7

Higher-Order Vagueness for Partially Defined Predicates

Scott Soames

1. Background

In this paper I will talk about a perplexing problem that arises for the theory of vague and partially defined predicates that I sketched in my book *Understanding Truth*, and which can, I think, be expected to arise for other theories that employ partially defined predicates.[1] The problem is that of making sense of so-called *higher-order vagueness*. This problem is often regarded as the chief difficulty facing analyses which treat vague predicates as partially defined. Although I can't claim to have solved the problem, I will argue today that it is more tractable than it is often taken to be. I begin by rehearsing the basic framework that I presuppose.

The central idea is that vague predicates are context-sensitive and partially defined. To say that a predicate P is partially defined is to say that it is governed by linguistic rules that provide sufficient conditions for it to apply to an object, and sufficient conditions for it to fail to apply, but no conditions that are both individually sufficient and jointly necessary for it to apply, or fail to apply. Because the conditions are mutually exclusive, but not jointly exhaustive, there will be objects not covered by the rules for which there are no possible grounds for accepting either the claim that P applies to them, or the claim that it does not. P is said to be *undefined* for these objects. Its *extension*

Thanks to Alexis Burgess for comments on an earlier draft.

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

is the collection of things to which it applies, and its *antiextension* is the collection of things to which it doesn't apply. The system is disquotational in that for any name *n*, we accept the statement ⌜'P' applies to *n*⌝ just in case we accept ⌜'P*n*' is true⌝, which we accept just in case we accept ⌜P*n*⌝.[1] When P is undefined for the referent of *n*, we do not accept ⌜P*n*⌝, ⌜'P*n*' is true⌝, or ⌜'P' applies to *n*⌝, nor do we accept the negations of these claims. We regard it as a mistake to do otherwise, since (i) none of these claims is a necessary consequence of the set of underlying non-linguistic facts together with the rules of the language governing the expressions they contain, and (ii) given the rules governing the predicates, even one who was omniscient about all non-linguistic facts would have no grounds for accepting them.

A distinction is made between the extension of P and its determinate extension, the latter being the set of objects *o*, such that the claim that P applies to *o* is a necessary consequence of the rules of the language plus the set of underlying non-linguistic facts. This distinction results from the fact that there are some objects *o*, such that the claim that *o* is not in the determinate-extension of P is true, whereas the claim that *o* is not in the extension of P is to be rejected because the predicate ⌜is in the extension of 'P'⌝ is, like the predicate P, undefined for *o*. Similar remarks apply to the distinction between the antiextension and the *determinate-antiextension* of P. Corresponding to these distinctions, there is also a distinction between truth and *determinate truth*.[2]

In addition to being undefined, vague predicates are also context-sensitive. Given such a predicate P, one begins with a pair of sets. One, the *default determinate-extension* of P, is the set of things to which the rules of the language, together with the underlying non-linguistic facts, determine that P applies. The other, the *default determinate-antiextension* of P, is the set of things to which the rules of the language plus the underlying facts determine that P does not apply. For all objects *o*, P is undefined for *o* just in case *o* is in neither of these sets. Since these sets don't exhaust all cases, speakers have the discretion of adjusting the extension and antiextension so as to include initially undefined cases. Often they do this by explicitly predicating P of an object *o*, or by explicitly denying such a predication. When a speaker does this, and other conversational participants go along, the extension (or antiextension) of the predicate in the context is adjusted so as to include *o*, plus all objects that bear a certain relation of similarity to *o*.

[2] See ibid., ch. 6.

We can illustrate these points with the help of an example. The model is clearest and most intuitive with simple observation predicates like *blue*. A characteristic feature of these predicates is that we learn them not by being given verbal definitions, but by being given clear and obvious examples of things to which they apply and things to which they don't. We are told when presented with some reasonable range of objects *this is blue* and *that is not*—or, if we are not explicitly told, we note that there are certain objects that everyone we encounter seems ready to call *blue* and other objects that everyone we encounter seems ready to characterize as *not blue*. These learning experiences give rise to beliefs about conditions for proper application of the predicate. Think of it this way: People say of a certain object that it is blue. We observe the object, which is perceptually represented to us as being a certain shade of color. Call this shade *B1*. They say of a different object that it is not blue. We observe that object, which we perceive to be of a different shade—call it *NB1*. On the basis of experiences like these, we form the belief (which virtually everyone we encounter seems to share) that objects of the first shade—B1—are objects to which *blue* applies, and objects of the second shade—NB1—are objects to which the predicate does not apply. We may idealize this situation by saying that we first entertain, and then come to accept, the hypothesis that the following pair of rules governs the application of the predicate *blue* in the language.

Blue 1

(a) If an object is B1, then *blue* applies to the object
(b) If an object is NB1, then *blue* does not apply to the object.

In saying that the agent first entertains, and then comes to accept, the hypothesis that these rules govern the predicate in the language of his community, I don't mean that the agent formulates these rules in words. Most likely the agent has no words, at least no non-indexical words, that stand for these specific shades. Rather, he comes to accept the propositions expressed by (a) and (b). Other learning experiences with *blue* lead the agent to accept other pairs of rules, involving different shades, as governing the predicate, as well. At some point in this process, the agent is counted as having successfully learned the meaning of the word, as it is used in his linguistic community. At this point, the agent will have accepted a set of rules Blue 1—Blue *n*, the (a) versions of which provide sufficient conditions for *blue* to apply to an object and the (b) versions of which provide suffi-

cient conditions for *blue* not to apply to an object. However, although these conditions will be mutually exclusive, the requirement that they be mutually agreed upon and generally adhered to by the overwhelming majority of speakers, no matter what the context, will ensure that they are not jointly exhaustive. Since there will be shades of color, and objects having those shades, about which the rules say nothing, the rules do not provide a set of conditions which are individually sufficient and jointly necessary for *blue* to apply to an object, or for it not to apply. This illustrates the partiality of the predicate in the language of the speaker's community.

