

## Vagueness Without Paradox

Diana Raffman

*The Philosophical Review*, Vol. 103, No. 1. (Jan., 1994), pp. 41-74.

Stable URL:

<http://links.jstor.org/sici?sici=0031-8108%28199401%29103%3A1%3C41%3AVWP%3E2.0.CO%3B2-O>

*The Philosophical Review* is currently published by Cornell University.



---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/sageschool.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

## Vagueness Without Paradox

Diana Raffman

Words like 'rich', 'heap', 'red', and even 'looks red', are vague. That is, they have blurred boundaries of application: there is no sharp division between cases in which they clearly apply and cases in which they don't.<sup>1</sup> There is, for example, no sharp division between objects that are clearly red and objects that aren't (clearly red), people who are clearly rich and people who aren't. The vagueness of these predicates goes hand in hand with what Crispin Wright has aptly called their "tolerance": they tolerate marginal changes in the parameters decisive of their application. That is to say, there is with respect to any vague predicate the notion of a "positive degree of change . . . insufficient to alter the justice with which [the predicate] is applied."<sup>2</sup> Thus an object just noticeably different in color from a red object is itself red; a person possessed of one dollar less than a rich person is himself rich. And so on.

The trouble with vague or tolerant predicates, of course, is that

---

<sup>1</sup>Here I follow Crispin Wright's characterization of vagueness in his well-known papers on the subject, "On the Coherence of Vague Predicates," *Synthese* 30 (1975): 325–64; "Language Mastery and the Sorites Paradox" (hereafter, LM), in *Truth and Meaning*, ed. Gareth Evans and John McDowell (Oxford: Oxford University Press, 1976), 223–47; and "Further Reflections on the Sorites Paradox" (hereafter, FR), *Philosophical Topics* 15 (1987): 227–90. As Wright explains, having fuzzy boundaries amounts to more than the admission of borderline cases: "there is no clear reason why possession of borderline-cases should entail possession of blurred boundaries. If, following Frege, we assimilate a predicate to a *function* taking objects as arguments and yielding a truth value as value, then a predicate with borderline-cases may be seen simply as a partial such function—which is consistent with the existence of a perfectly sharp distinction between cases for which it is defined and cases for which it is not" (LM, 226). So it isn't merely that no sharp line can be drawn between (clear) red and (clear) non-red, for example, but rather that no sharp line can be drawn between (clear) red and *any* other category—even a "borderline" category.

<sup>2</sup>Wright, LM, 229. Wright ultimately jettisons the notion of tolerance, claiming that it originates in a wrongheaded picture of our employment of vague terms as governed by semantic *rules*. Though I cannot treat Wright's view here, the present story will show that he is right about the inadequacy of a certain type of purely rule-based account, but wrong to abandon tolerance.

they generate sorites paradoxes. Consider, for example, a series of fifty colored patches, each just noticeably different in color from the next, ordered so as to progress from a clear case of red to a clear case of orange. Paradoxically, if 'red' is vague then it seems we can construct the following argument:

- (1) Patch #1 is red.
- (2) For any  $n$ , if patch # $n$  is red, then patch # $(n+1)$  is red.
- (3) Therefore patch #50 is red.<sup>3</sup>

But by hypothesis patch #50 is orange. Seemingly impeccable reasoning from seemingly impeccable premises yields a patent falsehood. Such anomalous results owe at least in part to the nontransitivity of marginal difference: a series of insignificant differences "add up" to a significant one.<sup>4</sup>

Thus are we caught in a peculiar bind: vagueness seems at once an essential yet paradoxical feature of the natural language. On the one hand, only by using vague predicates can we give voice to the relatively coarse-grained character of much of our commerce with the world around us. As Wright observes, "[w]e should have no use for a precisely demarcated analogue [of a vague predicate] in contexts in which the word is typically used. It would, for example, be ridiculous to force the question of obedience to the command 'pour out a heap of sand here', to turn on a count of the grains" (LM, 231). At the same time, their lack of sharp boundaries seems to render our use of vague terms incoherent, thus precluding any consistent logic or rules of application for them. Indeed, our successful employment of this chunk of the natural language comes to seem a feat of magic. Or so many philosophers have thought.

---

<sup>3</sup>The argument can also be formulated as a series of applications of *modus ponens*; hence worries about mathematical induction *per se* are misplaced. For discussion see Michael Dummett, "Wang's Paradox," *Synthese* 30 (1975): 304–5, and Hans Kamp, "The Paradox of the Heap" (hereafter, PH), in *Aspects of Philosophical Logic*, ed. U. Mönnich (Dordrecht, Holland: D. Reidel, 1981), 225–28.

<sup>4</sup>Indeed, it is familiar that the members of a sorites series may even be indistinguishable one from the next; indistinguishability too is nontransitive. For ease of discussion I shall follow Wright in using the term 'marginally different' to refer to indistinguishable items as well as slightly or just noticeably different ones; see Wright, LM, 233–34, 236–37.

In this paper I shall present a strategy for dismantling the paradox. The principal challenge of the sorites is to explain, in an *intuitively* compelling way, how a difference in kind can obtain between the endpoints (among others) of an effectively continuous series. That is the really hard part, and to my mind, no account thus far proposed has succeeded in it. In what follows, I hope to show how it might be accomplished.

I begin with the paradox for 'looks red'. Then I show how my solution works for 'is red' and other vague predicates generally. My story is at bottom a psychological one, resting on a hypothesis about the mental representations that underlie our usage of vague words. In its appeal to psychological considerations, the account advanced here differs markedly from previous contenders; for the most part, assaults on the paradox have been strictly "logico-semantic" in their methods and presuppositions.<sup>5</sup> If my solution is correct, then where vague predicates are concerned, logic and semantics are more intimately entwined with psychology than might otherwise have been supposed. I am not the first to suspect such a connection; despairing of any consistent set of rules for the employment of vague predicates, Wright recommends that we seek "a more behavioristic semantics," one according to which our competence with vague words is better conceived "on the model of a practical skill, comparable to the ability to hit a good cross-court backhand or ride a bicycle. . . . [T]he kind of sensitivity to

---

<sup>5</sup>In addition to the Wright, Dummett, and Kamp works cited above, see for example Kit Fine, "Vagueness, Truth, and Logic," *Synthese* 30 (1975): 301–24; Hilary Putnam, "Vagueness and Alternative Logic," *Erkenntnis* 19 (1983): 297–314; and J. Goguen, "The Logic of Inexact Concepts," *Synthese* 19 (1969): 325–73. Unfortunately I shall not be able to discuss the details of these various approaches here, but for helpful overview and evaluation see Wright, FR, sec. 5; Kamp, PH, sec. 3; Terence Horgan, "Psychologistic Semantics, Robust Vagueness, and the Philosophy of Language" (hereafter, PS), in *Meanings and Prototypes: Studies in Linguistic Categorization*, ed. S. L. Tsohatzidis (London: Routledge, 1990), 542–44; and Terence Horgan, "Robust Vagueness and the Forced-March Sorites Paradox" (hereafter, RV), in *Philosophical Perspectives* 8: *Logic and Language*, ed. J. Tomberlin (Northridge, Calif.: Ridgeview Publishing Company, 1994), 159–88, especially secs. 1 and 5–7. Much of the interest in sorites puzzles concerns their implications for the status of classical logic; again, I shall not be able to examine those issues here, but see David Sanford, "Competing Semantics of Vagueness: Many Values vs. Super-Truth," *Synthese* 33 (1976): 195–210, and Kamp, PH, for a survey of various positions.

normally-sighted) must have just *does* operate selectively among items which, in respect of apparent color, cannot be told apart” (FR, 238, 277–78). Though the present story will not force this latter counterintuitive conclusion, it is committed to the possibility that a semantics for vague predicates will find its roots in the churning machinery of the mind-brain.<sup>6</sup> Not surprisingly, our progress here will often call for substantial revision of the usual way of looking at things. But if we are willing to venture into some unaccustomed terrain, we will, I think, be rewarded.

# 1.

One of my principal contentions in this paper will be that an adequate treatment of vague predicates and their sorites puzzles must appeal to the character of our judgments about the items in the series. That makes ‘looks red’ a good starting place, with its undeniable sensitivity to judgmental factors; as Wright explains, “for *x* to look red just *is* for subjects to be willing to assent to that judgment [when relevant conditions of attentiveness, comprehension, visibility, and so forth] are met” (FR, 276). For present purposes we shall envision such a judgment as the activation of a *psychological category*, LOOKS RED.<sup>7</sup> The judgment that a given patch looks red will consist in the activation, by (the perception of) that-

---

<sup>6</sup>Terence Horgan also endorses a psychologistic approach to the semantics of vague predicates (see his PS); however, his treatment of the sorites itself is quite different from the one I advance here, focusing on the logic, rather than the psychology, of the application of vague words (see his RV). For a psychologistic treatment of the sorites, see C. L. Hardin, “Phenomenal Colors and Sorites”, *Noûs* 22 (1988): 213–34. Following J. A. Swets (*Signal Detection and Recognition by Human Observers* (New York: John Wiley and Sons, 1964)), Hardin argues on statistical grounds that any two different wavelengths of light are distinguishable, and hence that no series of the sort required for the sorites—that is, a series of pairwise indistinguishable colored patches—can be constructed. His thought is that given a sufficient number of pairwise discriminations, any physical difference between two patches would eventually manifest itself in subjects’ judgments in a statistically significant way. Hardin’s intriguing argument stumbles at a number of places, most importantly in its failure to touch the sorites as formulated for a series of just noticeably different patches.

