Further Reflections on the Sorites Paradox
Author(s): CRISPIN WRIGHT
Source: *Philosophical Topics,* Vol. 15, No. 1, Philosophy of Language (SPRING 1987), pp. 227-290
Published by: University of Arkansas Press
Stable URL: https://www.jstor.org/stable/43153999
Accessed: 22-10-2018 06:33 UTC

Further Reflections on the Sorites Paradox

CRISPIN WRIGHT
University of St. Andrews/University of Michigan

Preamble

In two papers published just over a decade ago,[1] I argued that the major premises — usually universally quantified conditionals[2] — which feature in Sorites paradoxes owe their plausibility to an assumption which it then seemed appropriate, from the perspective of a philosopher of language in Oxford, to dub the "governing view." The governing view was a combination of two claims. The first was that a master of a language is someone who has, at some deep level, internalized a definite set of semantic and syntactic rules, definitive of the language. The second was something of a *pot pourri* of specific ideas about how such rules might be brought into the light or more fully characterized and understood. It included the admissibility of, for instance, considerations concerning speakers' known limitations (of memory or perceptual acuity, for example), standardly accepted criteria of (mis-)understanding, salient features of standard linguistic training, and the purposes and role which a particular distinction might have. The argument was that once the *relevance* of such considerations was granted, the major premises for Sorites paradoxes could be seen as guaranteed by aspects of the semantic rules governing the relevant expressions (together with such general facts as the non-transitivity of indiscriminability). Since the other premises involved in such paradoxes, and the principles of inference which they utilize, seem incontestable, the conclusion suggested was that the governing view was responsible for them. That (at least some) Sorites paradoxes are *genuine* — do no more than reflect features of the meanings of the affected expressions and aspects of the world, and so commit no identifiable fallacy nor involve any identifiable error — is a view which has been adopted by a number of writers, most

227

notably by Dummett.[3] My suggestion, in contrast, was that this is an illusion, generated by the governing view; specifically, that its second, epistemological ingredient is unacceptable, and that its first ingredient — the overall conception of language mastery which it embodies — needs, at the least, severe qualification.

Despite their antiquity, the Sorites paradoxes had received comparatively little attention in the literature prior to 1975. Since then, there has been something of an explosion. One purpose of the present paper is to try to respond to various criticisms of my earlier efforts which have appeared, notably in Christopher Peacocke's very interesting [1981].[4] I now wish to restrict the scope of my original proposal. Major premises for Sorites paradoxes can actually be motivated in a number of quite different ways, which it is essential to distinguish. Only for a limited class of cases does it now seem to me that my earlier diagnosis is apt. The widespread conviction that the paradox afflicts all vague expressions (and hence virtually all our language) is best represented as the product of a different and, I believe, confused line of thought. And we have to reckon also with a third line of thought, leading to what I have called in this paper the Tachometer Paradox. Moreover this threefold division is probably not exhaustive.

Even in the restricted class of cases in which I still wish to press my original proposal, I am no longer content with the kind of formulation which I offered. Where earlier I spoke, vaguely, of the inappositeness of the comparison of language with a game, and of the need for a more purely 'behavioristic' conception of language-mastery, I now wish to offer something different. It is less that I think these formulations wrong than that they appeal to contrasts which may seem unclear and which are anyway very difficult to sharpen. A first statement of the particular moral which these formulations tried to capture would better proceed, it seems to me now, in somewhat different terms.

I have tried, so far as is possible, to write the present paper in a fashion which will not presuppose that a reader is familiar with my earlier two papers. But since my principal objective has been to improve on them, it is unavoidably, to that extent, a sequel.

I

The Governing View and the Sorites Paradox

There is an idea about the kind of ability which mastery of a language (basically) is which would figure in virtually anyone's first philosophical

228

thoughts on the topic. Suppose we had to teach somebody to play chess without any kind of verbal or written exchange. Could it be done? There seems no reason why not; no reason, that is, why an intelligent and receptive subject could not acquire a knowledge of the moves and of checkmate and stalemate, and so on, just by immersion in the practice of the game, suitably reinforced by nods and shakes of the head or punishment and reward. Such a subject might become an average or better than average chess player without ever learning any of the vocabulary which we use in description of the game. However once there was unmistakable evidence of mastery, we would not hesitate to attribute to the trainee the knowledge which *we* express by a statement of the rules and object of the game and also, if the trainee were an accomplished player, the knowledge which we can express by a description of certain strategies and of their rationale. We would think of the trainee as having this knowledge although unable — for the time being anyway — to talk chess-theory at all. Such knowledge would be viewed as *implicit*; but its content would coincide with that of the sort of explicit articulate knowledge which is normally acquired when someone learns to play chess.

This way of describing the example seems natural, almost inevitable, because the ability in question is par excellence a rational one, involving intelligence, insight and purpose. If recourse to rationalistic explanation — explanation involving attribution of beliefs, goals and intentions — were for some reason excluded, we simply should not know how to begin to describe the trainee's accomplishment. But beliefs, e.g., are contentful states; and we should be equally at a loss if none of the beliefs allowably attributable to the trainee could have a content which we would specify using the standard vocabulary of chess. So provided the eventual performance were good enough, the idea that the trainee possessed implicit knowledge of the rules of chess would be just as attractive if he or she mastered chess before acquiring *any* linguistic competence; or, indeed, if the trainee were a deaf mute whose prospects of over understanding a language were slight.