Context-sensitivity results from the fact that speakers have the discretion to apply the predicate *blue*, or its negation *not blue*, to objects for which it is undefined by the rules of the language. Often they do this by asserting that some contextually salient object is blue, or that it isn't. Consider the positive case. If the other conversational participants accept the characterization of the object *o* as blue, then the extension of the predicate in the context is adjusted to include *o* and all objects that bear a certain relation of similarity to it. In general, the relation involved in these contextual adjustments is determined by the meaning of the predicate together with the intentions of speakers and hearers in the context. Putting aside various complications, let us suppose that when an agent characterizes as blue an object *o* for which the predicate is initially undefined, he adopts a contextual standard that counts *o*, all objects uncontroversially regarded to be bluer than *o*, as well as all objects that are pairwise indiscriminable from *o* by ordinary observation in good conditions, as being blue as well. Let B*c* be a particular shade that applies to precisely this class of objects. We may then characterize what has happened in the context as a result of the speaker's predicating *blue* of *o*: as a result of doing this, the speaker has adopted a rule governing *blue* in the context that contains the following condition for positive application of the predicate.

> If an object is B*c*, then *blue* applies to it.

Although not a rule of the language governing the predicate, this rule is one that speakers are free to adopt at their discretion in particular contexts of utterance.

We have now illustrated both parts—partiality and context-sensitivity—of the semantic analysis of vague predicates that I will presuppose in what follows. In my opinion these two features of the analysis naturally go

together, and are mutually reinforcing. Given an analysis that posits one, we can find substantial reasons for adopting the other as well.[3]

2. Consequences for the Sorites Paradox

This brings me to the Sorites paradox. Since semantic theories of vagueness are often judged by the solutions they provide to the paradox, I will say a few words about this. Although all Sorites predicates are vague, not all vague predicates are natural Sorites predicates, with application conditions based on the position of objects in a more or less single and unified underlying continuum. Since the semantic analysis of vagueness is intended to apply to all vague predicates, it should be motivated to a substantial degree by considerations independent of the Sorites. Any light it sheds on the paradox is an extra benefit. In the case of the analysis I advocate, there are two general consequences that the model has for the paradox. First, the fact that vague predicates are partially defined means that the semantic categorization imposed on the world by such a predicate will include more than two categories. There may well be sharp and precise lines dividing the objects in different categories, but typically these lines are not properly characterized as separating objects to which the predicate applies from those to which it does not apply. Second, context-sensitivity tells us that the lines are movable. When one looks closely at the mechanisms by which these lines are adjusted in conversational contexts, one finds that in many cases the mechanism makes it practically impossible to display them; any attempt to display the precise line dividing objects to which the predicate applies (or doesn't apply) from objects for which it is undefined has the effect of moving the line elsewhere. This constant and elusive movement creates the illusion that there are no sharp lines to be drawn.

For example, let us assume that the predicate *blue* is partially defined, with a default determinate-extension and a default determinate-antiextension, plus a range of objects for which, absent temporary conversational adjustments, it is undefined. Suppose further that the conventions governing the predicate include constraints on how its extension and antiextension may be adjusted within this range. In particular, it is accepted that, typically, one who

[3] This is argued in the final section of my 'Replies' in the symposium *Understanding Truth* in *Philosophy and Phenomenological Research*, 65/2 (Sept. 2002).

explicitly characterizes something x as blue on the basis of ordinary perceptual evidence is, all other things considered equal, committed to a contextual standard that counts all objects that look bluer than x, plus objects perceptually indistinguishable in color from x (when paired with x and viewed together) as blue. Finally, suppose that two stimuli can be perceptually indistinguishable in this sense even though they differ slightly in the physical characteristics that cause them to look blue. Given this supposition, we can construct a sequence connecting x_1—which definitely looks and is blue—to x_n—which definitely is not, and does not look, blue—in which any two adjacent items in the sequence are (pairwise) perceptually indistinguishable in color. When an agent characterizes an object x_i—for which the predicate is initially undefined—as blue, the (determinate) extension of the predicate is adjusted to include x_i, all earlier items in the sequence, plus x_{i+1}—which is perceptually indistinguishable in color from x_i. As a result of this adjustment there is now a sharp line between x_{i+1} and x_{i+2} separating items to which the predicate determinately applies (in the context) from items for which it remains undefined. However, if one attempts to display this line, by showing the agent x_{i+1} and x_{i+2} together, and asking him to characterize them, he will, quite properly, resist the invitation to treat them differently. For if he now explicitly endorses his previously implicit commitment to counting x_{i+1} as blue, then his assertion that x_{i+1} is blue will have the immediate effect of adjusting the contextual standards so as to count x_{i+2} as blue as well. By focusing on and making judgements about what had been the line separating objects to which the predicate (determinately) applied from those for which it was undefined, the agent has imperceptibly moved the line, thereby engendering the illusion that there was no sharp and precise line in the first place.

In my opinion, this analysis has illuminating implications for different versions of the Sorites paradox. For example, in dynamic versions of the paradox an agent presented with a Sorites sequence about which he is asked to make judgements can easily be pressured into making a series of positive claims $\ulcorner x_1$ is F, x_2 is F, ..., x_i is F$\urcorner$ that comes to an end when he refuses to go further, and either assents to a negation $\ulcorner x_k$ isn't F$\urcorner$ or refuses to make any judgement at all. At this point, pressure can be generated in the opposite direction, with the result that the agent will dissent from, or withhold judgement on, sentences $\ulcorner x_j$ is F$\urcorner$ to which he previously assented. The semantic model of vague predicates just sketched indicates how and why this pressure is generated, and explains why such an agent need not be viewed

as contradicting himself or going back on something he originally asserted. He need not be seen as having done these things because the different judgements he makes change the extension and antiextension of the predicate in such a way that the proposition expressed by ⌜x_j is F⌝ when he assents to it differs from, and is compatible with, the proposition it expresses when he dissents from or withholds judgement about it. The semantic analysis also points to a useful lesson: although there is something about the meanings of many vague predicates that resists drawing stable boundary lines for applying them, the semantic rules governing such predicates are coherent as they stand, and there is no compelling practical or theoretical need for stable boundaries.[4]