<sup>7</sup>The precise nature of psychological categories is a matter of some controversy. (Psychologists sometimes call them “concepts,” reserving the term ‘category’ for the corresponding classes or types; however, because

patch, of the LOOKS RED category; and to look red will just *be* to activate that category. (As will emerge, this identification of looking red with activating a perceptual category is, strictly speaking, inessential to my account; it proves convenient for present purposes, however.) With this framework in place, I shall speak indifferently of patches' *looking red* and *being judged to look red*; and my talk of 'looks red' reports will be a merely convenient if elliptical way of referring to the judgments or activations that would tend to bring about such reports.<sup>8</sup> Also, I shall often write as if the patches are in, or are members of, one category or another, but these are *façons de parler*—the categories and their activations live in the mind.

Preliminaries aside, let us turn to the business at hand. The paradoxical argument for 'looks red' runs as follows:

- (1) Patch #1 looks red.
- (2) For any  $n$ , if patch # $n$  looks red, then patch # $(n+1)$  looks red.
- (3) Therefore patch #50 looks red.

But #50, in fact, looks orange. Where have we gone wrong?

Consider a normal competent subject faced with the task of judging, of each patch in the series, whether it looks red or looks orange. (For ease of discussion I shall suppose that the subject is faced with a two-alternative forced choice; but we could as easily allow him however many response categories he likes—'looks orangeish red', 'looks non-red', 'borderline', 'not sure', 'no fact of

---

the term 'concept' sustains a bewildering variety of senses in the philosophical literature, I refrain from using it here.) See *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, ed. Ulric Neisser (hereafter, *CCD*) (Cambridge: Cambridge University Press, 1987) for an overview of recent literature on categorial structure; then see C. L. Hardin, *Color for Philosophers* (hereafter, *CP*) (Indianapolis: Hackett Publishers, 1988) for an enlightening introduction to color categories in particular. For present purposes we can think of the LOOKS RED category as a long-term sensory-perceptual representation of "what red looks like." It is supposedly in virtue of possessing such a category that we are able to recognize red-looking things as such. I discuss the role of categorial structures in other perceptual domains, notably audition, in *Language, Music, and Mind* (Cambridge: MIT Press, Bradford Books, 1993).

<sup>8</sup>I have no special share in the term 'judgment'. For present purposes we could as easily speak of perceiving or believing that a patch looks red, or simply of categorizing it that way.

the matter', etc.<sup>9</sup> Even silence could count as a response. As far as the paradox is concerned, the same questions will always arise concerning the location of the last patch in any given category.) Viewing conditions are normal (whatever exactly that comes to), and the entire series is presented at all times. To keep things simple, let us suppose that the subject begins at #1, which he judges to look red, and then proceeds by consecutive patches toward #50. Now one thing we know is that at some point on each (complete) run of judgments along the series,<sup>10</sup> a category shift just does occur: the subject's slide down the slippery slope is broken by his eventual landing at a patch that activates a new category—here, LOOKS ORANGE. Just where the shift occurs on any given run will depend on a constellation of factors, including the direction in which the subject proceeds along the series, where in the series he begins his judgments, his perceptual state at the time, and so forth.<sup>11</sup> But occur it does, on every run. Why then does intuition tell us that marginal differences between patches cannot constitute categorial differences, when in some cases, manifestly, they do?

Imagine yourself the subject. "#1 looks red, #2 looks red, #3 looks red . . . ." Up through, say, #26, the patches look red. Perhaps your assessments become increasingly tentative as you go along. Then, after some moments of deliberation, you decide that #27 looks orange. How then can #26 and #27 be marginally different, *ipso facto* category-identical in respect of apparent-color, when they are, manifestly, in different categories? Ah, but there's the catch: they are *not* in different categories—or, rather, they are not in different categories when judged *pairwise* (that is, two at a time). Recall the specification of the paradoxical series: each patch

---

<sup>9</sup>The psychology is equally catholic in this regard; there is, for example, no requirement that the categories in question be "natural kind" categories like, say, the eleven so-called basic (species-wide) color categories. See, for example, Lawrence Barsalou, "The Instability of Graded Structure: Implications for the Nature of Concepts" (hereafter, IGS), in Neisser, *CCD*, 101–40.

<sup>10</sup>By 'run' I shall mean a unidirectional series of judgments upon consecutive patches. Just how much time can elapse between judgments on a single run is a matter I prefer to leave to intuition; presumably a three-second interval is acceptable and a two-hour lunch break is not.

<sup>11</sup>Among other things, some sort of anchoring heuristic is probably involved; see, for example, Amos Tversky and Daniel Kahneman, "Judgment Under Uncertainty: Heuristics and Biases" (hereafter, JUJ), *Science* 185 (1974): 1124–31, for relevant discussion.

is to look marginally different from *the next*; that is to say, the marginal difference relation holds of adjacent patches judged pairwise. In “intuiting” that every  $\#n$  and  $\#(n+1)$  are category-identical, we rightly envision ourselves as judging each two pairwise, simultaneously, one aside the other. So what (reexamined) intuition tells us, and what is true, is that for any  $n$ , if patch  $\#n$  looks red then patch  $\#(n+1)$  looks red, *insofar as  $\#n$  and  $\#(n+1)$  are judged pairwise*. Judged *singly* (that is, one at a time), we shall see, adjacent patches can be category-different.

What I am contending is that as it stands, the inductive premise of the paradoxical argument (call it “IP”) is false. However, it is easily mistaken for the true pairwise claim just floated; call this latter “IP\*.” We shall soon discover that IP\* does not yield the false conclusion required to generate the paradox. But first things first. Our immediate concern is to understand how our pairwise and singular judgments can diverge in the manner described above. That adjacent patches are pairwise category-identical is hardly surprising; but it remains a mystery how, even judged singly, they can be category-different.

As a start, it is helpful to suspend the familiar Cartesian vision of a unified and transparent mind. Imagine instead that your ‘looks red’/‘looks orange’ judgments are the coordinated output of a pair of somewhat independent homunculi—one who *categorizes* and one who *discriminates* (that is, makes same/different judgments). As long as you judge the patches singly, your categorizer has free rein: he may, for example, categorize #26 as looking red, #27 orange, #28 red, and #29 and everything thereafter orange. When adjacent patches are judged *pairwise*, however, your discriminator takes over. Finding a pair to be marginally different, he will constrain his colleague to categorize them identically: “Categorize them as you like,” he’ll say, “but categorize them together.” Thus when you judge #26 and #27 pairwise, you will place them both in the same category, no matter how they look when you judge them singly. Where #26 and #27 together look orange after the shift, #26 will evidently have shifted its category membership. If on the other hand you make all of your judgments along the series pairwise ( $\#1/\#2$ ,  $\#2/\#3$ ,  $\#3/\#4$ , etc.), the shift from ‘looks red’ to ‘looks orange’ will occur between *pairs*—say, between the  $\#25/\#26$  pair and the  $\#26/\#27$  pair. In that case



the shared patch (again, #26) will shift its category membership from one pairing to the next.<sup>12</sup>

Of course I shall need to say a good deal more to make *intuitively* plausible the idea that a category difference could obtain between adjacent patches judged singly. However, let me pause here to mention two factors that lend credence to my homuncular duo quite apart from their role in dispelling the paradox.

