3. Supervaluationism

Words don't get their meaning by magic. We give them their meaning. Oh, to be sure we might not do all the work. We call some stuff of a certain kind 'water', and the *world* makes it the case that some other things satisfy 'water', because they have the same kind of structure as the things that we called 'water', and other stuff does not satisfy this predicate because it has a different structure, whatever its surface similarity to water. But the initial work is done by us, and the world helps us fill in some of the gaps.

Now, most of the time we can use a term even if we haven't decided just how, or whether, it applies in all sorts of hard cases. Take a more contentious term, 'religion'. I can still use this word to impart some information even if we haven't exactly decided where the boundary is between religions and non-religions. If I tell you that such-and-such is a religion, then you can infer that we (the relevant linguistic community) has decided that the term's extension is wide enough that it encompasses this activity. And if you know something about where that extension falls, you can infer some things about the activity. On the other hand, if I tell you that such-and-such is not a religion, then you can infer that we (the relevant linguistic community) has decided that the term's extension is not wide enough to encompasses this activity, which again might tell you something about the activity, if you know enough about the linguistic community's application of the term. I assume all throughout here that you take my word on what is and is not a religion, perhaps not the best strategy in life, but still one we assume here. And I ignore potential complications about what counts as the relevant linguistic community - as we'll see later in the course, some people think it is rather small. To this point, the story I've told is not particularly controversial. Supervaluationists add a few distinctive spins.

First, they say that in cases where we do not seem to have decided whether a term applies to a particular (kind of) object, this is not because we have decided to give the term a vague meaning, but because we have failed to decide which precise meaning to give it.

Secondly, they say that this type of semantic indecision is constitutive of vagueness, though they leave open the possibility that not *all* semantic indecision really amounts to vagueness. Some semantic indecision may be Quinean indeterminacy, and Quinean indeterminacy *might* not be vagueness.

Thirdly, as part of the consensus, they say that the fact that we haven't made these decisions is consistent with the idea that the decisions that we have made can make some sentences true. So if we have decided that any object with property *F* satisfies the predicate 'is a religion', and *x* is *F*, and *NN* is a name for *x*, then ⌜*NN* is a religion⌝ is true.

Most distinctively, they say that the truth of ⌜*NN* is a religion⌝ is a special case of how sentences containing these terms with incomplete meanings get their truth value. Say a **precisification** of a language is a way of assigning a precise meaning to every word in the language consistent with the decisions that we have already made. So if we have decided that any object that is *F* satisfies 'is a religion', and no object that is *G* does, then on all precisifications, 'is a religion' will only be satisfied by objects that are *F* but not *G*. But if we have not decided whether objects that are *H* satisfy this predicate, then there will be some precisifications on which they do, and some on which they do not. The distinctive supervaluationist claims are that

(a) there are many precisifications of language, because semantic indecision is pervasive;
(b) a sentence is true if it is true on all these precisifications;
(c) a sentence is false if it is false on all these precisifications; and
(d) if a sentence is true on some precisifications and false on others, then it is neither true nor false.

The last claim is a mild exaggeration. All supervaluationists accept (a) to (c), but some do not accept (d), as we shall see.

A sentence that is true on all precisifications is said to be **supertrue**, and a sentence that is false on all precisifications is said to be **superfalse**. So if (a) through (d) are all true, then truth is supertruth, and falsity is superfalsity. These claims are sometimes taken to be distinctive of supervaluationism, but since they only hold if (d) is true, and some interesting supervaluational theories reject (d), I will not say that the claims are constitutive of supervaluationism.

In terms of distinguishing supervaluationism from the many-valued theory, it is important to note that (b) and (c) apply to compound sentences as well as simple sentences. Although supervaluationists believe in truth-value gaps, they do not believe in 3-valued, or many-valued, truth-tables. Rather, they think the truth value of a compound sentence is determined by its truth on every precisification, and on every precisification its truth is determined by the 2-valued truth tables. So, for example, however we draw the line between the religions and the non-religions, (1) will be true, since on every way of making the sentence precise, it is true.

(1) If communism is a religion, then communism is a religion.

More controversially, (2) and (3) are true on every way of making the terms in them precise, so they are true.

(2) It is not the case that communism is a religion and communism is not a religion.
(3) Either communism is a religion or communism is not a religion.

It might be surprising that (3) is true, since we haven't yet said enough to guarantee that either disjunct is true. (I assume here that it is vague whether communism counts as a religion. This is a kind of silly example, but it gets talked about a bit in the literature, and I suppose once Castro dies we won't be able to use it any more, so get a last run out of it before it is too late.) But on every precisification, one or other disjunct will be true. Different disjuncts will be true on different precisifications, but as long as every precisification delivers a true disjunct, the whole disjunction is true.

Finally, supervaluationists hold that the semantic decisions that we have made include not just categorical decisions, like that things with such-and-such property satisfy this-or-that predicate, but conditional decisions, like if something satisfies this predicate, it doesn't satisfy that predicate, or that if things with some property satisfy a particular predicate, then so do things with suitably related properties. That sounds kind of abstract, but hopefully it is clearer with some examples.

For an example of the first kind of decision, note that the border between the blue things and the green things is fairly vague. But, according to many people, it is impossible that a (monochromatic) thing could be both blue and green. So in borderline cases, there may be some precisifications on which a particular thing satisfies the predicate 'is blue', and some precisifications on which it satisfies the predicate 'is green', but none on which it satisfies both predicates. The acceptable precisifications of these terms are not independent.

For an example of the second, note that we accept, as a general principle, that if *x* is tall for an adult male, and *y* an adult male who is taller than *x*, then *y* is tall for an adult male. Whatever we may have left undecided in the meaning of 'tall', we have decided that it links to the taller than relation in this way. Again, this leads to a restriction on what counts as an acceptable precisification of 'tall'. So while there is a precisification of 'tall' on which 181cm tall adult males are not tall for adult males, and another on which 180cm tall adult males are tall for adult males, there is no precisification on which both of these things hold.

These kinds of restrictions on precisifications are called penumbral connections. We already made one appeal to penumbral connections in the discussion of Sorenson's example of the twin basketball players. We said there that however vague it was whether the twins were tall, it was definitely true that if the shorter one were true, then so was the taller one. By acknowledging these conditional semantic decisions as well as categorical semantic decisions, supervaluationism can account for these facts.

The big advantage of supervaluationism is that, putting things a little colourfully, it manages to be conservative but open-minded. It is conservative in that it says that we don't need to change our logic because of vagueness. Anything that classical logic says is a logical truth, or indeed any inference that classical logic says is valid, is valid according to supervaluational theory. It is open-minded in that it doesn't get this result by denying, or even explaining away, the data. There really are no precise facts about when someone has grown enough to be tall or fat, or how far away you have to walk before you are no longer around here. It is nice to be able to steer down these channels so nicely. The costs of the theory are a little harder to state, and we shall spend longer on them. But it is worth keeping in mind from the top just how attractive the theory looks at first, especially compared to the myriad difficulties immediately confronting the many-valued theories.

When surveying the costs of supervaluationism, I will frequently refer to the responses to the various objections made by Rosanna Keefe in her book *Theory of Vagueness*. This is in part because Keefe actually believes a fairly orthodox version of supervaluationism, so she's probably in a better position than me to say what is in the best interests of the theory. And in part it is because I often agree with her about what is best things for supervaluationists to say in the face of a particular criticism. Which is not to say that I think her responses always work, or even frequently work.

3.1. Two Kinds of Truth

There are two ways of defining truth in supervaluational theories. These correspond to what Hartry Field calls the correspondence and disquotational conceptions of truth.

On the correspondence conception, truth is supertruth. A sentence is true if it is supertrue, false if it is superfalse, and *neither true nor false* otherwise. This is the standard account within supervaluationism, but it is not the only one we can define, nor is it necessarily the most natural.