In addition, the analysis provides the basis of rejecting the major premise MP of a generalized version of the Sorites paradox, while also explaining the deceptive plausibility it enjoys by virtue of its association with the more plausible and defensible premise MP* that arises directly from the rules for adjusting the extension of the vague predicate—in this case *blue*. (The background for the paradoxical argument includes the claim that there is a sequence S starting with something B that definitely looks and is blue, and ending with something NB that definitely is not and does not look blue. Moreover, for all members s_i and s_{i+1} of S, s_i is perceptually indistinguishable in color from x_{i+1} to competent observers in good light under normal conditions.)

(MP) For any two colored items x and y that are perceptually indistinguishable in color to competent observers in good light under normal conditions, x and y look to be and are of the same color. Hence for each s_i, s_{i+1} is blue, if s_i is blue.

(MP*) For any two colored items x and y that are perceptually indistinguishable in color to competent observers in good light under normal conditions, a person who characterizes *blue* as applying to x, (in such circumstances) is, all other things being equal, committed to a standard that counts *blue* as applying to y as well. Hence for each s_i, s_{i+1} is counted as blue, if s_i is explicitly characterized as blue.

By allowing us to distinguish the roughly correct MP* from the incorrect Sorites premise MP, the semantic model that I here presuppose is capable of

[4] See Soamer, *Understanding Truth*, ch. 7.

dispelling important and widespread confusions about standard versions of the Sorites paradox.[5]

Nevertheless, I don't regard the semantic model as providing a complete solution to the Sorites. As we will see, the model remains vulnerable to certain strengthened, revenge versions of the paradox, when we take higher-order vague predicates into account. It will be evident from the way these versions arise that there is a limit to how far one can go in defusing them by appealing to my semantic model. Even if the model is more or less correct, as I believe it to be, and even if it tells us important things about the Sorites, as I believe it to do, there remains a fundamental mystery brought out by the Sorites that the model does not resolve or illuminate. But that is getting ahead of ourselves. Higher-order vagueness can seem to be a perplexing problem for my semantic account from the very beginning. The central problem arises from treating vague predicates as partially defined. There is a natural line of reasoning arising from this characterization that makes it difficult to see how there could be any higher-order vagueness in the first place. What is the problem?

3. The Prima Facie Problem of Higher-Order Vagueness

Let P be a vague predicate that is undefined for objects that are in neither its default determinate-extension nor the default determinate-antiextension. Let ⌜is determinately P⌝ apply to an object *o* just in case *o* is in the determinate-extension of P.[6] This predicate applies neither to any object for which P is undefined nor to any object in its determinate-antiextension. Is it partially defined? There is reason to think that it can't be. In giving the analysis of P, we specified three and only three relevant categories of objects—those to which P determinately applies, those to which it determinately fails to apply, and those for which it is undefined. If these categories are jointly exhaustive, then ⌜is determinately P⌝ is totally defined, and so cannot be either vague or partial.

[5] See ibid., ch. 7. For objections and a reply, see Timothy Williamson, 'Soames on Vagueness,' and my 'Replies'.

[6] The *just in case* connective is used to form biconditionals that always have truth values when its arguments are undefined, or are otherwise such that we must reject assignments of truth values to them. (This is not a definition.) See *Understanding Truth*, ch. 6, for further discussion.

That sounds like a problem. The reason it is a problem is not, in my opinion, that there couldn't be vague predicates for which the relevant higher-order predicates were totally defined. It seems to me that we could, if we wanted, introduce an artificial predicate P which was both context-sensitive and partially defined, for which the higher-order predicate ⌜is determinately P⌝ was totally defined. On my view, P would then be vague, even though it would not give rise to higher-order vagueness. We could introduce P with this result, provided (i) that it was fully determinate what the rules governing our new predicate P were, and (ii) that these rules did not contain other vague or partially defined concepts, and so were not themselves vague or partial. Thus, higher-order vagueness is not a sine qua non for vagueness. However, the cases in which higher-order vagueness doesn't arise are special, and different from what we find with ordinary predicates like *bald*, *red*, *poor*, and *young*.

The problem is that higher-order predicates—⌜is determinately P⌝—corresponding to ordinary vague predicates like *bald* and *red* also appear to be vague. Not only is there no sharp and precise line dividing the objects to which *red* or *bald* apply from the objects to which they don't, there also seems to be no sharp and precise line dividing (i) the objects to which it is determined, by the rules of the language and the underlying non-linguistic facts, that these predicates apply from (ii) the objects for which this is not determined.[7] Thus, it would seem that the predicates *is determinately bald* and *is determinately red* are themselves partial. If they are also context-sensitive (which I will here assume), then they too should count as vague. This means that analyses of ordinary predicates like *bald* and *red* that treat them as partial and vague must explain how and why the higher-order predicates corresponding to them are also partial, and vague. How might this be done?

4. Proposed Explanation

4.1 The Idea

Think again about *bald*. There are some individuals to which it determinately applies, others to which it determinately does not apply, and still others for which it is indeterminate whether or not it applies, and so is undefined.