First is what might be called the “structural” difference between discriminatory and categorial judgments. Whereas the former pertain to comparisons of two presented patches, the latter presumably pertain to “comparisons” of a presented patch with some sort of standard or prototype in memory. Invoking a spatial metaphor, we can envision the two kinds of judgments as orthogonal to one another, with the discriminatory judgments represented horizontally as between adjacent patches, and the categorial judgments represented vertically as between each patch and the stored prototype. With this picture in mind, it is easier to see how the categorizer, assessing single patches in the series, could be indifferent to the proximity between neighbors: what concerns him is not the “horizontal” distance between patches, but rather the “vertical” distance between each patch and the prototype in memory. As we have noted, the baffling feature of the series is the existence of a categorial divide, despite the absence of any nonmarginal difference, between adjacent patches. Now we begin to see how such a divide could obtain. If we are mystified by our ability to draw categorial distinctions between patches we “can’t tell apart,” that is partly because we are seeing things, as it were, from the discriminator’s point of view. The discriminator has no memory to speak of, and certainly no memory of the sort required for categorial distinctions; hence he fails to notice the progressive change in the appearance of the patches as he moves along the series. The categorizer, on the other hand, has a rather good memory—or at least a good memory for the clearly red-looking patches like #1

---

<sup>12</sup>Some readers have feared that my homuncular metaphor simply relocates the original problem: how do the little men make *their* judgments? However, I employ the notion of a homunculus in the purely mechanical sense current in cognitive science; these little men in the head are, as Daniel Dennett puts it, “just picturesquely described parts of the switching machinery” (“A Cure for the Common Code?” in *Brainstorms* (Cambridge: MIT Press, Bradford Books, 1978), 102).

and #2. We might capture the idea by envisioning the category as a kind of mental elastic band anchored at one end to the stored prototype; the categorizer's job would then be to try to stretch the band far enough to embrace a given patch.<sup>13</sup> As a matter of brute mechanical fact, he is able to stretch the band only so far at any given time, and his performance will vary from run to run with his degree of fatigue, the current condition of the band, and so forth. Encountering a patch he cannot reach with his LOOKS RED elastic band, he will shift to a different band—here, the LOOKS ORANGE one. Conceived in this way, the categorizer does not strictly speaking *discriminate* between adjacent patches. Rather, he is concerned exclusively with the “vertical” distances between presented stimuli and information in memory.<sup>14</sup>

A second factor counting in favor of my homuncular pair is the apparent existence of special-purpose mechanisms for hue discrimination. As C. L. Hardin explains,

[s]upposing that there is a color-difference signal that receives separate treatment by the visual system is not quite as gratuitous as it may seem to be. . . . [A]s we know, the cortex contains numerous cells with “preferences” for certain features of the visual scene. And, in particular, there are cells that respond primarily to chromatic *difference* without caring much about the particular chromaticities involved. (CP, 181)

Granted, nothing I have suggested would demand specialized hard-

---

<sup>13</sup>Such a picture has a lot in common with so-called *prototype* theories of the psychology of concepts. See, for example, Eleanor Rosch, “Natural Categories,” *Cognitive Psychology* 4 (1973): 328–50, and “Cognitive Reference Points,” *Cognitive Psychology* 7 (1975): 532–47; then see Horgan, PS, 544–52, for general discussion of the potential relevance of prototype theory for a semantics of vagueness.

<sup>14</sup>This is not to suggest that the categorizer discriminates between vertical distances, either. In theory, a shift could occur even between physically identical patches. For example, it might be that the categorizer so exhausts himself reaching for one patch that he shifts to a new band for the next, identical patch, hoping to ease his task; or perhaps he simply gets bored with one band and so shifts to another. Less picturesquely: it might be that some shifts are triggered by a patch's sheer spatial proximity to #50, rather than any increased similarity in apparent-color. Be that as it may, we shall see that even a category shift between identical patches would pose no threat to the apparent continuity of the series. (Of course, in a sorites series, not all of the patches can be physically identical.)

ware. We are likely to find such mechanisms throughout perception, but there is no obvious reason to expect the same for non-observational words like ‘rich’ or ‘person’. A sufficient *functional* independence of the two kinds of judgments could underwrite the scenario I have in mind. My present point is simply that the existence of special-purpose machinery in the perceptual cases tends to support the hypothesis of an (at least) functionally similar arrangement across the board (see note 22 below).

Be the architectural details as they may, what is critical to my account is that our pairwise and singular categorical judgments can diverge in the manner described above. That is to say: judged pairwise, adjacent patches are category-identical; judged singly, they may be category-different. But of course I shall need to say more if the latter claim is to win us over. *Intuitively* speaking, how can adjacent patches, even judged singly, be category-different? How, that is, can IP be false?

Consult your intuitions for a moment. Suppose that upon judging #27 to look orange (singly), you are queried, “But didn’t you just say #26 looked red? How then can #27 look orange?” Doesn’t intuition suggest that you will retract one of the two judgments—most likely the earlier judgment that #26 looked red? Or, better, you will say that things seem different to you now, that your perspective has changed, that though #26 looked red before, it now looks orange, even judged singly. And similarly, moreover, for some number of its predecessors (#25, #24, #23, etc.). How far back does the shift in perspective extend? To the next such shift, at which point the entire process repeats in reverse. Suppose you judge ‘looks orange’ all the way back down to #19, but then #18 again looks red. In the simplest-case scenario, #19 will then look red both singly and pairwise against #18, #20 red singly and pairwise against #19, and so forth up the series until a new ‘looks red’/‘looks orange’ shift occurs.

As a first approximation (which will want refinement later on), I propose that a category shift consists in a shift of perspective in which the new category instantaneously “spreads backward” along a string of the preceding patches. In other words, the new category (LOOKS ORANGE, in the first shift at #27) is what I shall call “fluid”: it expands backward, instantaneously, to embrace some of the patches that formerly lay in the range of its competitor (LOOKS

RED).<sup>15</sup> Thus the category shift is not a local phenomenon; rather, it is spread out across some portion of the series, so that a string of patches shift their category *together*. In this way, #26 and #27 can be category-different without jeopardizing the visual continuity of the series, for they are never category-different *at the same time*. The category difference between them is wholly successive: when you judge that #27 looks orange, it's not as if #26 still looks red. On the contrary, by the time you arrive at #27, you have already entered LOOKS ORANGE territory (that is, you have already entered a region of patches that sometimes look orange), and so patches on both sides of #27 look orange at the moment of shift. (If you like, once the categorizer has switched to his LOOKS ORANGE elastic band, he will continue to use it as long as he can, even on patches that formerly looked red. Switching bands is a complicated business.) A category shift occurs between #26 and #27, to be sure, but it embraces *both* patches (and probably others).<sup>16</sup> Small wonder we thought IP true.

I should say straightaway that I know of no empirical research testing specifically whether subjects do thus alter their judgments when asked to proceed backward down a sorites series after a shift (though it is perhaps worth noting that the psychologists I have

---

<sup>15</sup>Such a view of categorial judgment accords well with recent studies indicating a high degree of "flexibility" in our psychological categories; see, for example, Barsalou, IGS.

<sup>16</sup>Of course, there is no ruling out a priori the possibility of (highly exceptional) instances in which #26 continues to look red, judged singly, even after a shift to 'looks orange' at #27. In such a case my view will need to say that the backward spread of the LOOKS ORANGE category is a limiting case of spread having zero places in length. (Even then, however, #26 and #27 will not be *simultaneously* category-different: if you judge #26 to look red immediately after judging #27 to look orange, then a second category shift, embracing #27 as well as #26 *inter alia*, will have been triggered by #26.) Granted, the adversion to limiting cases avoids being *ad hoc* only if the limiting cases are far and away the exceptional cases *and* the theory does a good job of explaining the ordinary ones. In the present instance we shall have reason to be optimistic on both counts.

Let me emphasize that after the shift at #27 you would not, or at least should not, say you were *mistaken* in your 'pre-shift judgment of #26 as looking red. This much seems to me intuitively obvious in any case; but more can be said. For consider that if your pre-shift judgment was mistaken—if #26 actually looked orange before the shift—then (trivially) you were mistaken in your pre-shift judgments of preceding patches back to some #n (either #26 or a predecessor). But then #n and #(n-1) flank a sharp boundary and the original puzzle simply reappears.

consulted find the idea quite plausible). Nevertheless, we can extrapolate directly from related findings. Among other things, it is widely acknowledged that subjects often employ so-called *anchoring heuristics*: they categorize stimuli relative to certain mental *reference points* or *anchors* and, as a result, are biased thereby. For instance, if you begin a run of judgments along the series at patch #1, then *ceteris paribus* you will employ a representation of a prototypical red-looking patch as your anchor, whereas if you begin at #50 you will similarly employ a representation of a prototypical orange-looking one. And an anchor, as its name suggests, sets up a kind of judgmental inertia. Hence you will tend to categorize a greater number of patches as looking red when you begin at #1 than when you begin at #50, and a greater number as looking orange when you begin at #50 than when you begin at #1. As Tversky and Kahneman explain, “different starting points yield different [judgments], which are biased toward the initial values. We call this phenomenon anchoring” (JUU, 1128).<sup>17</sup> What I am suggesting at present is that at the moment of shift, you (or, more properly, your categorizer) shifts to a new anchor—in my example, a representation of a prototypical orange-looking patch like #50. Your post-shift categorizations are then biased toward the new LOOKS ORANGE anchor, and upon being queried about preceding patches you will tend to judge at least some of them as looking orange.