On the disquotational conception, we preserve all instances of Tarski's schema T: '*A*' is true iff *A*, where we can substitute any object-language sentence for *A*. The idea is that if a sentence is supertrue then it is true, and if it is superfalse, then it is false, but if it is neither supertrue nor superfalse, then it is indeterminate whether it is true or false. To capture this formally, note that within any precisification, we can introduce a new sentential predicate 'true$_T$' that satisfies the T-schema. (Well, satisfies all versions of it that can consistently be satisfied. The issues raised by the liar paradox become rather pressing here, but we won't be pressed, at least not now!) If we interpret truth as truth$_T$, then since the (consistent instances of) the T-schema are true on all precisifications, they are true simpliciter. On the other hand, since on all precisifications there are no sentences that are neither true nor false, then it is true simpliciter that there are no sentences that are neither true nor false.

Matching these two conceptions of truth are two conceptions of validity, commonly called *local validity* and *global validity*. An argument is locally valid iff whenever the premises are true on a precisification, the conclusion is true on that precisification. An argument is globally valid iff whenever the premises are true on all precisifications, the conclusion is true on all precisifications. In an interesting sense, local validity is like the idea of validity as degree-of-truth preservation, and global validity is like the idea of validity as preservation of degree of truth 1. With a locally valid argument, you are guaranteed to never go from truth to falsity no matter how the terms get precisified, just like you are guaranteed to never move further from the truth no matter where you are in relation to it on the earlier conception. With a globally valid argument, you are only sure to preserve truth itself, but you may not preserve the property of being near to the truth.

If we think validity is truth preservation, then these two conceptions of validity match up with our two conceptions of truth. Global validity is preservation of supertruth, so if truth is supertruth, and validity is preservation of truth, then validity is global validity. Local validity is preservation of truth$_T$, so if truth is truth$_T$, and validity is preservation of truth, then validity is local validity.

Getting clear on these two conceptions of truth is important for understanding the varieties of supervaluationism. But it is also important because concerns about the way truth is handled raise some of the most immediate difficulties for supervaluationism. There will be different problems raised depending on which theory of truth we take, but as we shall see there are difficulties either way. We now turn to those problems.

3.2. Problems with Correspondence

There are three problems with taking truth to be supertruth. In rough order of importance these are: violation of intuitive semantic principles concerning disjunctions and existential quantification; abandonment of classical rules of inference; and, violation of core principles concerning the concept of truth.

It is very plausible that if a disjunction is true, then one or other disjunct is true. And it is very plausible that if an existentially quantified sentence is true, then one or other instance of it is true. If truth is supertruth, these plausible principles are in fact false. We have already seen how a disjunction can be supertrue even though neither disjunct is supertrue. And the core of the supervaluationist response to the Sorites is that an existentially quantified claim can be true even though no instance of it is true.

David Lewis notes that we can create similar puzzles in other areas. It might be true that I owe you a horse, but false of any particular horse that I owe you it. Once we have modal operators interacting with existential quantifiers, we have to be more careful than normal about which sentences are equivalent to other sentences. To spell out the analogy, in Lewis's example (4) might be true but (5) false, while, returning to our earlier example of the light rays, (6) might be true but (7) false on the supervaluationist theory.

(4) It is obligatory that there is a horse such that I give it to you.
(5) There is a horse such that it is obligatory that I give it to you.
(6) It is true that there are a pair of adjacent light rays such that one is red and the other is not.
(7) There are a pair of adjacent light rays such that it is true that one is red and the other is not.

Now there is no doubt we can create formal models for the logic of truth, mimicking perhaps the logic of obligation, on which (6) is true and (7) is not. The objection, though, is that intuitively, truth is not this kind of operator. How strong this intuition is, and whether it can be overridden by other theoretical considerations, will be left to you to decide.

One of the big selling points of supervaluationism is that it lets you keep classical logic. But strictly speaking, this is only true on a rather narrow conception of what counts as classical logic. If we identify classical logic with sequents, then we will say that we have kept classical logic provided that we have preserved all the valid sequents. But if we identify classical logic with rules of proof, then we will only say that we have kept classical logic if we have preserved all the classically admissible rules of proof. And if we take truth to be supertruth, and validity to be global validity, then we will not preserve all the classically admissible rules of inference. In particular, we will not preserve the following rules.

Conditional Proof: From $\Gamma, A \vdash B$ infer $\Gamma \vdash A \rightarrow B$
Reductio ad Absurdum: From $\Gamma, A \vdash B \wedge \neg B$ infer $\Gamma \vdash \neg A$
Argument by Cases From $\Gamma, A \vdash C$ and $\Gamma, B \vdash C$ and $\Gamma \vdash A \vee B$ infer $\Gamma \vdash C$

To create problems for all of these rules, note that the inference $A \vdash \ulcorner A$ is true$\urcorner$ is valid if truth is supertruth and validity is global validity. For in that case, the premise is true iff *A* is true on all precisifications, which is what the conclusion says. But, $\ulcorner A \rightarrow A$ is true$\urcorner$ is not a logical truth: if *A* is true on some but not all precisifications, then the consequent is false, i.e. superfalse, which means it is false on all precisifications, which means there are some precisifications on which the antecedent is true and the consequent false, which means there are some precisifications on which the conditional is not true, as required. Now for the counterexamples.

Conditional Proof: $A \vdash A$ is true but $\nvdash A \rightarrow A$ is true.
Reductio ad Absurdum: $\neg\ulcorner A$ is true$\urcorner$, $A \vdash A \wedge \neg A$ but $\neg\ulcorner A$ is true$\urcorner \nvdash \neg A$
Argument by Cases $A \vee B, A \vdash A$ is true or *B* is true
$A \vee B, B \vdash A$ is true or *B* is true
$A \vee B \vdash A \vee B$,
but $A \vee B \nvdash A$ is true or *B* is true

Rosanna Keefe responded to this problem by noting that in every case, a similar inference rule to the classical rule could be defended. If we let *DA* be short for *Definitely A*, which we take to be true iff *A* is supertrue, then Keefe shows that each of the following rules are acceptable within the supervaluational theory.

Conditional Proof: From $\Gamma, A \vdash B$ infer $\Gamma \vdash DA \rightarrow B$
Reductio ad Absurdum: From $\Gamma, A \vdash B \wedge \neg B$ infer $\Gamma \vdash \neg DA$
Argument by Cases From $\Gamma, A \vdash C$ and $\Gamma, B \vdash C$ and $\Gamma \vdash DA \vee DB$ infer $\Gamma \vdash C$

Two points can be made in response to this. First, it is not immediately obvious just what the benefit is of preserving something close to classical logic. If the aim was to keep classical logic, then the aim should be to keep classical logic, not a near approximation to it. Secondly, these new rules of proof are of dubious value themselves. The following is a short list of sequents which are valid according to the preferred supervaluational semantics, but which cannot be proven using Keefe's rules.

$\vdash p \rightarrow p$
$p \rightarrow r \vdash (p \wedge q) \rightarrow r$
$p \vee q \vdash q \vee p$
$p \vdash \neg\neg p$

I'd think we should be able to prove at least *some* of these.

The final problem, and the one that has received most attention in the literature, is simply that the correspondence concept of truth is not disquotational. Here is how Timothy Williamson puts the point.

> Truth is standardly assumed to have the disquotational property to which Tarski drew attention. 'Cascais is in Portugal' is true if and only if Cascais is in Portugal. More generally: '*A*' is true if and only if *A*. Here '*A*' may be replaced by a sentence of the object-language under study … The 'if and only if' is just the material biconditional. How much more there is to the concept of truth than the disquotational property is far from clear, but in most contexts truth is assumed to be at least disquotational, whatever else it is or is not. (162)

But as we saw above, supertruth is not disquotational. The conditional *If A then A is true* is not, in general, true if 'true' is interpreted as supertrue. Isn't this embarrassing? Keefe makes four responses to this.