[7] Here and in what follows, I will presuppose the default settings of ordinary vague predicates when talking about objects to which they apply, don't apply, or are undefined—

Although these three categories are mutually exclusive, we should not assume that they are jointly exhaustive; after all, there may be individuals *o* such that we can find no possible basis for asserting that the predicate *x is determinately bald* applies to *o*, that it doesn't apply to *o*, or even that it must either apply or not apply to *o*. The reason for this is that there may be no possible basis to assert either (i) that the claim that *bald* applies to *o* is a necessary consequence of all non-linguistic facts about *o* plus the rules of the language governing *bald*, or (ii) that the claim that *bald* applies to *o* is not a necessary consequence of all non-linguistic facts about *o* plus the rules of the language governing the predicate *bald*, or (iii) that one of these claims about necessary consequence must be true.

How could this be? We may think of the rules governing *bald* as being of the sort indicated by the pair B_{pos} and B_{neg}, where *so and so* and *such and such* in the antecedents of the two conditionals are mutually exclusive, but not jointly exhaustive.[8]

Rules governing bald

(B_{pos}) For all *o*, if *o* is so and so, then 'bald' applies to *o* (and so *o* is bald).

(B_{neg}) For all *o*, if *o* is such and such then 'bald' doesn't apply to *o* (and so *o* isn't bald).

The rules governing the predicate *determinately bald* are the rules governing *bald* plus the rules D_{pos} and D_{neg} governing *determinately*.[9]

Rules governing *determinately*

(D_{pos}) For all *o*, if *o* is such that the claim expressed by ⌜'P' applies to *x*⌝ relative to an assignment of *o* to '*x*' is a necessary consequence of

unless special contextual standards are explicitly indicated. Thus, when P is such a predicate, ⌜is determinately P⌝ will standardly be taken to apply to *o* just in case *o* is in the default determinate-extension of P. Of course, when the determinate-extension of P is contextually adjusted, the extension of ⌜is determinately P⌝ is also adjusted. But such cases will concern us only when explicitly indicated.

[8] All instances of these rules are assumed to have definite truth values.

[9] In giving the rules for *determinately*, I have simplified matters to focus on the most dramatic and important case—uses of a vague predicate in contexts in which it carries its default determinate-extension and antiextension. In such a context for an object *o* to be determinately bald is for the claim that 'bald' applies to *o* to be a necessary consequence of the rules of the language governing 'bald' plus the underlying facts. In a context in which speakers have already exercised their discretion by adjusting the extension of 'bald', for *o* to be determinately bald is for the claim that 'bald' applies to *o* to be a necessary consequence of the rules already in force in the context plus the underlying facts.

the set of all non-linguistic facts about *o* plus the rules of the language governing P, then ⌜determinately P⌝ applies to *o* (and the claim expressed by ⌜*x* is determinately P⌝ relative to an assignment of *o* to '*x*' is true).

(D_{neg}) For all *o*, if *o* is such that the claim expressed by ⌜'P' applies to *x*⌝ relative to an assignment of *o* to '*x*' is not a necessary consequence of the set of all non-linguistic facts about *o* plus the rules of the language governing P, then ⌜determinately P⌝ does not apply to *o* (and the claim expressed by ⌜*x* is not determinately P⌝ relative to an assignment of *o* to '*x*' is true).

These rules are sensitive to three things: (i) the set of all non-linguistic facts about *o*, (ii) the rules of the language governing P, and (iii) the relation of necessary consequence. I will take (i) and (ii) to be sets of propositions, and (iii) to be a relation holding between sets of propositions and individual propositions, which, when applied to propositions that are precise and non-vague, is itself precise and well defined. In order to simplify the discussion, I will further assume that the propositions in (i) are all fully defined, precise, and true—no vagueness allowed here. However, no such assumption will be made in the case of (ii). If there is any vagueness about what the rules of the language are, or if there is any vagueness in something which definitely is a rule of the language, then this may affect the results achievable by applying D_{pos} and D_{neg}.

With this in mind, we return to our question *How can the conditions in the antecedents of D_{pos} and D_{neg} be seen as anything other than jointly exhaustive when P is the predicate 'bald'?* The answer is that whether or not we can establish or correctly accept the claim that these conditions are jointly exhaustive depends on whether or not we can establish or correctly accept the claim that for each potential rule R of the form $<B_{pos}, B_{neg}>$, it is determinate whether or not R is a rule of the language governing *bald*. The crucial point is that we cannot do this. Graph G1 of the baldness continuum illustrates this point.

```
(G1) | bald        |  ?         |  undefined     |  ?             |  not bald
     |             |            |  bald          |                |
     |xxxxxxxxxxxx|xxxxxxxxxxx|xxxxxxxxxxxxxxxx|xxxxxxxxxxxxxx|xxxxxxxxxxxx
       Region 1      Region 2     Region 3         Region 4         Region 5
```

Region 1 consists of individuals who would be judged to be clearly bald by virtually every competent speaker, provided the speaker were fully apprised

of the relevant facts about them, for example, by observing them in normal conditions. There is no serious question about these individuals; they are bald. Similarly, there is no serious question about those rule candidates B_{pos} that classify only members of region 1 as individuals to which *bald* applies; such candidates are included in the rules of the language governing *bald*.

Region 2 consists of individuals about whom there is moderate uncertainty or disagreement. Most competent speakers would judge these individuals to be bald, and few if any would confidently characterize them as not bald, but a significant number would be uncertain whether they qualify as bald, and would be somewhat reluctant to pronounce judgement on them. This region of individuals gives rise to undefinedness in the predicate ⌜is a rule of the language governing the predicate 'bald'⌝. Rule-candidates B_{pos} that classify all members of region 1, some members of region 2, and no members of any other region as individuals to which *bald* applies are rules for which the predicate *is a rule of the language governing 'bald'* is undefined.

Region 3 contains paradigmatically borderline cases of baldness. There is great uncertainty and variation among speakers, and across time, regarding whether they classify individuals in this region as bald or not bald; and often they may be reluctant or unwilling to classify these individuals as either. Rule candidates B_{pos} that classify *bald* as applying to some individuals in region 3, as well as rule candidates B_{neg} that classify *bald* as not applying to these individuals, are **not** rules of the language governing *bald*. They may be rules that speakers have the discretion to adopt in particular conversational circumstances; however, they are not rules that are constitutive of the language itself. Regions 4–5 are mirror images of regions 1–2, with *not bald* replacing *bald* and B_{neg} replacing B_{pos}.