Remember that the shift occurs between borderline cases, not clear ones. These former are patches that can look either red or

---

<sup>17</sup>Analogous “order effects” are observed in psychophysical experiments. S. S. Stevens reports that

the order of presentation affects the apparent size of the successive intervals between the stimuli. For example, if three loudnesses are presented in an ascending order 1, 2, 3, the interval between 2 and 3 seems relatively larger. When the order is reversed, the interval between 2 and 1 seems relatively larger. Suppose we now ask the observer to adjust stimulus 2 so as to bisect the interval 1 to 3, or, in other words, to make the distance 1 to 2 equal the distance 2 to 3. In order to equate the apparent intervals in the ascending order, the observer must set stimulus 2 to a higher value than in the descending order (*Psychophysics* (New Brunswick: Transaction Books, 1986), 129).

Moreover, when subjects use a method of magnitude estimation to judge values along sensory-perceptual stimulus dimensions, “a stimulus tends to be judged lower when preceded by a lower stimulus than when preceded by a higher stimulus, and vice versa” (ibid., 275). This latter finding further corroborates my hypothesis that (for example) #26 will tend to look orange after a shift to LOOKS ORANGE at #27.

orange: #26, for example, looks red before the shift, orange afterward. More to the point, borderline cases are those with respect to which *looking red* and *looking orange* are very much alike (maybe even the same—see the following paragraph). To deny this is to forget that there is more to the categories at issue than their clear cases. When it seems to us that no category shift can occur in the series, we are, I suspect, thinking only of *clear* cases of looking red and looking orange; we are not thinking of borderline cases. Thus we are bound to overlook the fact that in the latter neighborhood, patches that look red look an awful lot like patches that look orange.

Indeed, carrying out a few homegrown trials of my own, I found myself inclined to say that in some quite straightforward sense, the appearance of the patches remained constant throughout. #26, for example, looked the same (almost the same?) when I judged it to look red as when I judged it to look orange. (Of course, it also looked different: it looked red beforehand, orange afterward. That is what makes the situation so peculiar.) To put the claim more starkly: my inclination was to say that one and the same *look* looked red before the shift, orange afterward; and similarly for some of the preceding patches. Ruth Barcan Marcus has suggested to me that such an effect may amount to a kind of Gestalt switch: there is a similar respect in which (for example) the duck-rabbit “looks the same” while yet “looking different” as its species fluctuates. I want to urge that both cases of shift mobilize a process of *seeing as*: just as we can see the duck-rabbit either *as a duck* or *as a rabbit*, we can see #26 and the others either *as looking red* or *as looking orange*. It is because we undergo such shifts that we perceive no discontinuities in the series.

If it continues to seem mysterious that #27 could trigger a category shift where #26 does not, keep in mind what such a shift consists in: it consists in the instantaneous backward spread of a psychological category. The point is, we have no *intuition* to the effect that if #26 does not trigger such a shift then neither will #27. What will or will not trigger a shift is not something to which we, as judging subjects, have access; so far as the subject is concerned, the shift simply occurs. (Perhaps it is the work of a categorizing homunculus; again, see note 14.) ‘Looks red’ is a so-called purely observational predicate insofar as we can tell, just by looking at a

given patch, whether it looks red then. What doesn't follow is that we can tell, just by looking, how the patch will look at other times.

## 2.

Now for a refinement of the story presented thus far. It may have occurred to the reader that my account simply relocates the original problem. For instance, at the moment of shift at #27, won't there be a sharp boundary between the orange-looking patches in the backward spread and the red-looking patches that precede them? Specifically, if the spread of LOOKS ORANGE on this run extends back through #19, won't there be a sharp boundary between #19 (which looks orange) and #18 (which looks red)? Evidently we cannot say that at the moment of shift to 'looks orange' at #27, #26 and the other patches in the spread look orange as well. On the other hand, we cannot say they look *red*, for then #26 and #27 themselves would flank a sharp boundary. Similarly, when the subject is judging #26 to look red, how does patch #27 look? If we say it looks orange, then #26 and #27 flank a boundary; if we say it looks red, we only push the boundary farther along the series toward #50. So how *do* the patches in question look? (As throughout, the introduction of further response categories is ineffectual.)

The right thing to say here, it seems to me, is that the patches have no apparent-color at all until the subject "gets there," as it were. Strictly speaking, a patch looks red or looks orange (/reddish orange/non-red/borderline, etc.) when, and only when, it is being so judged—in other words, when and only when it is activating the relevant psychological category. I shall put this by saying that the apparent-colors of the patches, indeed of objects generally, are *episodic*.<sup>18</sup> In a formal mode: 'looks red' and 'looks orange' apply *episodically*. At the moment of shift, #27 is the only patch in the series that has an apparent-color. Thus any simultaneous adjacent category differences (boundaries) are avoided: no categorical difference obtains between #27 and #26, nor between #27 and #28, because failing to have a category membership at any given time is not itself an apparent-color category. In the present context, to

---

<sup>18</sup>No doubt familiar problems will arise concerning the individuation of judgmental episodes, but we needn't solve those here. However the episodes are individuated, my story will remain essentially the same.

say that a patch fails to have a category membership (that is, fails to have an apparent-color) is simply to say that it is not currently being judged. And no one ever thought *that* sort of difference couldn't obtain between adjacent patches.

We must consider the following objection, however. Surely those patches that look red every time they are judged do look red nonepisodically: they look red at all times, whether or not they are being judged. And if that is true, then there will be two adjacent patches such that, simultaneously, the one looks red while the other looks orange. Suppose the patches up through #14 are invariably judged to look red, but #15 is so judged 95 percent of the time. In that case, at those rare times when #15 is being judged to look orange (for example, when it is the last patch in a backward spread of LOOKS ORANGE), aren't #14 and #15 simultaneously category-different?

In reply let me first point out that, as a matter of fact, no patch invariably looks red; run enough trials and any subject will eventually say 'looks orange' even of #1. (There are no perfect performers, after all.) What this objector is asking about, rather, are patches that look red every time they are *competently* judged. (Of course, the notion of competence is itself vague, but set that aside for the moment.) Now it may be natural to think of such "invariant" patches as looking red nonepisodically, and for practical purposes there is no harm in doing so. But strictly speaking what is true is only that such patches look red every time they are (competently) judged. *That* is their relevant nonepisodic property. Ordinarily when we ask whether an object *O* looks red, we are asking, elliptically, whether *O* would look red were it to be looked at; in other words, we are asking about the visual appearance *O* would present, not one it is now presenting. With respect to their apparent-colors, these invariant patches differ only in degree, not in kind, from the highly variable patches in the middle of the series. They are merely limiting cases—patches which, when competently judged, invariably look red.

If what I have just said is correct, then at no time are adjacent patches in the series simultaneously category-different. Admittedly, #14 has, while #15 simultaneously lacks, the nonepisodic property of invariably looking red when competently judged. (I shall explain in a moment how *that* kind of division can obtain in the series.) But that is not the sort of difference at issue in the objection: the



difference at issue there concerns not how #15 looks across runs, but merely how #15 looks at a particular time. And my point is, it poses no threat to our view that, at a given time, #15 looks orange simultaneously with #14's having the nonepisodic property of invariably looking red when judged. That is because (1) at the moment when #15 is judged to look orange, #14 has no apparent-color, and (2) when the subject subsequently reconsiders #15 after judging #14, he will virtually always judge it to look red as well.

Lastly: how can adjacent patches differ in the frequencies with which they look red? The answer is already in hand. As we have seen, there is no telling by looking, on a given run (a fortiori across all runs), which patch will trigger a category shift. The exact location of the shift is determined by brutally mechanical ("sub-personal") factors. (Perhaps your categorizer simply gets tired or bored at a certain point and decides to shift bands; again, see note 14, and also note 26.) Hence there is nothing strange about a sharp division between the last patch that looks red, say, 95 percent of the time and the first patch that doesn't, even where these are only marginally different. To put the point another way, the difference between the last patch that looks red 95 percent of the time and the first patch that doesn't need not be a *visible* difference. On the contrary, the visual appearance of a patch is at best good evidence for its category memberships across runs. Thus a boundary of the statistical sort presently contemplated poses no threat to the visual continuity of the series.

Perhaps what I have been saying seems like so much philosophical whitewash: if only one patch at a time has an apparent-color,<sup>19</sup> then *trivially* there is never a simultaneous category difference between neighbors. Solving the paradox can't be that easy! And of course, it is not. To solve the paradox we have also had, first, to explain how a category shift could occur *at all* in such a series, and, second, to do away with successive but *fixed* adjacent category dif-

---

<sup>19</sup>More properly: at any given time, only the patch(es) currently being judged has(have) an apparent-color; a pairwise judgment, for example, categorizes two patches at the same time (that is, in a single judgment). Indeed, there will be many ways of categorizing large numbers of patches, even the whole series, at the same time: one might, for instance, judge all of the patches to look red-or-orange-or-something-in-between. The only restriction is that no two adjacent patches activate "incompatible" categories at the same time.

ferences. It is not enough that (for example) #14 and #15 are never simultaneously category-different; it must further be the case that, typically, once #14 has been judged to look red, #15 will then look red (when judged) as well. That is where the category spread comes in. If #15 continued to look orange after #14 had been judged to look red, we would be saddled with a sharp and fixed division and the intuitive mystery would stand.