The first is to note that the disquotational concept looks incoherent when we start considering the paradoxical sentences, though it seems she can't quite decide how much of an issue this raises. After raising the point that the T-schema simply can't be right in general for the second time in three pages, she magnanimously agrees to waive the point, saying, "Again, I put aside the liar paradox." The whole passage comes off as a guilty attempt to smear the T-schema by association. Which is unfortunate, because the underlying point has *some* merit. Here's a more epistemic way of putting the point. The T-schema is, in general, mistaken, so the intuitions backing it are unreliable, so the fact that they support the T-schema in cases where it *could* be consistently applied *might* not have much evidential weight. On the other hand, it might, so perhaps we should look at other points. And Keefe never exactly shows that the concept of supertruth she uses is consistent, so at best we no longer *know* that we are using an inconsistent theory, which is a relatively small advantage I think.

The second is to note that at least the T-schema is never false. This isn't as useless a move as it was when Tye tried it in defence of the three-valued system, but still feels pretty feeble.

The third is to point out that something similar to the T-schema is true. As Keefe notes, van Fraassen back at the dawn of time (well, in 1966) pointed out that in most three-valued systems, the following mutual entailment principle holds, even though the T-schema does not.

(T*) $A \dashv\vdash$ '*A*' *is true*

We are then supposed to think that our disquotational intuitions will be satisfied by the fact that this mutual entailment hold. Perhaps when we think the T-schema is valid, we are just confusing it with this. (Or perhaps we recognise this and 'mistakenly' apply conditional proof. Who knows?)

Finally, Keefe thinks that the argument for the T-schema, as opposed to T*, begs the question against the person who believes in truth-value gaps. Keefe pictures the argument going as follows. First, we note that if *A* is true, then the biconditional *A iff A is true* is true. Second, we note that if *A* is false, then the biconditional *A iff A is true* is true. Then we claim, well that's all the possible cases, so it is always true that *A iff A is true*. And the last step is evidently question-begging against a determined promoter of intermediate truth values. Indeed, if we think about how this kind of reasoning should continue when faced with the possibility that *A* is neither true nor false, we might see that the schema is not always true.

The problem with this is that it is far from clear that is how we actually argue for the T-schema. In fact, it is far from clear that we accept it because of any kind of *argument*. The principle just seems intuitively obvious, and if it clashes with the principle that there are truth-value gaps, well all the worse for the principle that there are truth-value gaps. Of course, maybe the principle that there are truth-value gaps is intuitively compelling, though I have my doubts there. Or maybe this isn't intuitively compelling, but we need to accept it to accommodate other intuitions concerning, say, vagueness. And those other intuitions have a stronger claim to accommodation than the intuition that the T-schema holds. This kind of holistic justification ends up being Keefe's fall back position, and it has a considerable amount of merit. Of course, it needs a lot to go right for it to work.

First, it needs our attachment to the T-schema to be weak enough that we'd be prepared to trade it off for those other benefits. Speaking for myself, I think we should *be prepared to* make that trade, though I'm not sure Williamson would concede even that much.

Secondly, it needs there to be no other significant problems for supervaluationism. You can only play the 'this theory is the best of a bad lot' card so often. As we'll see, it isn't exactly obvious that that is true.

Finally, it only works if supervaluationism really is the best of a bad lot. So to tell whether this objection works we have to survey the entire range of theories about vagueness, and then worry about whether there's been any good theories not yet considered. That seems like a lot of work. Let's start with a theory very similar to Keefe's own that may do the job.

3.3. Problems with Disquotation

The most immediate problem facing Keefe's holistic justification of taking truth to be supertruth, the one Williamson immediately seizes on, is that keeping the rest of supervaluationism, but taking truth to be truth$_T$, promises to deliver all of the benefits of the supertruth theory without giving up the T-schema. We then get that the T-schema is always correct (modulo concerns about the semantic paradoxes), that validity is local validity, and that a rule of inference is acceptable iff it is classically acceptable. So what could go wrong?

Well, as Williamson points out, this theory is close to self-refuting. If we accept the T-schema, and the law of excluded middle, and all classical inference rules, we are committed to the following common formulation of the principle of bivalence: For any sentence *A*, either *A* is true or *A* is false. (I will argue later in the course that this is not the correct formulation of the principle of bivalence, that the right formulation is that, as the etymology suggests, there are exactly two truth values. We can accept that every sentence is either true or false, and deny that there are exactly two truth values, if we deny that *true* and *false* are truth values. Why would we do that? Good question - read on and *eventually* we will answer it!) That isn't too bad in itself, but Williamson argues that any supervaluationism that accepts bivalence is self-defeating.

We were a bit rough above with what counted as an acceptable precisification. The following looks like a *minimal* constraint. If *A* is true in English, and false according to a particular precisification, then that precisification is not an *acceptable* precisification. Conversely, if *A* is false in English, and true according to a particular precisification, that precisification is not acceptable either. Precisifications are meant to fill in the gaps, not to change the clear cases. The metaphor should already tell us what is going to go wrong - if there are no gaps, then there are no gaps to fill. Bivalence plus the above two constraints on precisifications tell us that for all acceptable precisifications, a sentence is true on the precisification iff it is true in English. And that means that there is only one acceptable precisification. And that means, so it seems, that the language is precise after all.

We can escape this trap only by denying the validity of the argument forms used to spring the trap. So we might want to give up excluded middle, or give up some of the principles concerning disjunction that let us get from the T-schema and excluded middle to bivalence, or the principles concerning quantification that let us get from bivalence to the uniqueness of the acceptable precisification. Is this any worse than what the supertruthers have to do? In one respect, no, in another respect, yes. Remember that there were a few problems with the supertruth approach: to avoid this trap the we merely need to adopt one of those problems. On the other hand, there was a principled reason within the supertruth approach to accept that these classical rules of inference fail. Within the truth$_T$ approach, we need some or other classical rule of inference to fail, but we haven't got a start of a principled reason as to why they should fail.

3.4. Fodor and Lepore's Objection

The above complaint is similar to one raised by Jerry Fodor and Ernest Lepore in their 1996 *Journal of Philosophy* paper "What Cannot be Valuated Cannot be Valuated, and it Cannot be Supervaluated Either". (They wanted a snappier version, but the Columbia philosophy department takes a principled stand against contractions.) They claim that the precisifications which are the core of this account do not satisfy a mandatory condition on models for a language: that all conceptual truths are true in the model. Because of this there is no reason to think 'truth in all precisifications' is really truth.

The objection is built up as follows. Vague terms like 'bald' have a penumbra. Assume for the sake of the argument that a person with 1/9 of their head covered with their own hair is in this penumbra, neither definitely bald nor definitely not bald. If the ratio is wrong, it can be changed. All that matters is that there is one, which everyone agrees. Let *S* be the sentence 'A person with a head-to-hair ratio of 1/9 is bald'. If our assumption is right, *S* is neither definitely true nor definitely false. Indeed, on standard accounts *S* is neither true nor false *simpliciter*.

That *S* is not definitely true is not something that we discovered by looking at the world. Of course we discovered what *S* means by looking at the world, but once we found that out we worked out by conceptual analysis alone that it is not definitely true. So '*S* is not definitely true' is a conceptual truth. Fodor and Lepore then wield what they call principle (P).

(P) Conceptual truths must be respected by all classical models, including classical valuations. (521).