On this way of looking at things, the rules of the language governing *bald* include many different pairs of positive and negative conditionals, even though one pair may subsume many others—i.e. cover every case that the others do, and more. So what are these rules? The rules governing *bald* include pairs of rules $< B_{pos}, B_{neg} >$ in which the antecedent of B_{pos} applies only to individuals in region 1 and the antecedent of B_{neg} applies only to individuals in region 5. The rules of the language governing *bald* do **not** include any pair in which either the antecedent of B_{pos} applies to individuals outside regions 1 and 2, or the antecedent of B_{neg} applies to individuals outside of regions 4 and 5 (though speakers may choose to adopt these rules in particular contexts). Any pair $< B_{pos}, B_{neg} >$ in which either (i) the antecedent of B_{pos} applies to individuals in region 2 (but none in regions

3–5), while the antecedent of B_{neg} applies only to individuals in regions 4 or 5, or (ii) the antecedent of B_{neg} applies individuals in region 4 (but none in regions 1–3), while the antecedent of B_{pos} applies only to individuals in regions 1 or 2, is such that we can draw no conclusion regarding whether or not it is a rule of the language governing *bald*. Speakers can decide to be guided by these rules in particular conversations, but, if they do, there will be individuals o_2 in region 2, or o_4 in region 4, such that we can establish no correct answer to the question *Are speakers' classifications of o_2 as bald, or o_4 as not bald, correct because they are consequences of the facts about these individuals plus the rules of the language governing 'bald', or are they correct because in making these classifications speakers have exercised their option of adopting extensions of the rules of the language?*

If this is right, then there is reason to resist the claim that *is determinately bald* is a totally defined predicate. The basis for the resistance is that for some rules there is simply no saying whether or not they are rules of the language governing *bald*. Let R be the class of such rules. For certain objects *o*—namely those in region 2 of G1—the question of whether the claim that *bald* applies to *o* is, or is not, a necessary consequence of the rules of the language governing the predicate can be answered only by assuming that certain members of R are rules of the language, or by assuming that they aren't. Since neither of these assumptions can be established, there is no possible justification for accepting them; thus, we should reject both the claim that these objects are determinately bald and the claim that these objects are not determinately bald, just as we rejected both the claim that they are bald and the claim that they are not bald. So, we reject the claim that *determinately bald* is totally defined.

The reason for this is that the rules governing the predicate ⌜determinately P⌝ make use of the predicate ⌜is a rule of the language governing 'P'⌝ which cannot be seen as total, when P is an ordinary vague predicate like *bald*. In saying this, I recognize that the picture I have sketched is incomplete, and that there are unfinished tasks that need to be pursued. Certainly, one would like more informative descriptions of different regions in the graph, including (non-circular?) explanations of crucial concepts—like that of being a competent speaker—employed in giving those descriptions. There is also the issue of locating vagueness in the descriptions of these regions, and exploring the sources and consequences of such vagueness. Despite these unresolved matters, I am not convinced that there is any irresolvable mystery here. As far as I can tell, the predicates *bald*, *determinately bald*, and *is a rule of the language*

governing 'bald' do fit the broad-brush picture I have sketched. Let us try to fill out that picture a little further.

4.2 Iterating 'Determinately'

The points we have made so far are visually represented by the graphs G1 for *bald*, G2 for *determinately bald*, and G3 for *determinately not bald*. (The question marks indicate that it is so far an open question how individuals in the region should be characterized.)

```
(G1) | bald          |  ?          | undefined      |  ?          | not bald
     |                             | bald           |             |
     |xxxxxxxxxxxx|xxxxxxxxxxxx|xxxxxxxxxxxxxxxx|xxxxxxxxxxx|xxxxxxxxxxxxxxxxx
       Region 1       Region 2     Region 3         Region 4       Region 5

(G2) | det.bald     |?/ undefined |                 not det.bald
     |              |det.bald     |
     |xxxxxxxxxxxx|xxxxxxxxxxxx|xxxxxxxxxxxxxxxx|xxxxxxxxxxx|xxxxxxxxxxxxxxxxx
       Region 1     Region 2                      Regions 3, 4, and 5

(G3) |                 not det.not bald          |?/undefined   | det.not bald
     |                                           |det.not bald  |
     |xxxxxxxxxxxx|xxxxxxxxxxxx|xxxxxxxxxxxxxxxx|xxxxxxxxxxx|xxxxxxxxxxxxxxxxx
                     Regions 1, 2 and 3            Region 4       Region 5
```

We have rejected the claim that *determinately bald* is totally defined. Should we accept the claim that it is partial (and presumably vague as well)? If so, do partiality and vagueness go even higher? Consider again the graphs and the question marks they contain. We know that the question marks in regions 2 and 4 of G1 do not indicate that *bald* is undefined for individuals in the regions. But, for all we have said up to now, the question marks in region 2 of G2 and region 4 of G3 might represent individuals for which the predicates *is determinately bald* and *is determinately not bald* are, respectively, undefined. Suppose this is so. We can then use (i) and (ii) to establish that the predicate *is determinately determinately bald* is totally defined.

(i) Just as the individuals of whom it can properly be said that they are determinately bald are the same as the individuals of whom it can properly be said that they are bald, so the individuals of whom it can properly be said that they are determinately determinately bald

are the same as the individuals of whom it can properly be said that they are determinately bald. Thus, the initial section of the graph for *determinately determinately bald* is the same as the initial section of the graph for *determinately bald*.