In the final analysis, how ought we to characterize the backward spread? In particular, if the patches in the series have apparent-colors only when being judged, what does it mean to say that a category-shift-*cum*-backward-spread has occurred? It means that the subject has entered a certain psychological state. Specifically, the occurrence of a shift-*cum*-backward-spread at some  $\#(n+1)$  consists in the subject's entering a psychological state such that he places  $\#(n+1)$  in a "new" category and *ceteris paribus* would, if queried, similarly categorize some number (typically greater than zero) of its predecessors. Our "first approximation" that a string of patches shift to looking orange together with #27 is thus recast as the claim that *ceteris paribus* the subject would judge that string of patches to look orange were he to proceed directly backward down the series. The patches actually look orange, however, only when *being* so judged. There are shifts—events that occur—but these are best viewed as Gestalt-like changes of "perspective" or "anchor," not as boundary crossings—at least not if by 'boundary' you mean something that installs a simultaneous and/or fixed category difference between adjacent patches.

### 3.

With the foregoing picture in mind, we can begin to appreciate how attenuated is the sense in which #26 is the last red-looking patch in the series. More than anything else, it is the notion of backward spread that reforms our conception of category shift in such a way as to dispel the intuitive mystery of the sorites (for 'looks red', anyway). Now we see how shifts can occur in an effectively continuous series. Now, that is, we see how IP could be false.

I have said that the conclusion of the paradoxical argument would not follow from the true pairwise claim IP\*. When IP is replaced by IP\*, and the first premise and conclusion are spelled out fully, the equivocation is clear:

- (1) Patch #1, judged singly, looks red.
- (2) For any  $n$ , if patch # $n$  looks red then patch # $(n+1)$  looks red, insofar as # $n$  and # $(n+1)$  are judged pairwise.
- (3) Therefore, patch #50, judged singly, looks red.

As I have argued, nothing about the singular category membership of a patch in the series follows from the category memberships it sustains when judged pairwise against its neighbors. Hence the argument does not go through.

Perhaps it will be thought that the paradox could be revived by appeal to a false pairwise conclusion: can't we construct a seemingly valid argument with “#49 looks red when judged pairwise against #50” as its (false) conclusion? No. Recall the scenario in which the subject makes all of his categorial judgments along the series pairwise. There the category shift evidently occurs between *pairs* (the #25/#26 pair and the #26/#27 pair, in my example). In such a case the pairs are single items, as far as the categorizer is concerned: just as category shifts can occur between adjacent patches judged singly, they can occur between adjacent *pairs* judged singly, even where those pairs share a member. Hence nothing about the category membership of one pair follows from that of any other pair. Furthermore, as we have seen, nothing about the pairwise category membership of adjacent patches follows from their respective singular memberships (recall, for example, the singleton #26 and the pair #26/#27). Indeed, as far as I can see, any effort to resurrect the paradox for ‘looks red’ will be similarly blocked.

As I have presented it, the foregoing account presupposes that looking red *consists in* activating a certain perceptual category. But as I noted at the outset, such a presupposition is inessential to my solution of the puzzle. The sorites for ‘looks red’ or any other vague word is solved independently of any particular meaning analysis of the predicate (or metaphysical account of the property it names). On the contrary, all that is required to solve the puzzle is a claim about the correct application or *extension*—as opposed to the meaning analysis or *intension*—of the predicate at issue. The solution for ‘is red’ makes this clear.

#### 4.

Resolving the paradox for ‘is red’ is complicated by our uncertainty as to what the predicate means. (From now on I shall typically call

it, simply, “red”). Should we say, for example, that ‘red’ means ‘is disposed to look red under normal viewing conditions’? Or ‘has that microstructural property that disposes objects to look red under normal conditions’? Or perhaps ‘reflects all and only light of precisely 650 nanometers’? Or ‘reflects all and only light of *roughly* 650 nm’? These do not exhaust the possibilities. I cannot here consider all the ways we might analyze ‘red’; and in any case, my solution to the puzzle, which concerns only the correct application or extension of the predicate, is neutral with regard to its meaning analysis. (More on this neutrality as we proceed.) Accordingly, I propose to examine two contrasting analyses of ‘red’ and show that my solution disposes handily of any associated paradoxes. Then I shall suggest some reasons for believing that any version of the paradox for ‘red’ could be cured in similar fashion.

Consider first a crude “physicalist” analysis:

- (R1) ‘X is red’ =df ‘X is disposed to reflect all and only light of precisely 650 nm’.

On this reading, ‘red’ is intolerant to marginal change: there is no positive degree of change in wavelength insufficient to alter the justice with which the predicate is applied. Thus such a ‘red’ has sharp and fixed boundaries and no paradox arises. (Of course, one might well insist that ‘reflect’ and ‘light’ are vague, and hence that ‘red’ as interpreted by R1 is vague after all. My present and obvious point is that *if* ‘red’ is non-vague, no paradox arises; you may appeal to whatever non-vague analysis you like.)

Naturally, those who think that ‘red’ is vague will be unpersuaded by R1. Consider instead R2, taken by many to be truistic:

- (R2) ‘X is red’ =df ‘X is disposed to look red under normal conditions’.<sup>20</sup>

---

<sup>20</sup>Just for example, Wright maintains that

to be red is to look red in circumstances of observation which leave nothing to be desired—however exactly that condition should be explicated. . . . Looking red suffices for being red when other things are equal (FR, 273).

A remark is perhaps in order about the apparent circularity of defining ‘red’, as does R2, in terms of ‘looks red’. I think the circle can be broken

'Normal conditions' here covers a variety of requirements on the character of the light, the nature of the intervening medium, the location and duration of the proximal stimulus to the retina, the composition of the visual field, the state of the subject's visual system, the character of the subject's judgmental faculties, and so forth. And a sorites series could be constructed along any of these many dimensions.

For simplicity's sake, let us consider a series of patches that vary in respect of the normalcy of the light in which they look red. Specifically, suppose that normal light for viewing these patches is midsummer daylight between 6 A.M. and 8 P.M. Suppose further that the series contains eighty-four patches, each paired with an indication of the time span of light in which it looks red, progressing from a first patch that looks red between 6 A.M. and 8 P.M. to a last patch that looks red between 7:50 P.M. and 8 P.M., so ordered that each time span of light is ten minutes shorter than the preceding. Suppose finally that the 7:50 P.M.–8 P.M. time span is clearly non-normal (hence that patch #84 is clearly not disposed to look red in normal light) and that the 1 P.M.–8 P.M. time span is a clear borderline case (hence that patch #43 is a clear borderline case of being so disposed).

---

by assigning different senses to 'red' in the two predicates: roughly, looking red will consist in presenting a certain visual appearance (here, in activating a certain visual category), while *being red* will consist in being disposed to present that appearance under normal conditions. The difference might be signaled by the introduction of a special predicate, 'looksred', to replace 'looks red'. Then the latter term would mean 'looks as if it is red (that is, as if it is disposed to lookred under normal conditions)'; and a patch could perfectly well lookred without looking red. Of course, the question then arises as to whether the circle recurs when we try to say *which* perceptual category is activated when something looksred. I think there are ways to handle the problem, or at least to show that color predicates are no worse off in respect of their definition than many other elements of the natural language. (In this connection, see Michael Watkins, "Dispositionalism, Ostension, and Austerity" (hereafter, DOA), *Philosophical Studies* (forthcoming); also, for relevant discussion, Christopher Peacocke, "Colour Concepts and Colour Experience," *Synthese* 58 (1984): 365–81; Colin McGinn, *The Subjective View: Secondary Qualities and Indexical Thoughts* (hereafter, SV) (Oxford: Oxford University Press, 1983), chap. 2; and Paul Boghossian and J. David Velleman, "Colour as a Secondary Quality," *Mind* 98 (1989): 81–103.) I shall not undertake that exercise here, however, since my strategy for dissolving the sorites holds independently of any particular claims about the 'looks red'/'is red' relation.

Which then is the last red patch in the series? IP tells us that a difference of ten minutes could never make the difference between a time span of light that is normal and one that is not. Yet we know that if we judge each consecutive patch-plus-time-span along the series, we will eventually reach one that activates our NON-NORMAL category. How can this be? The answer, as before, is that category identity is preserved for pairwise judgments: for any two adjacent time spans, either both are judged normal or neither one is—*eo ipso* either both patches are judged red or neither one is—insofar as the two are judged pairwise. With respect to singular judgments, however, a Gestalt-like shift occurs at the time span that triggers a backward spread of our NON-NORMAL category on that run. (Here the shift is presumably nonperceptual, mobilizing different ways of thinking or conceiving of a given time span of light—*conceiving of it as*, rather than *seeing it as*, normal or non-normal.) After the shift, our judgments are biased toward a new, NON-NORMAL anchor; hence *ceteris paribus* were we to proceed backward down the series, some of the preceding time spans would also be judged non-normal. At some point, a new shift would occur and the process would repeat in reverse.