The justification is that a purported model of a language which does not respect conceptual truths is not a genuine model. "If there are conceptual truths, then they determine what the topic under discussion *is*, so they must not be flouted on pain of equivocation." (521). But precisifications do not satisfy this criteria. To see this, note that in many precisifications, '*S* is not definitely true' comes out as false, despite being a conceptual truth. In all those precisifications in which *S* does come out as true, the conceptual truth '*S* is not definitely false' comes out as false. So there are no precisifications which satisfy principle (P).

As well as this general argument for (P), they have an *ad hominen* against the supervaluationist who does not accept it. Say, Al and Bill each have 1/9 of their head covered with their own hair. Then there are, according to supervaluationism, acceptable precisifications according to which 'Al is bald' comes out true. There are also acceptable precisifications according to which 'Bill is bald' comes out false. But there is no acceptable precisification according to which 'Al is bald' comes out true and 'Bill is bald' comes out false. Referring to Fine (1975), Fodor and Lepore claim that the supervaluationists' reason for this is that precisifications must preserve conceptual truths, in this case the conceptual truth that baldness supervenes on head-to-hair ratio. So by their own lights, supervaluationists are committed to (P). But there are no precisifications that are acceptable according to (P).

The supertruth theorist should have little difficulty responding to Fodor and Lepore. Even if *S* is true according to a particular precisification, that alone is no reason to think that *S is true* is true according to that precisification, let alone *S is definitely true*. Remember, these people do not believe in the T-schema! In short, supertruth theorists do not believe in the autonomy of precisifications. Whether *S is true* is true according to a precisification depends not only on how that precisification lines up the words, but on how the others do as well.

Matters are more complicated for the truth$_T$ theorist. Still, we can make a three-fold response on their behalf. First, some reasons for thinking that (P) need not be satisfied by models are discussed. The basic point is that precisifications of English are not meant to be meaning preserving at the level they are discussing. It's no news to say 'bald' in a precisification means something different from 'bald' in English because the former is precise and the latter is vague. The second is that we can distinguish acceptable from unacceptable precisifications without relying on (P). Finally, it is argued that there is no way to make sense of Fodor and Lepore's positive suggestion, which is that *S* is gappy, and must be so on all models, without using supervaluations. In sum, we won't dispute that giving up principles like (P) is part of the cost of adopting a supervaluational account, but I think the cost can be shown to be rather small, and the benefits rather large.

Fodor and Lepore treat precisifications as languages, so we can talk about the meaning of a word in a precisification. This is a very natural way of putting things, but it will ease presentation here if we treat them as sets of true sentences, or equivalently as (complete) functions from sentences to truth-values. In any case, we ought to be able to determine the meaning of a word in a precisification from the set of sentences containing that word which are true. Let *E* be the set of first-order truths of English. So 'Snow is white' is in *E*, but '"Snow is white" is true' is not in *E*, and nor is "It is true that snow is white". In short, *E* is the set of true sentences that do not employ any semantic machinery, and I'm for convenience calling these first-order sentences. Let E^* be any maximally consistent superset of *E* which is closed in the following ways:

$A \,\&\, B \in E^*$ iff $A \in E^*$ and $B \in E^*$
$A \vee B \in E^*$ iff $A \in E^*$ or $B \in E^*$
$\neg A \in E^*$ iff $A \notin E^*$
If $A \rightarrow B \in E^*$ and $A \in E^*$ then $B \in E^*$
'A' is true $\in E^*$ iff $A \in E^*$

E^* is a precisification of English iff it satisfies all of these conditions. (I'm intending '→' here to be read as a natural language conditional; the condition regarding it is redundant if it is read as a material implication.) This definition is intended to perform two jobs. First, any (first-order) truths of English are true in all precisifications. So if Jack is bald then 'Jack is bald' will be an element of all precisifications. Secondly, precisifications preserve what Fine called the penumbral connections, like 'Taking someone's hair away doesn't change them from being bald to not-bald'. The way these were preserved by Fine suggested that supervaluationists were committed to (P). However, here they are preserved not because they are conceptual truths, but because they are first-order. For example, 'Baldness supervenes on hair-to-head ratio' is a first-order truth, so it will be in E^*. Returning to our example of Al and Bill, this implies that 'If Al is bald, Bill is bald' is in *E* and hence E^*. I'm assuming here that if *A* entails *B* then 'If *A*, *B*' is true. Hence there can be no precisification in which Al is bald and Bill is not bald. Similarly, we can find general (perhaps conceptual) first-order truths which imply that bald people can't have a higher hair-to-head ratio than non-bald people.

It's a trivial fact that for one object to model another, it doesn't have to have all the properties of the object being modelled, or indeed all the essential properties. Consider the use of crash test dummies to model the behaviour of humans in car crashes. So on a natural reading of 'model', there is no reason to say that precisifications are not models of English just because they lack essential properties of English. There are two ways in which breach of (P) by precisifications would cause problems, but neither seem to be realistic possibilities.

First, if someone were claiming 'bald' in *E** means the same as 'bald' in English, then breaches of (P) would be problematic. Meaning-preserving translations ought to preserve conceptual truths. But there is a bigger problem with this approach. It would imply that we can work out the meaning of 'bald' by just stipulating a cut-off point. Since any stipulation would provide the meaning, this would lead to blatant inconsistencies. This clearly isn't what supervaluationists are trying to do. The meaning of 'bald' isn't given by its behaviour in a particular precisification, but in the set of them.

Secondly, there might be a difficulty if there were permutation problems. Granted that *E** is a model (not an analysis or translation) of English, we have to determine which English words are being modelled by particular words in *E**. Ideally there will be a function from words in *E** to words in English. However, there might be multiple plausible functions. Were this to occur then *E** wouldn't be a good model, and there might be wholesale difficulties for the supervaluationist, because it wouldn't be clear if the equivalent sentence in the model to a particular sentence of English were true or not. However, there is little evidence that there are such problems, and hopefully the preservation of all first-order truths in all precisifications prevents such a difficulty occurring. If there is a problem on these lines, no objector has yet shown it.

From this we can determine that the status of 'Jack is not definitely bald' in a precisification will depend on how we read 'definitely'. If we read it as a function from predicates to predicates (like 'very') we will assume that this is a first-order truth (if Jack is in the penumbra of bald), and so it will be true in all precisifications. On the other hand, if by this we mean '"Jack is bald" is not definitely true', then it is a second order truth. This will be true in some models and not in others. I assume that what is true in a model is, for reasons given in previous sections, definitely true in that model. So this conceptual truth will not be preserved in all models, but we have reasons for thinking models need not preserve conceptual truths.

There remains a troubling question, which *might* be the question that Fodor and Lepore most wanted to raise. Even if we discover that a sentence is true on all models thus defined, why should we care? It's all very well to compare precisifications to crash-test dummies, and boldly assert that we can in principle learn something from their behaviour even though they differ in essential properties from the things they are modelling. It's another to actually make a convincing case that this works in practice. This might sound familiar, and depressing, but as long as supervaluational theory is the best of a bad lot, I think that case is made. So we (still) have to go through the whole problem of seeing whether supervaluational theory as a whole is better than its rivals. We better get on with it.

3.5. Sorites problems

It's strange, but for all the time I've spent studying supervaluationism (and I want you to stop and reflect here on just how long that is) I've never quite figured out what the supervaluational solution to the Sorites paradox is meant to be. On the supertruth account, we can get some of the way to a solution by noting the distinction between the truth of an existentially quantified sentence and the truth of one of its instances. This distinction won't be available on the truth$_T$ account, but maybe that's a reason to accept the supertruth version. The following (long) quote from Keefe *appears* to state how she thinks the paradox gets solved. (I apologise for the length, but this is an instructive passage.)