(ii) Just as the individuals of whom it can properly be said that they are not determinately bald include all and only those of whom it can properly be said either that they are not bald or that it is undefined whether or not they are bald, so the individuals of whom it can properly be said that they are not determinately determinately bald include all and only those of whom it can properly be said either that they are not determinately bald or that it is undefined whether or not they are determinately bald. So, if we accept the claim that the predicate *determinately bald* is undefined for every individual in region 2 of G2, and hence that every individual in that region is one of which it can properly be said that it is undefined whether or not that individual is determinately bald, then we get the graph G4 for *determinately determinately bald*, which is the graph of a totally defined predicate.

(G4)

```
| det.det.bald |                    not det.det.bald
|
|xxxxxxxxxxxx|xxxxxxxxxxx|xxxxxxxxxxxxxxx|xxxxxxxxxx|xxxxxxxxxxxxxx
  Region 1                  Regions 2, 3, 4, and 5
```

That is a surprising result. How is that when we start with *bald* and add *determinately* we get a predicate which cannot correctly be characterized as total, whereas when we start with that predicate and iterate *determinately*, we do get a totally defined predicate? The answer is that we have made a mistake. The crucial assumption, used in (ii), is that every individual *o* is either determinately bald, not determinately bald, or such that the predicate *determinately bald* is undefined for *o*. How do we know that? If at an earlier stage—in moving from *bald* to *determinately bald*—we had started with the assumption that every individual *o* is either bald, not bald, or such that *bald* is undefined for *o*, we would have reached the conclusion that *determinately bald* was totally defined—which we certainly did not. But if we didn't make that assumption in the previous case, in moving from G1 to G2, why should we make the corresponding assumption in this case, in moving from G2 to G4? The issue concerns the regions in the graphs labeled with question marks. All we know so far is that when, in G1, *o* is an individual in one of these regions, we reject

the claim that *o* is bald, we reject the claim that *o* is not bald, and we reject the claim that the predicate *bald* is undefined for *o*—all for the same reason, we see that it is impossible in principle to justify these claims. This being so, we need to clarify the status of the regions in the other graphs presently marked '*?/undefined*'.

The individuals in region 2 are undefined for *determinately bald* just in case those individuals are not determinately determinately bald, which will be so just in case *determinately determinately bald* is a totally defined predicate. How do we evaluate the claim that it is such a predicate? The first thing to notice is that the rules governing *determinately determinately bald* are the same as the rules governing *determinately bald*; they are the rules governing *bald* plus the rules D_{pos} and D_{neg} governing *determinately*, given earlier. In the present case, we apply the rules twice, once letting P be the predicate *bald*, and once letting P be the predicate *determinately bald*. The reasoning is given in (i) and (ii), and the results are summarized in (iii).

(i) Suppose we are given that the claim that *bald* applies to *o* is a necessary consequence of the rules governing *bald* plus the underlying non-linguistic facts about *o*. Then, using D_{pos} we derive that *determinately bald* applies to *o* (and hence that *o* is determinately bald). Since the rules governing *determinately bald*—namely, the rules $< D_{pos}, D_{neg} >$ governing *determinately* plus the rules governing *bald*—include the rules used in the foregoing derivation, this means that the claim that *determinately bald* applies to *o* is a necessary consequence of the rules governing *determinately bald* plus the underlying non-linguistic facts about *o*. But then, using D_{pos} again, we get the result that *determinately determinately bald* applies to *o*, and hence that *o* is determinately determinately bald.

(ii) Suppose we are given that the claim that *bald* applies to *o* is not a necessary consequence of the rules governing *bald* plus the underlying non-linguistic facts about *o*. Then, using D_{neg} we derive that *determinately bald* does not apply to *o* (and hence that *o* is not determinately bald). Since the rules governing *determinately bald* include those used in the foregoing derivation, this means that the claim that *determinately bald* does not apply to *o* is a necessary consequence of the rules governing *determinately bald* plus the underlying non-linguistic facts about *o*. But then, given the consistency of these rules (with the underlying non-linguistic facts), we conclude that the negation of

that claim—namely, the claim that *determinately bald* applies to *o* is not a consequence of the rules governing *determinately bald* plus the underlying non-linguistic facts about *o*. Finally using D_{neg} again, we get the result that *determinately determinately bald* does not apply to *o*, and hence that *o* is not determinately determinately bald.

(iii) When we are not given either (i) that the claim that *bald* applies to *o* is a necessary consequence of the rules governing *bald* plus the underlying non-linguistic facts about *o*, or (ii) that the claim that *bald* applies to *o* is not a necessary consequence of the rules governing *bald* plus the underlying non-linguistic facts about *o*, we cannot use the rules $< D_{pos}, D_{neg} >$ governing *determinately* to get any result. We conclude that the rules of the language together with the underlying non-linguistic facts give us the same results for *determinately bald* and *determinately determinately bald*. Since it is impossible to justify the claim that either predicate is totally defined, we reject this claim, and for any individual *o*, we accept the claim that *o* is (is not) determinately bald just in case we accept the claim that *o* is (is not) determinately determinately bald. The iteration of *determinately* does nothing.

4.3 What Not to Say

We have rejected the claim that *determinately bald* and *determinately determinately bald* are totally defined predicates. Are they partially defined? There is reason not to say this. I have said that partially defined predicates are those that are undefined for some objects, and that totally defined predicates are those that are not undefined for any object, where by *undefined* I have meant the following:

> *Undefinedness*. P is undefined for *o* just in case the rules of the language governing P together with the underlying non-linguistic facts about *o* do not determine either that P applies to *o* or that P does not apply to *o*—which in turn holds just in case neither the claim that P applies to *o* nor the claim that P does not apply to *o* is a necessary consequence of the rules governing P together with the non-linguistic facts about *o* (i.e. just in case neither ⌜x is determinately P⌝ nor ⌜x is determinately not P⌝ expresses a truth relative to an assignment of *o* to 'x').