Now at this juncture in the analogous story for ‘looks red’, we concluded that IP was false. That is, we drew a conclusion about the correct application of ‘looks red’ from a story about our *judgments* of the patches. There such a move seemed unproblematic insofar as ‘looks red’ is by definition a judgmentally sensitive predicate. However, it will be objected that ‘is red’ is different: IP for ‘is red’ is a claim about patches that *are* red (here, that *are* disposed to look red in normal light), not merely those *judged* to be red (judged to have that disposition). Intuitively speaking, it doesn’t make sense to say that a patch could be red when judged singly but non-red when judged pairwise against a neighbor; a patch either *is* red or it *isn’t*, or else there is no fact of the matter, or something!<sup>21</sup>

Appealing though the latter conviction may be, it is mistaken; and in what follows I shall show that the sorites can be dissolved in essentially the same manner for ‘is red’ as for ‘looks red’. Before turning to that task, however, let me indicate briefly how my judgmental story would go for a wholly nonobservational word like

---

<sup>21</sup>I thank Beth Cohen for this vigorous formulation of the worry at issue.

'rich'. (Inasmuch as my solution is supposed to accommodate vague predicates generally, we shall do well to keep a contrasting example in mind.) Consider a series of dollar amounts progressing from one that clearly makes a person rich (say, twenty million dollars) to one that clearly doesn't (say, one dollar), so ordered that each amount contains one dollar less than the preceding. Which is the last "rich" amount in the series? IP tells us that one dollar could never make the difference between being rich and being non-rich. Yet if we judge each consecutive amount along the series, we eventually reach one that activates our NON-RICH category. How is this possible? As in the observational cases, consecutive amounts are category-identical when judged pairwise: either both amounts are judged rich or neither one is. With respect to singular judgments, however, a Gestalt-like shift occurs at the amount that triggers a backward spread of our NON-RICH category. (Here again the shift is presumably nonperceptual.) *Ceteris paribus*, were we to proceed backward down the series, some of the preceding amounts would be judged non-rich as well.<sup>22</sup>

The question then is this: what, if anything, about the correct application of words like 'red' and 'rich' follows from these stories about our judgments?

## 5.

As everyone knows, the apparent-color of an object can vary with both viewing conditions and viewer. For example, an object may look one way (say, green) in sunlight and another way (say, brown) in incandescent light, even where both kinds of light are normal (see Hardin, *CP*, 47). In such a case, I think we will—or at least should—say that the color of the object varies as well: it is green in sunlight and brown in incandescent light. As Simon Blackburn notices,

---

<sup>22</sup>For empirical evidence that non-perceptual or "generic knowledge" categories (like RICH) function in a manner analogous to that of perceptual categories (like LOOKS RED) in respects relevant to the present discussion, see D. L. Medin and L. W. Barsalou, "Categorization Processes and Categorical Perception," in *Categorical Perception*, ed. Stevan Harnad (Cambridge: Cambridge University Press, 1987), 455–90.

[o]ur attachment to single colours for objects is quite casual. Metamerism is not usually described as things of the same colour, which appear different from each other in different lights, but things of the same colour in one light, and different colour in different lights.<sup>23</sup>

Or an object may look one way to one (normal competent) subject under normal conditions and another way to another subject. For example, Hardin cites a study of fifty normal subjects in which “unique green was located by five subjects at about 490 nm, by eleven at about 500 nm, by fifteen at around 503 nm, by twelve at about 507 nm, by five around 513 nm, and by two at 517 nm” (*CP*, 162). Here we should say for instance that an object that reflects 490 nm is both unique green (relative to the first group of subjects) and not unique green (relative to the others). As Colin McGinn observes, it is “entirely proper to speak of objects as red with respect to perceiver *x* and green with respect to perceiver *y*. . . . There is thus a sense in which an object has (or could have) *many contrary colours simultaneously*” (*SV*, 9–10; my emphasis).<sup>24</sup>

Granted, the variabilities just cited are peculiar to an R2-type or “response-dependent” analysis of ‘red’—that is, an analysis that makes essential appeal to the responses of some subject or subjects—whereas my account is meant to be neutral regarding the analysis of the term. However, at present I am simply trying to show that the idea of context-relative variation in the extension of ‘red’ is a commonplace. And in any case, such context relativity is not restricted to observational words. Just for example, a man with one million dollars can be both rich and non-rich—rich relative to, say, welfare recipients in the South Bronx, non-rich relative to denizens of Beverly Hills. Or an object that reflects light of 643 nm may reflect roughly 650 nm relative to one set of explanatory interests and not roughly 650 nm relative to another. (The latter example illustrates that even a ‘red’ that means ‘reflects roughly 650 nm’ will sustain context-relative variation in its extension; a response-dependent analysis is not required.) Indeed, presumably the extension of any vague predicate will exhibit a similar variability; as

---

<sup>23</sup>“Circles, Finks, Smells, and Biconditionals,” in *Philosophical Perspectives* 7: *Language and Logic*, ed. J. Tomberlin (Northridge: Ridgeview Publishing Company, 1993), 259–80.

<sup>24</sup>On some of the reasons why such a pluralism does not violate claims of color incompatibility, see McGinn, *SV*, chap. 3, and Watkins, *DOA*.



Hans Kamp remarks, “[i]t is typical of a vague predicate that what objects it is true of depends on the context in which it is used” (PH, 242).<sup>25</sup>

What is distinctive about the sorites is that it exposes the variability of the correct application of a vague predicate even *independently of any variations in “external” context* (including variations in the identity of the judging subject). Hold fast the subject and the entire external context, including the number and arrangement of patches in the series, the time spans of light in which they look red, the time span of light that is normal for viewing them, and so forth, and still the extension of ‘red’ will vary. Hold fast the entire external context, including the judging subject and such factors as the operative standard of wealth (are we talking about Beverly Hills or the South Bronx?), and still the extension of ‘rich’ will vary. Why? For at least three now familiar reasons.

First, the extension of a vague predicate varies with what I shall call the *presentational context* of the items in the series: are they being judged singly, or pairwise, or threewise, or otherwise? Second, the extension of a vague predicate varies with the *historical context* of the items on any given run. For example, in the series of eighty-four patches discussed above, #43 might be red relative to some judgmental histories and non-red relative to others. Perhaps it is red, when judged singly, given a history in which the subject has previously judged consecutive patches beginning with #1; or non-red, judged singly, given a history of consecutive patches beginning with #84; or red, judged pairwise against #44, given a history of consecutive pairs beginning with #15/#16, progressing up through #68/#69, and then doubling back down to #43/#44; and so on. Third, the extension of a vague predicate varies even given a single presentational context and judgmental history. For instance, the color of #43, judged singly, will vary from one run to the next even given the same history of consecutive singular judg-

---

<sup>25</sup>For empirical corroboration, see for example Barsalou, IGS; E. M. Roth and E. J. Shoben, “The Effect of Context on the Structure of Categories,” *Cognitive Psychology* 15 (1983): 346–78; and Kahneman and Tversky, “Prospect Theory: An Analysis of Decision Under Risk,” *Econometrica* 47 (1979): 286–88. In an especially apt passage, Kahneman and Tversky remark that “[t]he same level of wealth, for example, may imply abject poverty for one person and great riches for another—depending on their current assets” (277).

ments beginning with patch #1. Why? Because of brute mechanical variations in the operation of our categorizing homunculus: as a matter of psychological fact, our categorizer just does get tired, or bored, and so shifts to a new elastic band, at different places on different runs. The exact location of the category shift is, in this sense, arbitrary.<sup>26</sup> To refer to this third kind of variation, I shall say that the application of a vague predicate varies with the *categorical context* of the items in the series.

No doubt one could isolate still further species of contextual variation; the foregoing inventory is not meant to be exhaustive. But at present the important point is this: the extension of a vague predicate varies with more than just the relevant *external* contexts; in particular, it varies with (at least) the three kinds of contexts just described. These latter are plausibly conceived as purely judgmental or *psychological* contexts. Where the correct application of ‘red’ varies in any of these three ways, the only difference between circumstances in which (for example) #43 is red and circumstances in which it is non-red lives in the psychology of the competent judging subject. *Ex hypothesi* no variation occurs in the “extramental” world. Similarly, the extension of ‘rich’ will vary with the presentational, historical, and categorical contexts of the monetary amounts in its series, independently of any “external” variations. Granted, nothing in my story requires that these contexts be viewed as psychological per se; perhaps there are better ways to conceive them. All my story requires is that there *be* such contexts, psychological or otherwise, and that the extension of a vague predicate vary therewith in the manner outlined above. For present purposes, however, insofar as the extension of a vague predicate varies in any of the three ways just described, I shall say that it varies with the psychological or “internal” contexts of the items in the series.