> Existential quantification displays behaviour similar to that of disjunction: 'something is *F*' can be true though no substitution instance is true. Although there is no *h* for which 'people of height *h* are tall while people 0.01 inches shorter are not true' is true, the existentially quantified sentence formed from it, (H) 'there is a height *x* such that people of height *x* are tall while people 0.01 inches shorter are not tall' *is* true, since on each precisification some height or other makes it true. Similarly, we have the dual result that a universally quantified proposition can be false without any substitution instance being false. This is central to the treatment of the sorites paradox: the negation of (H) can be transformed into a quantified inductive premise of a sorites paradox for 'tall', namely, for appropriate series of x_i of gradually increasing height, 'for all *i*, $\neg(x_i$ is tall and x_{i+1} is not tall)'. This premise is false on the supervaluationist theory even though none of its substitution instances are false. (But objections to the supposed truth of sentences such as (H) will be considered in §5.) Versions of the paradox using a conditional instead of a negated conjunction similarly have a false premise. And when the sorites paradox is expressed in terms of a *series* of conditionals, there will be no one conditional which is false, but there will be some that are neither true nor false and on every precisification there will be one that is false. So again the supervaluationist will avoid the paradox by refusing to accept all of the premises. In short, supervaluationism solves the sorites paradox in all its forms.

There's some interesting points here, and certainly a supervaluationist solution to the sorites had better start from these points, but nothing here to me even *suggests* that there is a supervaluational solution to the sorites. To see why I'm underwhelmed, compare the sorites to another well-known puzzle: the problem of evil. To press the analogy, let's express the problem of evil as a paradox.

(P1) An omnipotent, benevolent God exists.
(P2) If an omnipotent, benevolent God exists, then this is the best of all possible worlds.
(P3) This is not the best of all possible worlds.

This isn't quite as pressing a paradox as the sorites for many because in the sorites we have an inconsistent set of propositions that are all intuitively compelling, whereas some people don't find every proposition here intuitively compelling. But that just means that it isn't as compelling a paradox for those people, not that we can't describe it as a paradox. And for people for whom all three propositions are intuitively compelling, it is a very pressing paradox.

Now one way *not* to solve this paradox is simply to announce to the world which of the propositions that you plan to reject. That is, you can't solve the paradox this way unless you think that making this announcement is sufficient to undermine whatever intuitive case there was for that proposition there was. If, for example, we plan to solve the paradox by rejecting (P3), as I guess a few theists still attempt to do, we have to do a *lot* more work than simply denying it. We have to do a lot of work to show why our initial belief that it was true was mistaken. And if we end up with a solution to the paradox, *all* the work in the solution will have been done by this explanation, and little or none by the mere announcement that it is (P3) that is false.

It seems to me that, more or less, all Keefe does in the above passage is announce which is the paradoxical propositions she rejects, and issues a promissory note about an eventual explanation of the permissibility of this move. It is as if the theist had said, "Supertheism says that this is the best of all possible worlds. (But objections to the supposed truth of sentences such as (P3) will be considered in §5.) So supertheism solves the paradox of evil." More or less, because Keefe's position is *even worse* in one respect than the theist, but when we get to the explanation, it is a little better than your average apologetic.

Unless we start giving up on classical logic, or asserting that blue light is red, we are forced very quickly to the view one of the premises in the first Sorites argument I gave is not true. The obvious task, then, is to explain why they all seemed true. Maybe this will be a useful move in the end, but I'd have thought it unwise to start the explanation by asserting that in fact *very very many* of them are not true. Do that and you're just making life harder for yourself. But that is exactly what the supervaluationist, or at least the supertruth supervaluationist, does. Note that according to Keefe, whenever a sentence is not true on some precisification, it is not true *simpliciter*. And whenever *x* nm is in the penumbra between red and not-red light (on the low wavelength side), the conditional *If light of x nm is red, then light of x* - 0.1 *nm is red*, will fail to be true on at least one precisification. So hundreds, maybe thousands, of the premises in that argument are not true. It really is an unsound argument! Maybe it will all turn out for the best, but this doesn't feel like progress to me. Note that if we take truth to be truth$_T$ rather than supertruth, we do not have this problem: the number of premises that are not true is back down to one. If only the truth$_T$ theory were coherent…

When Keefe gets to the explanation for the surprising existential claims that supervaluationism makes, she is on safer ground than the theists to whom I was comparing her. After some initial skirmishing regarding the fact that since this is a paradox we have to say *something* counterintuitive, she settles on a defence of these existentials that is due largely to Kit Fine. Consider the *D* operator we discussed above, which produces sentences meaning roughly "It is definitely the case that…" When we consider a normal subject-predicate sentence *a is F*, we should only be happy to assert this if *a* is definitely *F*, or, symbolically, if *DFa*. We don't say of a penumbral case of *F*-ness that it is *F*; we only say of definite cases of *F*-ness that they are *F*. You might have thought that this was because, in general, we assert sentences only when we think they are definitely true. This looks to be mistaken. In fact, it fails in just the way supervaluationists need it to fail to explain the (merely!) apparent falsity of these surprising existential claims.

We need two claims about existentially quantified sentences and to get out of this difficulty. For simplicity, say that a pair of light rays, a *G*, is *F* iff one is red and the other is not red. Intuitively, we accept *No G is F* and reject *Some G is F*. This can be explained if the following three claims are true:

(a) When we reject *Some G is F*, we are rejecting its proper assertability, not its truth;
(b) We accept *No G is F* because we reject *Some G is F*
(c) The criteria for accepting *Some G is F* as being properly assertable is that *Some G is DF* is true.

So (a) and (b) are claims about what the intuitions we have about these sentences should be taken to be evidence of. In short, the supervaluationist takes themselves to have explained the intuitions if they have provided a theory on which the intuitions are *correct* once it is realised what they are about, which is proper assertability. (This kind of move is very common - it's at the heart of Grice's pragmatic explanation of the Wittgensteinian data he wanted to explain away.) And (c) is a special instance of a rather plausible principle concerning proper assertions. I won't go into that here because we'll have a whole section on that later in the course. Suffice to say, the kind of theory that Keefe is relying on here works. In fact it works spectacularly well.

None of this undercuts my earlier criticism. If the supervaluationist has a solution to the sorites, and I'm still waiting for a story about why so many premises are not true, or for that matter what we should say about the forced march sorites, it is in the details of this Gricean story, not in the comments Keefe makes above.

3.6. Higher-Order Vagueness

At first glance, it seems supervaluationism cannot be applied to all sentences. If it is, we at least seem to get the absurd result that *Every sentence has precise truth conditions* 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. I am sure that if we had to rely on such rules of interpretation, we 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. Or even better, it would be nice to have a theory of how and when people make these judgements not to apply the supervaluational theory, and say *that* is our official theory of vagueness.

Perhaps, though, things are not as they seem. Assume that we have in the object-language a 'definitely' operator, so that *Louis is definitely bald* is true just in case *Louis is bald* is true on all precisifications. We will assume, somewhat contrary to initial syntactic impressions, that *definitely* is a sentential operator. And it will be important in what follows that *definitely* is an operator, rather than a predicate. It takes sentences as inputs and has sentences as outputs, if it were a predicate it would take names of sentences as inputs and produce sentences as outputs, just as any other predicate, like 'is tall' takes names of entities as inputs and produces sentences as outputs.

Also assume that we have an 'accessibility' relation between precisifications. To help get clear on the picture, think that one precisification is accessible from another if the second resembles the first in some salient respects. It is important that this is only a picture, not a definition, because we may want to reject the apparent implication that the relation is symmetric. Now say that the sentence *Definitely S* is true on a precisification iff *S* is true on all precisifications accessible from *S*. Finally, say that a sentence *S* has precise truth conditions iff *Definitely S or Definitely not S* is a necessary truth. This will be true if the accessibility relation is the identity relation, which is effectively what I assumed above. (And, it seems, what Lewis was assuming if he thought that we needed a pragmatic explanation for the oddity of *Every sentence has precise truth conditions.*) But it will not be true in general. So if this picture, which is the picture Williamson suggests supervaluationists accept as a theory of higher-order vagueness in §5.6 of *Vagueness*, is right, then it is not true in general that every sentence has precise truth conditions.

