LEM
(3)	$\sim H2 \vee H2$	LEM
(4)	$H3$	Premise(!)
(5)	$\sim H1 \vee (\sim H0 \wedge H1)$	1, 2
(6)	$\sim H2 \vee (\sim H1 \wedge H2) \vee (\sim H0 \wedge H1)$	3, 5
(7)	$(\sim H2 \wedge H3) \vee (\sim H1 \wedge H2) \vee (\sim H0 \wedge H1)$	4, 6
(8)	$\exists n(\sim Hn \wedge Hn+1)$	7, proof by cases

Of course for the argument to be plausible we'd need to choose a much larger number than 3 for the last premise and many more instances of LEM, but the idea is the same.

Does that mean that if we accept LEM we have to say that there is a sharp cutoff or boundary between the heaps and the non-heaps? No—'sharp' and 'blurred' are metaphors, and probably everyone with a theory of what vagueness is will want to use that theory to cash out the metaphor.

3 The Lindenbaum-Tarski lemma

Definitions: A set of sentences is *closed under logical consequence* iff whenever a sentence is a logical consequence of some of its members, that sentence is itself one of its members. A set of sentences is *maximal consistent* iff (i) it is logically consistent, and (ii) it is not a proper subset of any logically consistent set.

The Lindenbaum-Tarski Lemma: for any set D of sentences which is closed under logical consequence, there is some class Γ of maximal consistent sets of sentences such that D is the intersection of the members of Γ.

> *Proof:* Take Γ to be the set of all maximal consistent sets of sentences that include D. So it's immediate that every member of D is a member of each member of Γ. In the other direction: suppose that s is not a member of D. Then s is not a logical consequence of the members of D (since D is closed under logical consequence). So the negation of s is consistent with the members of D. So there is a consistent set of sentences that includes all the members of D and the negation of s. We can extend this set to a maximal consistent set of sentences that includes all the members of D and the negation of s. So there is a *maximal* consistent set of sentences that includes all the members of D and does not include s. So there is a member of Γ of which s is not a member.

The argument depends only on the following claim about logical consequence: a sentence is a logical consequence of the sentences in some set iff its negation is not consistent with them. This is controversial, but holds in classical logic.

Extensions: (i) Suppose **D** is a *property* of sentences such that necessarily, the set of sentences that have **D** is closed under logical consequence. Then there is some set Γ of properties of sentences, such that for every $\mathbf{P} \in \Gamma$, necessarily the set of sentences that has **P** is maximal consistent, and such that necessarily, a sentence has **D** iff it has all the properties in Γ.

(ii) Instead of looking at sets/properties of *sentences*, we can look at sets/properties of ordered n-tuples each of which has an $n-1$-adic predicate as its first member. (This gives us a way of talking about the 'true of' relation.) Logical consequence and logical consistency can be defined on such n-tuples in various ways any of which would work equally well in the lemma.

- One needs a rather strong and philosophically controversial property theory to make this go through. But I won't talk about this now.
- If one wants to allow quantification as a source of vagueness, one won't want to include certain characteristic quantifier-based inferences in the notions of "consequence" and "consistency" for sets of such n-tuples.

4 Supervaluationist theses

The mildest claims: Necessarily, a sentence is (strongly) true-in-English iff it is true in all precisifications-of-English. Necessarily, a sentence is borderline-in-English iff it is true in some precisifications-of-English, and false in others.

- What kinds of entities are precisifications-of-English? One convenient answer: they are *properties* of sentences (strings of symbols). A sentence is "true in" a given precisification iff it *has* the property that is the precisification.
- The same account can be extended to "fully interpreted" languages in general: a special hallmark of *precise* languages is maximal consistency.
- We may want to expand this analysis along the lines of extension (ii) from the previous section, to allow for talk of a predicate being *true of* an n-tuples of things in a language.

Thanks to the Lindenbaum-Tarski lemma, we know that provided the property of (strong) truth-in-English is necessarily closed under classical logical concsquence, we can always find an interpretation of 'precisification-of-English' under which these claims are true. We can simply take 'precisification-of-English' to mean 'precise language in which every sentence that is (strongly) true in English is true'.

Some supervaluationists want to make a stronger claim: that *what it is* for a sentence to be true-in-English is for it to be true in all precisifications-of-English. ('Truth is supertruth.')

- This is such a small analytic step forward that it hardly matters. The central question will still be, 'In virtue of what is it the case that we are speaking English rather than some other language'?