Lest you harbor any doubts that the *correct* application of a vague word varies with psychological context, consider the following. First, if the extension of a predicate does not vary in the ways just described, or something very like them, then the predicate is non-vague. Think of it this way: if the extension of the predicate is invariant with psychological context, then once the relevant exter-

---

<sup>26</sup>Incidentally, such a claim neither precludes the possibility of error nor impugns the rationality of dispute concerning the application of vague words; see section 6 for elaboration.

nal, or nonpsychological, contexts have been fixed, the category shift will occur at the same place on every run; in other words, there will be a sharp and fixed boundary in the series. Second, since our actual applications of vague predicates vary with psychological context, it follows that if the (true) extensions of these predicates do not thus vary, then we are linguistically incompetent in their use. But of course we are not incompetent in the use of these words. Finally, apart from the psychological considerations lately discussed, what sorts of “facts” could determine the extension of a vague predicate in borderline cases? (I shall explore this question further in section 6.) And for that matter, if there were such facts, how could we know about them?<sup>27</sup> The difficulties seem overwhelming.

*N.B.* I do not claim that the *meaning analysis* or *intension* of a vague predicate includes a judgmental element. For instance, I do not claim that in calling an object red one means or is saying, in either the “speaker” or “semantic” sense, that the object has a judgment-dependent property, or, for example, that the object is merely red-relative-to-me-now or red-relative-to-such-and-such-a-context. Rather, in calling an object red, one is saying whatever the correct analysis of ‘red’ specifies—that is, that the object looks red under normal conditions, or reflects roughly 650 nm, or has a certain microstructural property, etc. I chose R2 as my exemplar simply because of its prevalence in the philosophical literature, and nothing in my account depends upon its special features. Rather, I claim that the *extension* of ‘red’—the class of objects that satisfy the predicate—is always relativized to certain psychological (and nonpsychological) contexts. The sorites is a puzzle about the correct *application* of vague predicates, and that is all my story addresses. All that was required to generate the paradox, remember, was the claim that a given predicate applied equally to both of two marginally different items.

A number of questions clamor to be answered here, and I shall address those in section 6 below. For now, so as not to lose the thread of our discussion, let me proceed with my solution to the paradox.

---

<sup>27</sup>See Horgan, RV, section 5, for trenchant criticism of the related idea that “hidden precisification-facts” determine sharp boundaries for the application of vague predicates.

For all I have said so far, the application of ‘red’, like that of ‘looks red’, is episodic: I said, for instance, that #43 might be red *on one run* and non-red *on another*. However, an episodicity claim seems implausible for ‘red’ (if there is no such word as ‘episodicity’ I hereby coin it); in other words, it seems implausible to suppose that color predicates, a fortiori vague predicates generally, apply to objects only during occurrent episodes of judgment. Intuition dictates that any visible object has some color or other at all times, whether or not it is being judged.<sup>28</sup> Happily, nothing in my view requires an episodicity claim. For ease of exposition I have spoken of particular series of applications of ‘red’ and ‘non-red’ to the patches, but my account can be “de-temporized” straightforwardly: just as an object may have many contrary colors simultaneously *relative to* distinct external contexts, it may have many contrary colors simultaneously *relative to* distinct psychological or internal contexts. For example, #43 may be red relative to one historical context and non-red relative to another. More importantly, *adjacent* patches may satisfy incompatible predicates simultaneously relative to distinct psychological contexts. For instance, if a category shift occurs (or would occur) between #43 and #44 on some run along the series, then #43 is red relative to one categorial context while simultaneously #44 is non-red relative to another. Clear cases are simply limiting cases that lie in the same category relative to all psychological contexts.<sup>29</sup> Thus relativity to psychological context is just a further, rather fine-grained species of contextual relativity.

Not surprisingly, the individuation of psychological contexts—especially categorial ones—promises to be a delicate affair. For ex-

---

<sup>28</sup>I say that an episodicity claim *seems* implausible for ‘is red’. In fact, I think an episodicity claim is correct for vague predicates generally, not just for apparent-color words. However, because such a view is initially counterintuitive, and because my solution to the sorites does not require it, I take a more moderate line here. I defend an episodicity claim for vague predicates in general in another paper, “Episodic Properties” (in progress).

<sup>29</sup>Indeed, a patch like #1 is presumably red relative to all relevant psychological and *nonpsychological* contexts. One might introduce a special predicate, ‘super-red’, to distinguish patches like #1 from “multi-colored” patches like, clearly, #43. Of course, ‘super-red’ will itself be vague; hence just where in the series the super-red patches end and the multicolored patches begin will itself vary from context to context.

ample, does the pre-shift portion of each run on which the shift occurs *at a certain patch* (say, at #55) instantiate the same categorial context type, or are the categorial contexts instantiated by the pre-shift portions of all runs type-identical? In a different vein, consider again the run described above, on which the shift to ‘non-red’ occurs at #44. Consider the categorial context, instantiated on that run, relative to which #43 is red. What color is #44 relative to *that* context? (In fact I think we should say that #44 has no color relative to that context; but I shall not argue the point here.)<sup>30</sup> One could go on. Fortunately our solution to the sorites need resolve none of these complexities. All it requires is that any differences in color between adjacent patches be relativized to distinct contexts, and of that we can rest assured. At a minimum, any such adjacent difference is grounded in the occurrence of a category shift, and a category shift *always introduces a new categorial context*. Whatever color(s) #44 may be, it is not non-red (nor any other color incompatible with red) relative to any context relative to which #43 is red. More generally, adjacent patches do not differ in color relative to the same categorial context.

Now we can solve the puzzle. The paradoxical argument for ‘red’, like that for ‘looks red’, is guilty of an equivocation:

- (1) Patch #1 is red, relative to a singular presentational context (and any given historical and categorial contexts).
- (2) For any  $n$ , if patch # $n$  is red then patch # $(n+1)$  is red, relative to a pairwise presentational context (and any given historical and categorial contexts).
- (3) Therefore patch #84 is red, relative to a singular presentational context (and any given historical and categorial contexts).

The absurd conclusion follows only on the false “singular” construal of IP—that is, the construal of IP as claiming that for any  $n$ , if patch # $n$  is red relative to a singular presentational context, then patch # $(n+1)$  is red relative to a singular presentational context. As we have seen, relative to singular presentational contexts, adjacent patches can

---

<sup>30</sup>This is not to suggest that #44 has no color(s) *at all*; my thought, rather, is that #44 has its color(s) relative to contexts other than the one at issue. But of course I would need to say more.

differ in color. This is because a category shift occurs (or would occur) between some two adjacent\* patches on any complete run of singular judgments along the series; hence any (standing, “de-temporized”) categorial differences in the colors of adjacent patches result from the application of the corresponding predicates *relative to different contexts*. For example, #43 is red relative to one categorial context and #44 non-red relative to another. Generally speaking, adjacent items can lie in the extensions of incompatible predicates relative to different psychological contexts. And no one ever thought *that* sort of difference couldn’t obtain in the series.

Thus the paradox for ‘red’ dissolves. Similarly, *mutatis mutandis*, for ‘rich’ and other vague predicates. The puzzle for ‘heap’, for example, is solved by appeal to the fluid categories HEAP and NON-HEAP, the puzzle for ‘person’ by appeal to PERSON and NON-PERSON. In addition, it should now be obvious that anyone troubled by our episodicity claim for ‘looks red’, and/or our *identification* of looking red with activating a perceptual category, can solve the puzzle for ‘looks red’ in the same way we have solved it for ‘is red’—that is, by ascribing multiple simultaneous context-relative *apparent*-colors to the patches in the series. As I have emphasized, the sorites is solved independently of any particular analysis of the predicate (or metaphysical account of the corresponding property) at issue. Moreover, unlike some other treatments of the puzzle, ours has no need of such legislations as a weakened validity for modus ponens or a nonstandard semantics for the universal quantifier, nor for the machinery of definiteness operators or degrees of truth. Best of all, there is no need to deny what seems undeniable—that predicates like ‘red’ and ‘rich’ apply equally, if at all, to marginally different items. Properly understood, the falsehood of IP poses no threat to our pretheoretic intuitions.

I conclude by addressing some outstanding questions.