In short, we analyse higher order vagueness by taking *Definitely* to be a modal operator, with precisifications playing the role of points in a Kripke-model for the logic KT. In such a model, we have a set of points, W, a reflexive accessibility relation R on the worlds, and a valuation function V from atomic sentences to subsets of W. An atomic sentence p is true at a point $w \in W$ iff $w \in V(p)$. We then provide a recursive definition of truth for compound sentences. The clauses for the truth-functional connectives are familiar from Tarski's theory of truth. And the clause for the modal operator, traditionally a box but here a D (for *Definitely*), is that DA is true at w iff for all w' such that wRw', A is true at w'. This seems to give us everything we want in a supervaluational account of higher-order vagueness. Every substitution-instance of classical theorems from the D-free language is a logical truth, as is $DA \rightarrow A$, but neither (8), (9) nor (10) is a theorem.

(8) $DA \vee D\neg A$
(9) $DDA \vee D\neg DA$
(10) $DD\neg A \vee D\neg D\neg A$

If (8) were a theorem, then for any sentence *S*, it would either be definitely true or definitely false, so there would be no first-order vagueness. If (9) were a theorem, then for any sentence it would either be definitely definitely true, or definitely not definitely true, so there would be a sharp boundary at the 'upper end', between the cases where *A* is definitely true, and the cases where *A* is not definitely true. If (10) were a theorem, then for any sentence it would either be definitely definitely false, or definitely not definitely false, so there would be a sharp boundary at the 'lower end', between the cases where *A* is definitely true, and the cases where *A* is not definitely true. If (9) and (10) are both theorems, then any kind of higher-order vagueness would seem to be ruled out. (Keefe, as we will see, denies this, but there is clearly a *prima facie* case that it is so.) Hence it is rather pleasant to have a formal theory where none of these troublesome claims are logical truths.

Despite these initial attractions, it is not entirely clear that this approach will deliver everything that supervaluationists desire. It might be wondered just where in this formal model we are to find natural languages like English. Consider, for example, what happens if we identify English with one of these points. Let *S* be any sentence that in English is intuitively indefinite, and let *w* be the point that we identify with English. If from *w* there are accessible points where *S* is true, and other accessible points where *S* is false, then neither *DS* nor *D¬S* will be true at *w*, which seems to be the result that we want. Note, however, that either *S* or *¬S* will be true at *w*. Each of these points are complete: every sentence is either true or false at each of them. And this is not what supervaluationists want. If English is *w*, then there is a fact about whether *S* is true, it just might not be a definite fact. As Williamson has stressed, this idea of a fact that is not a definite fact is a little hard to comprehend, if *Definitely* is meant to be a *semantic* operator. (It is easy enough to interpret if *Definitely* means something like *Knowably*, as Williamson will argue it does.) Even if we can comprehend it, this is not a position that supervaluationists should find attractive. The position is not just that certain sentences, like (8), are false. It is that there are no facts of a certain kind, so a formal model which allows that there are these facts (even if they aren't reflected in the behaviour of object-language operators like *Definitely*), is unacceptable.

What if, instead of identifying English with a point *w* in the model, we identify it with a set of points *E* in the model? The idea here is that natural languages are not the analogues of possible worlds, in a familiar interpretation of the Kripke semantics, but of what Humberstone calls possibilities. As Humberstone shows, we can provide a fairly natural semantics for truth according to a possibility, which does not guarantee that any sentence *S* is either true according to a given possibility or false according to that possibility. (For our purposes we can take possibilities to be sets of possible worlds, though this is not Humberstone's purpose. He shows how we can take possibilities to be primitive and do away with possible worlds, and provides some interesting reasons for thinking that we should do this.) Now it seems we do not have the troubling result that a sentence is either true or false in the language, even if we cannot state which, or even the existence of this fact, in the object language. We do, however, have a result that should be equally troubling. Now for any sentence, there is a fact about whether it is true, or false, or neither true nor false. True, we cannot state the existence of this fact by asserting the conjunction of (9) and (10), just as in the earlier picture we could not state the existence of the meddlesome fact using (8). But if English really is *E*, and truth in English just is truth throughout *E*, then there will be such a fact, and its existence rules out any kind of higher-order vagueness.

Well, perhaps we can say that English is to be identified with a set of precisifications, but there is no fact of the matter as to which it is. This avoids the earlier problems, but at some cost. Now we have a harder time saying just what it is for a sentence to be true in English. We have a theory of what it is for a sentence to be true on a precisification, or on a set of precisifications, but if English is not (definitely) any one of these, then the question of what it is for a sentence to be true in English has been left unresolved. The natural response here is to adopt a kind of supervaluational approach. If there is a set *E′* of sets of precisifications, and it is indefinite just which element of *E′* is English, then we might say that a sentence *S* is (definitely) true iff it is true in every element of *E′*, (definitely) neither true nor false if it is neither true nor false in every element of *E′*, and (definitely) false iff it is false in every element of *E′*. This, of course, will not do because, for reasons that should be obvious by now, it rules out the possibility of third order vagueness.

What seems to be needed is a theory in which English is a set of precisifications, but there is no fact of the matter as to which set it is, and no fact of the matter as to which is the set of sets between which there is no fact of the matter which set it is, and no fact of the matter as to which is the set of sets of sets between which…and so on for all orders of vagueness. This is a mouthful (and then some), but it seems to allow *n*th-order vagueness for every natural number *n*, which is an improvement over what has come so far. Unfortunately, there are a few objections, five or so of which are presented here. (‘Or so’ because individuation of objections is a tricky business - maybe we should come back to this when we discuss vague identity.)

Objection One: The formalism, and by extension the supervaluational theory, is doing no work. All the theoretical work is being done by the relationship between the formalism and the language.

Mark Sainsbury presses something like this objection to a similar proposal, and it seems to have some merits. As it stands, the theory relies on our making sense of the nature of the relationship between English and these sets of sets of precisifications, such that it is a fact that English is one or other of them, but not a fact that it is any particular one. In other words, it looks like the theory requires us to make sense of something like ontological vagueness, when all it has told us is how to process linguistic, or at best semantic, vagueness. Even if we do implicitly understand how to do this, the task of a philosophical theory is to make explicit these implicit understandings, and the theory fails to do this, unless it can be extended from a theory of vague representations to a theory of vague facts as well.

Note that it is not much good to say *here* that the theory is at least better than a trivial theory that only explicitly utters contentless sentences, and leaves all the work to our implicit understanding of these sentences. The imagined responder says that there is still a lot of theory to go once we get past this mysterious connection between the formalism and real languages, and that theory might do *some* work. (Keefe makes this kind of response to a Sainsbury-style criticism of a *similar*, though I stress not *identical*, proposal. We shall discuss Keefe’s own proposal below.) The objection is that the theory does not meet one of its duties. That it could have been even more negligent is not a great defence. (“I admit I was speeding Your Honour, but at least I was sober while doing it.”)

Objection Two: On this theory, English becomes a vague object, and Evans showed that there can be no vague objects.