A more ambitious analytic claim: necessarily, a sentence is (definitely) true in some arbitrary language iff it is true in all precisifications of that language.

- But since it is clearly possible for there to be a language in which sentences of the form '$P\vee\sim P$' are not all true, one had better not require that the precisifications of arbitrary languages meet any syntactic test for precision!
- This isn't to rule out giving some interesting independent characterisation of what makes it be the case that some people are speaking a language of which a given other language is a "precisification", on which it is an empirical truth about us that we are speaking a language whose precisifications all have the property that whenever they don't assign truth to a sentence, they assign truth to its negation.

5 "Semantic indecision"

Fine: what is special about the precisifications is this: if we were to modify the meanings of our words in such a way as to make them more precise in those ways, that would not count as an outright "change" of meaning but only as a "refinement". If one has some independent grip on that contrast, it could be taken as the basis for some interesting analytic claim.

Lewis: 'Vagueness is semantic indecision': the precisifications of our language are languages such that we have made no collective decision (or convention) as regards which of them we are to speak. A given language is one of the precisifications of the language we actually speak iff one could adopt the beliefs and motivations characteristic of a speaker of that language without violating any linguistic conventions.

6 Truth

I've been assuming, so far, that borderline sentences are neither true nor false. But if one wanted instances of the disquotation schema like

> 'Harvey is bald' is true iff Harvey is bald.

to come out always true, and never borderline, one might give this up.

If truth obeys the disquotational principle, then we can infer from 'Harvey is either bald or not bald' to 'Either "Harvey is bald" is true or "Harvey is not bald" is true', i.e. 'Either "Harvey is bald" is true or it is false.' This is an instance of the principle of *bivalence*: every meaningful sentence is either true or false. Presumably we will accept at least some restricted version of this principle—at least vagueness will no longer give us any distinctive source of counterexamples to it.

Get used to the idea: Although 'The patch is red' is borderline, it is either true or false. Likewise for 'the patch is not red'. Hence some borderline sentences are true, and others are false.

Of all the precisifications of our language, one is special in that it assigns truth to all and only those sentences that are in fact true. (But it is indefinite which one it is!)

Question: why should we *care* about borderlineness? If I could express a truth by asserting either 'The patch is red' or 'The patch is not red', why should I be held back from doing so by the consideration that if I said either of these things I would be saying something borderline?

Fine claims that we could stipulatively introduce a predicate 'true$_D$' that obeys the disquotation schema even if we started out only understanding the "strong" notion of truth as inconsistent with borderlineness.

- Perhaps such a stipulation would work for our own language, but it doesn't tell us what to say about sentences of other languages. Or if 'true$_D$' is taken as a predicate of *utterances*, it doesn't tell us what to say about utterances in other languages.

7 Validity

Assume that validity is (necessary) preservation of truth. If truth is disquotational, we get something for which classical logic is entirely acceptable as a theory—call it "disquotational validity".

- This is not the same as what Williamson calls "local validity": preservation of truth in each precisification. An argument is locally valid iff it is definitely disquoatationally valid. EG: suppose it is indefinite whether 100 grains can make a heap. Then 'x contains 100 grains, therefore x is not a heap' is not locally valid; it is indefinite whether it is disquotationally valid.
- But it's still true that all classically valid rules are locally valid, and all classically valid metarules preserve local validity.

If we take truth to be supertruth, on the other hand, things get a bit more complicated: we get what Williamson calls "global validity". Obviously any argument that is definitely disquotationally valid is globally valid. So, since all classically valid arguments are definitely disquotationally valid, they are all globally valid. But there are globally valid arguments that are *not* definitely disquotationally valid. For example: 'Soros is a heap; therefore Soros contains more than 100 grains'. (Hard question: are there globally valid arguments that are definitely *not* disquotationally valid?)

Because of this, classical *metarules* fail for global validity. A metarule says: if such-and-such arguments have been established to be valid, conclude that such-and-such other argument is also valid. EG: Although 'Soros is a heap; therefore Soros contains more than 100 grains' is globally valid, the sentence 'If Soros is a heap, Soros contains more than 100 grains' is not globally valid since it is not necessarily true on all precisifications.