Given this, one cannot correctly say that *determinately bald* is undefined for *o*. For if one does say this, one must then admit that *o* is not determinately

determinately bald. But that conflicts with what we have just found—namely, that just as we must reject, as unjustifiable, the claim that *o* is determinately bald, without accepting its negation, so we must reject the claim that *o* is determinately determinately bald, without accepting its negation. So is *determinately bald* undefined for *o* or not? Since neither claim can be justified, we have no option but to reject both. A similar result holds for (i) the claim that *determinately bald* and *determinately determinately bald* are partially defined predicates, and (ii) the claim they are not. Since our characterization of what it is to be a partially defined predicate requires the predicate to be undefined for some objects, we must reject the claim that these are partially defined, while continuing to reject the claim that they are totally defined. What can we positively assert about these predicates, and about the regions on the graphs for them that are labeled with question marks? As for the predicates, though they cannot correctly be characterized as partial in the original sense, they can be characterized as partial in a weaker and extended sense.

4.4 What we Can Say: Weak Partiality

A predicate P is *weakly partial* just in case there are some objects *o* such that, no matter how much information one is given about the rules of the language and the underlying non-linguistic facts, one cannot correctly accept either the claim that P applies to *o* or the claim that P does not apply to *o* (or the claim that either P applies to *o* or it doesn't). Ordinary, partially defined predicates like *bald* are weakly partial, as are the corresponding higher-order predicates formed by attaching one or more occurrences of *determinately* to them. The difference between partiality and weak partiality can be illuminated by considering the contrast between regions 2 and 3 on the graph G1 for *bald*. We consider a pair of claims—the claim that *bald* applies to o_2, and the claim that *bald* applies to o_3—where o_2 and o_3 are individuals in regions 2 and 3, respectively. Neither claim can be accepted because neither can be justified. But the reasons for the lack of justification are different in the two cases. In both cases, in order to justify the claim that the predicate applies to the object one has to establish the premise that there is a rule of the language governing *bald* which characterizes the predicate as applying to the object. In the case of o_3 we can refute this needed premise. In the case of o_2 we can neither refute it nor establish it. What the cases have in common is that since the needed premise can't be established, one in possession of all the facts cannot be

justified in accepting the claim that the predicate applies to the object, even though in neither case can one be justified in accepting the negation of that claim either. Genuinely partial predicates always include cases like o_3; predicates which are only weakly partial include cases like o_2, but none like o_3.

As for the regions on the graphs labeled with question marks, let us take region 2 of the graph G1 for *bald* as a representative example. Let *o* be an individual in this region. We can't correctly say that *bald* is undefined for *o* because there are pairs $< B_{pos}, B_{neg} >$ which are candidates for being rules of the language governing *bald* according to which *bald* does apply to *o*—where candidates are rules which we cannot show not to govern the predicate in the language. Since we can't show this, we cannot correctly say that *bald* is undefined for *o*. Of course, we also cannot correctly say that *bald* applies to *o*, because there is no pair of rules $< B_{pos}, B_{neg} >$ which characterize *bald* as applying to *o* that we can show to be rules of the language that do govern the predicate.

It is helpful in summarizing this situation to introduce the notion of a predicate P being undefined for an object *o* relative to a rule R.

> *Relative Undefinedness*. P is undefined for *o* relative to a rule R: $<P_{pos}, P_{neg}>$ iff neither the claim that P applies to *o* nor the claim that P doesn't apply to *o* is a necessary consequence of R plus the set of underlying non-linguistic facts about *o*. P is defined for *o* relative to R just in case P is not undefined for *o* relative to R.

Absolute undefinedness is defined in terms of relative undefinedness.

> *Absolute Undefinedness*. P is undefined for *o* iff (i) for all rules R which are such that we can, in principle, establish that R is a rule of the language governing P, P is undefined for *o* relative to R, and (ii) there is no rule R which is a candidate for being a rule of the language governing P, relative to which P is defined for *o*. (A candidate is a rule which we cannot, in principle, show not to be a rule of the language governing the predicate.)

In the presence of natural background assumptions—e.g. the assumption that if two rules are such that they should both be accepted as rules of the language, then they don't give conflicting characterizations of whether a predicate applies to any object—this definition gives the same results as the characterization of undefinedness given earlier. With this in mind, we can characterize each individual *o* in region 2 of the graphs as follows:

Individuals o in Region 2

(i) Every rule R which is such that we can establish that R is a rule of the language that governs the predicate *bald* is such that *bald* is undefined for *o* relative to R.

(ii) Nevertheless, there remain candidates for being a rule governing *bald* which characterize *bald* as applying to *o*.

(iii) For these reasons, we cannot establish, or correctly accept, any of the following claims: that *bald* applies to *o*, that *bald* is undefined for *o*, that *determinately* bald applies to *o*, that *determinately bald* does not apply to *o*, that *determinately bald* is undefined for *o* (ditto for *determinately determinately bald*).

(iv) It is the case, however, that *o* is not determinately not bald. (See G3.)

We have now distinguished predicates which are merely weakly partial from predicates which are (also) partial in the original sense. Ordinary vague predicates like *red* and *bald* are partial without qualification. Higher-order predicates built from them using the determinately operator are weakly partial (and correspondingly weakly vague). Is this the end of the story? Is there anything more to say about higher-order vagueness for partially defined predicates? I suspect there is.