A full discussion of this objection would require a full discussion of Evans-style arguments against vague identity, and that will be a chapter of its own later, so this treatment will be a little quick. The core of the Evans argument is that there cannot be definitely vague identity claims where we have referring (as opposed to merely denoting) terms either side of the identity sign. The proof is a snappy reduction - assume that $a = b$ is such a sentence. Then since *a* is indefinitely identical with *b*, but *b* is not indefinitely identical with itself (it is definitely identical with itself, after all), then $a \neq b$ by Leibniz's Law, and since every premise in this argument was definitely true, the conclusion is definitely true as well, refuting the possibility that $a = b$ might have been indefinite. Let E_N be a language with first-order vagueness but no higher-order vagueness, such that it is represented in our formalism by the set E of precisifications, and such that E is a set such that it is indefinite whether English is represented by that set. Now there seem to be only three options here, none of them particularly happy: either *English is* E_N is true, false or indefinite.

We can rule out the first one pretty quickly, since English has higher-order vagueness and E_N does not, the two are not identical. The second is not exactly great. Remember that it is indefinite whether *every sentence* in English is synonymous with the homonymous sentence in E_N. (This is because it is indefinite whether English is represented by E, and if it is, then every sentence in English is synonymous with the homonymous sentence in E_N.) So if it is false that the two languages are identical, then it cannot be true that any two languages in which homonymous sentences are synonymous are identical. To be fair, this principle might still be indefinite, but it is a little unfortunate that it cannot be true.

What of the third option, that the sentence *English is* E_N is indefinite? This breaks down into two options. First, we could say this is indefinite despite containing two referring terms, and hence that the Evans argument fails somewhere. This is a fairly popular option amongst proponents of deviant logics (see e.g. Parsons 2000) but should not be a happy result for supervaluationists. Alternatively, we could say that one of the two terms is not really a name, but is shorthand for a description or some other part of language. (I have no idea what it could be if not a description, but let's keep our options open, shall we?) This might be the best way out. As we shall see in chapter 10, it is rather hard for supervaluationists, or even their fellow travellers, to accept that there are vague proper names, because of the problems posed by Evans-style argument. If supervaluationists must accept there are no vague names, then accepting that our own language cannot be named, though it can be described, might be an acceptable option here.

Option Three: Introduce a new operator D^* into the language such that D^*A is true iff A is true, and DA is true, and DDA is true, and so on. One might well assume that there will be vague sentences of the form D^*A, but there is no way to represent this vagueness within the formalism.

This is one of the major arguments Williamson runs against this view of higher-order vagueness, but it is not one that strikes me as particularly powerful. (And I don't think the equivalent argument against Williamson's own theory, as developed by Mario Gomez-Torrente and Delia Graff, is particularly persuasive either. We shall get to that argument in the next chapter.) There are four ways out, and at least two of them strike me as being perfectly acceptable.

First, we could insist that D^*A is still susceptible to a kind of vagueness. It is a theorem that $D^*A \rightarrow DD^*A$, and this a major reason why it is thought that D^*A could not be vague. But it is not a theorem that $\neg D^*A \rightarrow D\neg D^*A$, so we cannot infer from the (logical) truth that $D^*A \vee \neg D^*A$ to the claim that D^*A is not vague: $DD^*A \vee D\neg D^*A$. As Williamson notes, there are two reasons for being unhappy with this response. First, although there can be a failure of determinacy here, it is strictly one-sided. If it is indefinite whether D^*A is true, then D^*A is not true. (Contraposing one of the theorems about, we get $\neg DD^*A \rightarrow \neg D^*A$.) It is odd, at least, to have an indeterminacy with this characteristic. Secondly, if the initial accessibility relation R is necessarily symmetric, then $\neg D^*A \rightarrow D\neg D^*A$, and hence $DD^*A \vee D\neg D^*A$, suddenly are logical truths. And there seems to be no good grounds for denying that R should be symmetric. So this response is perhaps not the best response we have.

Secondly, we could say that for any contingent claim *A*, *D*A* is just false. There is no sharp boundary between the women who are definitely* tall and those that are not, because none are definitely* tall. (Or there is a sharp boundary 'at infinity'. Whether one thinks of this as a sharp boundary or not seems to be a relatively uninteresting terminological dispute.) One can argue for this claim using a 'margin-of-error' principle for definiteness, modelled on Williamson's margin-of-error principles for knowledge. For some shorthand, say *x* is definitely^1 tall iff *x* is definitely tall, and definitely^{n+1} tall iff she is definitely definitely^n tall, for any natural *n*. Then the following principle has some plausibility: If *x* is definitely^{n+1} tall, then a woman only one nanometre shorter than *x* is definitely^n tall. Despite initial appearances, accepting this principle does not commit us to any Sorites problems. We cannot use this to argue from the fact that a woman of height two million and two nanometres is definitely definitely tall to the claim that a woman of height one million and two nanometres is definitely definitely tall. By applying the principle once we get that a woman of height two million and one nanometres is definitely tall, and by applying it again we get that a woman of height two million nanometres is tall, but these are hardly implausible claims. And it quickly follows from the principle that for any finite height, a woman of that height is not definitely* tall. There may be some difficulty with predicates that have natural stopping points (e.g. bald, short), but in general it doesn't seem too bad to reject the possibility of borderline cases for *D*A* by rejecting the possibility of positive cases for it, where *A* is any contingent claim.

Thirdly, we could just say that it is impossible for *D*A* to be vague for more traditional reasons, i.e. that there is a sharp boundary between, for example, the determinately* tall women and the not determinately* tall women, and say that our intuition that it could be vague is just caused by our inability to think clearly about the infinite. Of course, there are other possible explanations for this intuition, perhaps that we are ignorant of where the boundary lies, but this one strikes me as the most plausible. Williamson says that this would be a concession that vagueness is not a 'deep' phenomenon, but this seems a trifle unfair. After all, we are conceding that there are infinitely many layers of vagueness - isn't that deep enough?

Finally, and this is the option Williamson advises supervaluationists to adopt, we could say that there is a kind of vagueness in *D*A*, but this cannot be represented using the same operator. The idea is that we adopt a new operator, *D!*, and say that for *D*A* to be precise $D!D^*A \vee D!\neg D^*A$ would have to be a theorem, which it is not. Williamson notes that we will have to adopt a new operator to explicate the vagueness in *D!**, and then another for the next layer of vagueness, and so on, so this kind of move will massively increase the size of our languages. Keefe suggests, plausibly to my mind, that if we are to introduce new connectives to represent the vagueness in *D*A*, we could just have started by introducing new connectives to represent the vagueness in *DA*, rather than representing this by iterating one connective. So this option does not look like the best one for supervaluationists. But the second and third responses did seem plausible, so this objection does not look too telling.

Objection Four: When we tried to embarrass supervaluationists by noting that their theory abandoned the T-schema, they said that at least the two sides of the schema entailed the other. But now we don't even have the mutual entailment. How much will it take to embarrass some people?

The objection assumes a couple of things, which will probably be obvious to most readers. First, it assumes that validity is truth-preservation. Secondly, it assumes that a single-premise argument *A*, therefore *B* is truth preserving iff the conditional *If it is true that A then it is true that B* is necessary. Thirdly, it assumes that *It is true that A* is true at a point in the model iff *DA* is true at that point. These do not seem like substantive assumptions in this context. The first is something that supervaluationists accepted, the second looks like a definition, and the third is the supervaluationists' own definition of truth. (Why identify *It is true that A* with *DA* rather than with *A* alone? Because we are trying to develop a theory in which there are truth-value gaps, which means that we must be working with the truth as supertruth theory, not the truth as truth$_T$ theory, and *A* is supertrue iff *DA*.) Since *DA* → *DDA* is not a theorem, the logic does not guarantee that *A*, therefore *A* is true is truth-preserving, so this argument is not valid. I played down the significance of the mutual-entailment interpretation of the T-schema above, so by rights I shouldn't make too much of this option. But it should be noted that having no account of the T-schema is worse, perhaps by an order of magnitude, than having a weak account of it like the mutual-entailment interpretation.

Objection Five: First-order vagueness is *sui generis* on this approach.