- I've been assuming that we're using 'valid' in such a way as to encompass "analytic" validity in general. What I said may not be true for a *narrowly logical* notion of validity. But it's disputed whether there's any interesting, non-arbitrary distinction between the narrowly logical validities and more generaly analytic validities, and if so what it includes. (Are the validities of epistemic "logic" narrowly valid? Modal logic? Deontic logic? The logic of 'definitely'?)
- At one point in Fine there is an offhand suggestion (made in connection with the 3-valued theory rather than supervaluationism, but I don't see why it wouldn't generalise) that we should take validity to be preservation of what we might call "super-duper-truth": definite definite definite... truth. Whether this gives an interesting notion depends on how easy we think it is for a sentence to be super-duper-true. In my own view, it's impossible: no sentence whatsoever is super-duper-true.

8 Vague propositions?

Three views about the vagueness of 'true proposition' and 'expresses' (as in "sentence S expresses proposition P").

(i) 'true proposition' is precise: it has no borderline cases, necessarily. Nevertheless, 'The proposition that Harvey is bald is a true proposition' is vague: this is due entirely to vagueness in the singular term 'the proposition that Harvey is bald'. Since ' "Harvey is bald" expresses the proposition that Harvey is bald' is (presumably) definitely true, 'expresses' is also vague.

- Consequences for "propositional attitude ascriptions" like 'John asserted the proposition that Harvey is bald'. Either John asserted very many propositions, or there is no proposition that John definitely asserted.

(ii) 'expresses' is precise: it has no borderline cases, necessarily. For each meaningful sentence, there is some proposition that it definitely expresses. Likewise, terms like 'the proposition that Harvey is bald' are not vague: each such term determinately refers to the one specific proposition that its sentential component determinately expresses. 'is a true proposition' is vague.

(iii) A hybrid view: 'is a true proposition' and 'expresses' are both vague. 'The proposition that Harvey is bald' may be indeterminate in reference between several propositions, and some of these propositions may be neither definitely true nor definitely false.

It would be natural, though not inevitable, for a proponent of option (ii) to claim that the *central* notion of vagueness is that of vagueness in propositions: vagueness in language simply consists of the expressing of vague propositions. A proponent of this view could still accept the apparatus of supervaluationism as an account of vagueness in propositions. The account could go something like this: a *precisification* is a function that takes each proposition to some precise proposition. The nature of propositions is such as to make some precisifications admissible, others not. Necessarily, a proposition is (definitely) true iff every admissible precisification maps it to a true proposition.

- One might imagine "constructing" such propositions out of precise propositions. One could take vague propositions to be sets of precise propositions—but that won't allow for penumbral connections. Perhaps one could take a vague proposition to be something like a set of n-tuples of precise propositions, together with a number $m \leq n$? But on any such constructive account, admissibility will itself come out as precise. If we want propositions to be higher-order-vague, it's very hard to see what admissibility could *be*

9 Higher-order vagueness

We have been reasoning classically about precisifications, assuming for example that there is a *unique* set of admissible precisifications of every given language. The great virtue of supervaluationism is the fact that even though 'admissible precisification' is surely vague, as is 'borderline sentence', recognising this vagueness doesn't force us to stop reasoning classically about it as we have been doing. 'There is a unique set of admissible precisifications' is true on every precisification, although different precisifications assign different extensions to the expression 'admissible precisification'. Some precisifications happen to assign the set of admissible precisifications to the expression 'admissible precisification', just as some precisifications happen to assign the set of heaps to the expression 'heap'. But it's indefinite which precisifications have this feature!

One doesn't have to do anything *extra* to make the supervaluational approach compatible with higher-order vagueness.

Williamson suggests that the supervaluationist will want to elaborate his theory to include a "semantics" for the "definitely"-operator. But it's not at all clear what it means to give such a semantics. On one view, 'definitely' is surreptitiously metalinguistic: 'definitely P' just means ' "P" is true on all precisifications'. On such an account there is no need to say anything *special* about the definitely operator: different precisifications assign different truth-values to sentences involving it in whatever way they assign different truth-values to any other sentences.

On another account, 'definitely' is primitive. In that case, the only sense in which we can give a "semantics" for it is that of providing some formal method that lets us distinguish between valid and invalid inferences involving it.

It's not clear what the other options are! And the precedent of the possible-worlds semantics for modal logic is worthless as an explanation for what one might be doing in providing the distinctive "semantics" for the definitely operator—for no-one has a clear sense what the possible worlds semantics might be if not an analysis of 'necessarily' as 'true in all possible worlds', and not a mere calculus for determining validity.

10 Williamson on ‘definitely*’