4.5 Superundefinedness, Superdeterminateness, and Sharp Lines

Call the individuals in regions 2 and 4 of G1 *superundefined*, meaning by this that they are individuals of whom we cannot, in principle, establish that *bald* applies to them, that *bald* doesn't apply to them, or that *bald* is undefined for them, no matter how much information we are given. Since we cannot establish any of these claims, we cannot justifiably accept them. More precisely, we cannot accept them while maintaining that in so doing we are not exercising our discretion by contextually changing the conversational standards governing the predicate *bald*. Call objects that have this status objects for which the predicate 'bald' is *superundefined*. More generally, when an object *o* has this status for an arbitrary predicate P, we say that P is *superundefined* for *o*. With this definition in place, it seems plausible to suppose that for any predicate P and object *o*, either (i) P applies to *o*, (ii) P does not apply to *o*, (iii) P is undefined for *o*, or (iv) P is superundefined for *o*. These categories really do seem to be jointly exhaustive. Supposing that they are, we

may introduce an operator which attaches to a predicate P to form a totally defined predicate ⌜superdeterminately P.⌝

> *Superdeterminately Predicates*. The predicate ⌜superdeterminately P⌝ applies to an object *o* just in case it is not the case either that (i) P does not apply to *o*, or that (ii) P is undefined for *o*, or that (iii) P is superundefined for *o*.

Would it be a bad result if there really turned out to be such predicates? I don't see that it would.

The point of our discussion of higher-order vagueness for partially defined predicates has not been to avoid drawing sharp lines between all categories of objects to which one might think of applying a vague predicate. The point has been to accommodate what appears to be the genuine sense in which the higher-order predicate ⌜determinately P⌝ is vague (more precisely, weakly vague) when P is an ordinary vague predicate, like *bald*, or *red*. We have done that. As for sharp lines, the important questions are *If they exist, what do they separate?* and *How do they arise?* The lines I have been concerned with arise from the nature of contextual theories—theories that hold that there is a range of discretion within which speakers may acceptably adjust the contextual standards of what counts as red, bald, and the like. Since there are limits to the range of discretion that speakers have, there must be some individuals for which the rules of the language allow no discretion. For example, there must be some individuals for which any characterization conflicting with the characterization that the predicate applies to them is incorrect, no matter what the context.

Let us focus on this class of individuals, and the line separating them from the next class of individuals. This is the line between regions 1 and 2 in the graph G1 for *bald*. Individuals in region 1 are such that it is determinate that *bald* applies to them; hence, speakers have no option to characterize them in any other way. Since we know that individuals in region 2 are not determinately not bald, we know that one can correctly characterize the predicate as applying to them. However, if one does characterize *bald* as applying to these individuals, we can't say whether the rules of the language governing the predicate leave one any discretion to do otherwise. We may put this by saying that the individuals in region 2 are such that it is always correct to characterize *bald* as applying to them, but we cannot say whether the reason this is correct is because the rules of the language determine this characterization, or because in characterizing the predicate as applying to these individuals one is adopting a contextual standard that makes it correct. The line between

these things in region 2 (which may always correctly be said to be bald) and the things in region 1 (which may also always be correctly said to be bald) may very well be sharp. However, it is a line which, by its very nature, one would not expect speakers to notice. Hence, it is no embarrassment to the theory that they don't.

4.6 Implications for the Sorites

If I am right, then semantic models of vague predicates as both partial and context-sensitive do not allow one to avoid the conclusion that the meanings of these predicates impose classifications of individuals in their domains of potential application into sharply defined categories. Because of this, strengthened versions of the Sorites paradox can be constructed exploiting this fact.

A Strengthened Sorites Argument

A man with no hair is superdeterminately bald.
For all *x*, if *x* is superdeterminately bald, then a man with one more hair is too.
So everyone is superdeterminately bald.

Because of this one might wonder whether in using the semantic model I have defended we have made any progress in defusing the paradox. In my opinion we have, though we certainly have not fully resolved it. The puzzle that remains is how the linguistic behavior on which the semantics of our language supervenes results in such fine-grained classifications of the objects in the domains of our predicates. This is a problem for all theories of vague terms, and nothing I have said constitutes an answer to it.

However, if I am right about the semantics of these terms, then, it seems to me, these fine-grained classifications turn out to be less paradoxical and problematic than they were before. In particular, they do not pose the threat to our notion of linguistic competence that would be posed by a sharp, fine-grained bifurcation of the domain into objects to which a predicate definitely applies and those to which it definitely does not apply. The distinction between truth and falsity, or truth and untruth, is very important to speakers; and the norms of language use presuppose that we are able to closely track the truth. One lesson that has sometimes been drawn from traditional versions of the Sorites is that in order to avoid absurdity, we must embrace a semantic theory that distinguishes between those objects of which

a predicate is true and those of which it is not true in such a precise and fine-grained way that we can no longer view ordinary speakers who understand the predicate as competent to make the distinction, or as able to track the truth of statements made using it. That is paradoxical. How can a distinction based on meaning that is so important to language use be opaque to fully competent speakers who understand the meanings of their words? If the meaning of an ordinary predicate imposed a precise, fine-grained classification between objects to which it applied and those to which it did not, wouldn't fully competent speakers know this, and be able to locate the boundary with a high degree of accuracy? The virtue of the semantic account I have sketched is that it does not provoke these questions.[10] The distinction between truth and falsity is important enough to speakers that we expect an account of meaning (which is grasped by competent speakers) to classify statements into those categories in ways that fully competent speakers in possession of all relevant non-linguistic facts are able to approximate. By contrast, the sharp distinction between

(i) statements the truth of which are determined by the rules of one's language together with non-linguistic facts

and

(ii) statements for which there is no saying whether their truth is so determined or whether their truth results from the exercise of speaker discretion in adjusting the boundaries of context-sensitive predicates

is a highly theoretical one, of which speakers need have no clear and precise pretheoretical grasp. Since their shaky grasp of this distinction in no way impugns their competence, it is not paradoxical. Although all sharp, fine-grained distinctions imposed by the semantics of vague predicates are theoretically puzzling, they need not be paradoxical.

[10] More precisely, it doesn't provoke these questions for ordinary predicates like *red* and *bald*. Although related questions may arise for technical predicates, like *superdeterminately bald*, the sharp distinctions between things to which these predicates apply and those to which they don't are defined in terms of the theoretically less troubling distinctions corresponding to the ordinary vague predicates they arise from.