The theory has us account for first-order vagueness by representing natural languages as sets of precisifications. It then accounts for higher-order vagueness by saying there is no fact of the matter as to which set of precisifications it is. Why the bifurcation? Why not represent the language as a single precisification, with a *D* operator that behaves as described, and say that *all* vagueness comes about from there being no fact of the matter as to which such precisification it is? Such an approach may not even have the weak response we offered to objection one (the 'it is only contributory negligence' response), but we could compensate for that if the theory has other theoretical virtues. And having a unified treatment of first-order vagueness and higher-order vagueness seems to be such a virtue.

In fact, if we go this way, we end up back in a fairly familiar position. Presumably if English is a precisification *P* (though there is no fact about which one), then a sentence is true in English iff it is true in *P*. Since *P* is complete, this means that we have ruled out borderline cases. (Again.) We are back, in effect, to the position that truth is truth$_T$. So we have another theoretical advantage for this interpretation of truth over the account of truth as supertruth - it permits a unified treatment of higher-order and first-order vagueness. The point that within a supervaluational framework it is incoherent to suppose that truth is truth$_T$ still stands, but the supposition looks pretty good, apart from its incoherence. As we shall see in chapter 9, by throwing away parts of the supervaluational apparatus, we can remove the incoherence, so this point is worth remembering.

Rosanna Keefe offers a slightly different account of higher-order vagueness. As we noted, she says that if supervaluationists take Williamson's advice over objection three, they could have made this move much earlier, and had different concepts of definiteness for each different order of vagueness. One reason that Keefe makes this move is that she is worried by objection four - she accepts that if we take truth to be supertruth (as she does) then the argument *A*, therefore *It is true that A* is invalid unless *DA* → *DDA* is a logical truth. And she accepts that if this is a logical truth, then we cannot represent the vagueness in *DA* by using the same operator *D*. So we have to have a new operator for representing this vagueness. All the moves in the reasoning here look correct, except perhaps for the decision to use modus ponens at every step rather than occasionally reaching for modus tollens. Keefe's position ends up with a hierarchy of definiteness operators, like the hierarchy of truth operators in a Tarskian theory of truth. And it seems to face the same kinds of objections that are raised against Tarskian theories of truth (by, for example, McGee 1990 and Soames 1998). First, the hierarchical model posits a kind of ambiguity in our word *definitely*, but there is no evidence for this ambiguity. A sentence like *What Jack said is definitely true, and so is what Jill said*, is not even mildly zeugmatic, even if we know that Jack said some ordinary sentence, like *Grass is green*, and Jill said something involving the (first-order) operator *definitely*, like *Grass is definitely green*. Secondly, we can use the term *definitely* (and the term *true*) to pick out concepts outside this hierarchy, as we seem to do in *Everything that God believes is definitely true*. So this approach seems unlikely to succeed.

There is one last, seemingly desperate, option for supervaluationists to take. They could simply deny that there is any higher-order vagueness. As I noted above, it seems to be a live option to accept finite-order vagueness, and explain away the appearance of infinite-order vagueness. Perhaps it is also a live option to accept first-order vagueness and explain away the appearance of second-order vagueness. Indeed, Koons 1994 does just this. He suggests that there are really borderline cases of various concepts, but there are no borderline borderline cases, we simply do not know where the borderline cases start. Now it might seem like there is a rather compelling response to this kind of position. If there really is a sharp boundary between the definite cases of tallness (or whatever) and the borderline cases, why couldn't we cut out the middleman and say that there is a sharp boundary between the tall and the not tall? If the only reason we had for believing there is no such sharp boundary is our inability to locate it, then this rhetorical flourish would seem to constitute a more or less compelling objection to Koons's approach. To close this chapter, I want to note a way for a Koons-style approach to survive such a challenge. What I will suggest is that the most powerful argument for believing in first-order vagueness does not *obviously* extend to an argument for believing in higher-order vagueness.

As far as I can tell, the following argument schema, due to Ted Sider, captures the essence of the most powerful arguments that a term T is semantically indeterminate.

1. There exist multiple candidate meanings for T, corresponding to the conflicting theories about T
2. None of these T-candidates fits *use* better than the rest
3. None of these T-candidates is more *eligible* than the rest
4. No other T-candidate combines eligibility and fit with use as well as these T-candidates
5. Meaning is determined by use plus eligibility

C. *Therefore*, T is indeterminate in meaning among T-candidates corresponding to the conflicting theories of T, and so there is no fact of the matter which of these theories is correct. (Sider 2001: 189-90)

The concept of *eligibility* that is in play here derives from David Lewis's response to Putnam's model-theoretic argument against realism, and Kripke's argument for massive semantic indeterminacy. The thought is that there is something *natural* about certain properties and objects that makes them, as opposed to certain Skolemised or Goodmanised variants, the references of our predicates and names. The concept of naturalness could do with some more spelling out – it would be nice to know, for instance, just how it relates to ideal physics, or whether it can play a role in solving Goodman's riddle, or finding the one true Carnapian *c*-function. But I have nothing of any interest to say on these points, so I just refer the interested reader to what Lewis and Sider have said about eligibility in other places.

The argument, especially at premise 5, also draws heavily on those responses of Lewis. The idea behind it is that while Putnam is wrong to say that "We interpret our language or nothing does" (Putnam 1980: 482), his instinct is in the right place. While we aren't the only source of content, the list of sources of content should be quite short. There isn't a magical source of content that makes 'mass' in Newton's language mean *rest mass* rather than *inertial mass*. The world outside the head has a role to play in determining content, even of what's in the head, but its role is limited to privileging certain objects and properties, not, say, providing primitive semantic laws. As the old saying goes, intentionality doesn't go that deep.

Sider intends this argument to apply to various *indeterminate* terms, like 'same person' as it is used in the metaphysics of personal identity, or indeed 'mass' as it was used in Newtonian mechanics. The argument generalises easily enough to show that vagueness generates semantic indeterminacy.

Consider a typical vague term, like *tall for a woman*. If we think there is no semantic indeterminacy in this expression, we must think that there is a height *x nm* such that a woman is tall iff she is at least *x nm* tall. But for any plausible candidate meaning of this form, there is a rival candidate meaning, one that makes a woman tall iff she is above *x*+1 *nm* tall. And that rival will do just as well as capturing our uses of *tall*, even including our dispositions to use the word. And presumably it will be just as eligible a referent as the original, whatever eligibility comes to. So, Sider's argument teaches us, the meaning of the word *tall* must be indeterminate between, *inter alia*, these candidates.

Note, however, that the argument does not carry over as easily to an argument against the determinacy of higher-order expressions. We can obviously substitute *definitely tall* for T throughout the argument and get a conclusion that *definitely tall* is indeterminate in content, and hence there is higher-order vagueness. The problem is that, when we make this substitution, premise three no longer seems as plausible. The function from uses to meanings (or, better yet, to acceptable precisifications) is more natural than its nearby neighbours. After all, it is relevant to good semantic theorising; the others just get in the way. If we think there is a theory of meaning to be found, then we should think there is a reason why the acceptable precisifications are acceptable and why the unacceptable ones are not. This reason is sufficient for the one true function from use to meaning to be natural and its rivals to be unnatural. And this naturalness might be sufficient to determine the boundary between the determinately tall and the not determinately tall.

How plausible is the theory with which we end up? I am inclined to think *not very*, but I don't have a powerful argument for this, just a feeling. Philosophers tend not to like 'mixed' solutions to problems, and the problems associated with vagueness are no exception. But are there reasons for this dislike, or is it just a prejudice? It is worthwhile to read Graham Priest's arguments from the principle to uniform solution to the truth of various contradictions to get a full appreciation of the consequences of this methodological move. Having said that, I will try to look for a theory that treats first-order and higher-order vagueness alike.