## 6.

It follows on my view that once the relevant external, presentational, and historical—in other words, all but the categorial—contexts have been fixed, an item lies in a given category if and only if the relevant competent subject(s) would judge it to lie in that

category.<sup>31</sup> One might have thought that such a biconditional (call it *B*) holds only for borderline cases, that the status of clear cases is wholly judgment-independent: for example, it is just a “mind-independent fact” that patch #1 is red, or that J. Paul Getty was rich, or that J. Paul Getty was a person. However, the relevant difference between borderline and clear cases lies in the truth *makers*, not the truth *values*, of their respective biconditionals. Specifically, *B* is true with respect to borderline cases because our competent judgments of borderline cases determine their category memberships; conversely, *B* is true with respect to clear cases because clear cases determine what counts as a competent judgment. We are inclined to say that judging patch #1 red or J. Paul Getty rich is (partially) *definitive* of our competence with the corresponding predicates. Thus the category memberships of borderline and clear cases alike covary with our competent judgments (that is, *B* is true with respect to both borderline and clear cases), but the “direction of determination” is reversed between the two. Of course, just where in the series that direction shifts, that is, just where in the series the clear cases end and the borderline cases begin, is itself vague and hence variable from context to context (again, see note 31.)

Doesn’t an endorsement of biconditional *B* amount to claiming that once the relevant external, presentational, and historical contexts have been fixed, a competent subject cannot be mistaken in his application of a vague predicate? Yes—but the infallibility is

---

<sup>31</sup>Of course, ‘is competent’ is itself a vague predicate; hence on my view someone is a competent subject if and only if competent subjects would judge him a competent subject. Are we faced with a regress? Yes, but not a vicious or otherwise problematic one. In the first place, remember that the biconditional just stated is not a meaning analysis; rather, it expresses a (mere) covariation relation between competent judgments and the *extension* of the vague predicate ‘is competent’. (And in any case, it is no part of a solution to the sorites that one specify, or even be able to specify, who the competent subjects are.) Secondly, and more importantly, the regress is stanchied at any stage by a relativization to *our own* judgments as to who is and is not competent—in other words, to our own application of the predicate relative to a given context. Naturally, whether we count someone a competent subject depends largely on whether he agrees *with us* in the relevant cases; but that is as it should and *must be*. *Ceteris paribus*, we must take ourselves to be competent subjects; we must, as it were, stand somewhere. What alternative do we have?

highly circumscribed. Among other things, if he judges that patch #1 is non-red, he is incompetent, and if he is wearing rose-colored glasses, then all bets are off. To boot, he can always relativize to the wrong context—for example, to Beverly Hills instead of the South Bronx, or to a pairwise presentational context instead of a singular one, or to a judgmental history beginning with the first item in a series instead of the last. Opportunities for error abound. My present claim is simply that where the relevant contexts have been established, the possibility of error does not arise. We can be wrong about whether an object reflects precisely 650 nm, or whether a man has precisely one million dollars, or whether a fetus is precisely seventeen weeks old, but not about whether the object reflects roughly 650 nm, or whether the man is rich, or whether the fetus is a person.<sup>32</sup>

Perhaps it will be thought that the idea of a brutally mechanical, hence arbitrarily located, category shift is plausible for observational words like ‘red’ but not for nonobservational words like ‘rich’ and ‘person’. Such a worry is easily dispelled, however. For if the (exact) location of the shift is not arbitrary, if the shift is made at one dollar amount or fetal age # $n$  as opposed to # $(n+1)$  or # $(n-1)$  *for a reason*, as it were, then one of two things is the case: either (i) the predicate at issue is non-vague (for example, ‘rich’ defined as ‘possesses at least one million dollars’), or (ii) the difference between adjacent items in the series is not sufficiently small, that is, not marginal in the sense required to generate a paradox. A sorites series just is a series in which the difference between adjacent items is sufficiently small that any categorial distinctions between them must be drawn arbitrarily. Crudely: when adjacent items are enough alike, it doesn’t matter where exactly you draw the line. (Of course, this arbitrariness goes hand in hand with the infallibility mentioned above: if the exact location of the shift is arbitrary, then there is no getting it wrong.) Granted, # $n$  must be a borderline case; in other words, there is reason to shift at, say, patch #44 rather than at patch #2, and you are liable to censure if you do otherwise. But ‘borderline case’ itself is vague, and so what counts as a borderline case,

---

<sup>32</sup>See McGinn, SV, chapter 4, on the infallibility of secondary-quality ascriptions; I extend the claim to vague predicates generally.



hence where reasons begin and end in the series, varies relative to both external and brute internal contexts.

Ironically, it may be thought a strike against my solution that it applies uniformly to predicates as diverse as 'red', 'rich', and 'person'. Surely the differences among these words—isn't 'rich' a "vaguer" predicate than 'red'? isn't there a fact of the matter as to whether something is a person, but not as to whether someone is rich? etc.—ought to be effective somewhere in my story. In reply, let me first point out that we should expect a uniform solution to the puzzle: again, all that was required to generate it was the claim that a given term applied equally to both of two marginally different items; and that much is true of any vague or tolerant predicate. But furthermore, I think the intuitive disparities among vague words can be accounted for independently of the details of my solution. Just for example, perhaps 'rich' seems "vaguer" than 'red' because 'rich' does not admit of paradigm cases in the way that 'red' does. Whereas the notion of a maximally (that is, pure or unique) red object makes perfectly good sense, the notion of a maximally rich person does not; by the same token, whereas an object cannot grow redder *ad indefinitum* (at least not in our experience), one can become richer and richer without end.<sup>33</sup> Perhaps the extension of 'person' seems judgment-independent because one so rarely encounters borderline cases of persons, or because so much moral and political freight rides on a decision in borderline cases, or because the law often *requires* a decision in borderline cases, or because we are subtly influenced by religious views that wrongly take 'person' to be non-vague. And so forth. (Of course, the frequent need for judicial decision in borderline cases of personhood only underscores the judgment-dependence of the predicate's application.)

Reflection on 'person' may raise a related worry that the brute psychological variability featured in my account undercuts the rationality of dispute over the correct application of vague words. We spend a lot of time and money arguing over the extension of 'person', for one; does my view suggest that such activity is so much idle mouthing?<sup>34</sup> Hardly. In the first place, I do not deny

---

<sup>33</sup>It may seem that having zero dollars would constitute a case of maximal *non*-richness; but of course one can always go into debt.

<sup>34</sup>I thank Michael Watkins for pointing out the need to address this question here.

the importance of the “rational” determinants of our judgments. As I have explained, the correct application of a vague predicate demands that we relativize to the right context, verify that conditions are normal, and so forth; and when we err in such matters we are liable to censure. But furthermore, even where disparities in our (correct) applications of a word like ‘person’ owe to brute psychological differences among us, the activity of dialectical interchange may be rational insofar as it fosters consensus. It may be biologically adaptive, and in any case practically necessary, for a community to reach consensus on issues of substantial social effect, even where the decision at issue must be largely arbitrary. To that end, you may try to influence the location of my category shift toward the “early” end of a sorites series for personhood by teaching me about fetal responses to noxious stimuli, while I may seek the opposite effect in you by noting the cognitive inferiority of human infants to adult horses and dogs. As Alan Gibbard observes, “we simply are influenced in a fundamental way by what those around us think.”<sup>35</sup>

Perhaps, after all, this ending seems a disappointment. A certain letdown often accompanies the death of a longstanding problem, especially one as venerable as the sorites. Worse yet, the paradox succumbs to largely empirical considerations: it is because of the way certain psychological categories behave that our use of vague predicates is not—at least not on account of any sorites paradox—incoherent. What greater affront to logical mystery than to be stripped by brute empirical fact?

Then again, perhaps it isn’t disappointment at all. Perhaps it is skepticism—skepticism that a story about category spreads and little men in the head should have any import whatever for a problem about the correct application of words. In reply let me say that I appreciate the worry. The consortium of semantics and psychology is hardly transparent, and I have but scratched its surface here; no doubt formidable difficulties lie in store for any explorations in the area. Nevertheless, I am convinced that a semantic theory for the terms at issue must accommodate the judg-

---

<sup>35</sup> *Wise Choices, Apt Feelings* (Cambridge: Harvard University Press, 1990), 177.

mental considerations assembled above. Quite possibly, the blurred semantic boundaries of a vague predicate reflect the fluid judgmental spreadings of a psychological category.<sup>36</sup>

*Ohio State University*

---

<sup>36</sup>I am grateful to Michael Watkins, George Schumm, and Donald Hubin for countless illuminating discussions; also, for generous and helpful commentary, to Ruth Barcan Marcus, William Lycan, Robert Kraut, Jefferson White, Stewart Shapiro, Beth Cohen, William Taschek, Jonathan Vogel, George Pappas, Barbara Scholz, Cy Anders, Robert Levine, Terry Horgan, Calvin Normore, Hans Kamp, Craig Roberts, Marshall Swain, Daniel Farrell, Simon Blackburn, Steven Boer, Jody Graham, Lester Krueger, Tom Nygren, and the participants at a conference on Theories of Partial Semantic Information at the Center for Cognitive Science, University of Texas at Austin, January 1990.