The first philosophical thought about language mastery to which I allude is, in effect, a large-scale analogue of the foregoing. Language mastery, like mastery of chess, is apt to seem an exemplar of rational ability, differing only in that the set of rules — syntactic, semantic, and pragmatic — whose knowledge it involves are inordinately more complex than the rules of chess, and the purposes to which it is adapted far more various. Yet we perforce acquire language mastery in much the manner of the

229

fictional chess trainee, at least when learning a first language: we are immersed in the practice of those who can already 'play', and explicit verbal instruction can be received only when enough of the language has already been grasped to enable an understanding of the terms in which it is couched. No doubt a good deal of the instruction *is* explicit — in typical school curricula a great deal is given in the way of exact definitions of terms, and precise characterizations of what is good grammar. But it remains true that the greater part of anyone's competence in their native language is not and could not be acquired that way.

In short: it comes naturally to think of someone with a good level of mastery of their native language as knowing a great deal more than they were ever explicitly taught; and it follows that there can be at best a contingent connection between the possession of this knowledge and the ability to articulate it. If the chess trainee's performance when playing the game is completely convincing, we shall not be any less inclined to credit him with knowledge of the rules of chess should it prove, when he acquires explicit chess vocabulary, that he finds it difficult to articulate them for himself. So too with language mastery. The idea of implicit knowledge thus has just the same attractiveness, just the same seemingly inevitable part to play in the explanation of linguistic ability, that it should have in accounting for the performance of our fictional chess trainee.

The conclusion of this way of thinking about the matter, then, is that, as someone proceeds towards mastery of a first language, it must be true at certain stages, and may be true at any later stage, that some or all of the knowledge which they have acquired is implicit. But that, it seems, need pose no barrier to the possibility of an explicit description, either in a different language or in a larger fragment of the same language, of what is thereby implicitly known. On the contrary, since this 'implicit knowledge' has been characterized from the outset as a contentful state, as in effect propositional knowledge, it must be possible in principle to specify the contents known by language learners, even if they themselves cannot yet do so or will never be able to do so. It would follow that there must be a formulable theory which stands to the competence of speakers of English rather as a complete codification of the rules of chess stands to the competence of someone whom we credit with a complete implicit knowledge of those rules but who is able to give, for whatever reason, at most a partial characterization of them. But there is a difference. Chess is a complex enough game, but it would nevertheless be surprising if, once introduced to the names of the pieces and other terms of art like 'castling'

230

and 'checkmate', a subject whose knowledge of chess had originally been implicit did not prove to be able, on relatively brief reflection, to produce a passable statement of the rules of chess. In sharp contrast, it is obvious even on superficial reflection that the task of producing an explicit statement of the rules governing just the fragment of English concerned with literal statement making would be one of the greatest difficulty, even if we restricted our concern to legality, as it were, and prescinded from all consideration of what constituted 'good play'. We are as if in the situation of people who have all been taught, by directly practical methods and with a minimum of formal explanation, to play a game hugely more complex than chess, but now find that all the instructors — those who might have produced a written codification of the rules of the game if asked — have disappeared. We therefore confront, if we choose to rise to the challenge, an intriguing project: work out what the rules are. The result, if the task could be accomplished for an entire natural language, would be something that nobody has ever beheld: a complete explicit formulation of the knowledge which any master of the language knows but knows (mostly) only implicitly.

These ideas represent one possible route into sympathy with the modern philosophical project of a 'theory of meaning'. (They may, indeed, represent the only such route, though that is a nice question.[5]) I am not recommending them; I hope only to have made it intelligible, if there is anyone to whom they have never occurred, how they might nevertheless constitute an attractive vision. Their significance in the present context is that the Sorites paradox, or so I have argued in my [1975] and [1976], casts serious doubt on their acceptability in general. That is one claim which I wish to review in this paper. The claim is important; for the most fundamental task in the philosophy of language is the achievement, in the most general terms, of an understanding of the nature of language mastery. It is here, if I am right, that the paradox has most to teach us.

If the tension is to be perceived, however, we have to ride the analogy a little bit harder. Consider the task of trying to draw up the rules of a game which we have all been able to play since childhood but of which there is no extant codification. How should we set about it? No doubt we should first need to invent some vocabulary to describe the materials and situations which are specific to the game; but then what? Recognition of one's own hitherto unformulated intentions is a peculiar business in any case; but the natural approach would be for a number of competent players to draw up severally as complete a description as they could by reflection,

231

and then to compare notes. But when, as is likely, differences of opinion are revealed, how should they be resolved? The extent of the consensus on a particular point would be important, of course. But if somebody, e.g., listed a rule which no-one else had, that would not necessarily be decisive against that particular rule. If, for example, there was universal consensus that behavior in the course of the game which was in fact a breach of the putative rule would be regarded as outré and unacceptable, we would probably take that as strong prima facie evidence that the rule should be included and that the majority who did not list it had overlooked something. If, on the other hand, the putative rule seemed too complex, or incorporated distinctions too subtle to be remembered with any degree of surety, or relied on our ability to make contrasts which we felt that the direct, 'immersive' training which we had received could not have possibly got across — if any consideration of that sort applied, there would be a strong case for discounting the rule.

Perhaps the most intriguing thing about Sorites paradoxes is that an important class of them seem to arise very directly from the application of this sort of thinking to the rules of which implicit knowledge is supposed, on the other side of the analogy, to constitute our language mastery. Once the content of those rules is supposed to be constrained by considerations like the limitations of the methods whereby we were trained in them, and our limitations, e.g., of memory, and the criteria for correct understanding of them which we now find intuitively satisfying, it is relatively straight-forward to argue that they imply the unrestricted acceptability of the major premises — standardly, universally quantified conditionals — involved in Sorites paradoxes.

The arguments to this effect developed in my [1975] and [1976] were various and I shall give only the most general indication of them now. One thought was that it could not be part of understanding an expression to be able to make *unmemorable* distinctions if the learning and use of the expression standardly involves no reliance on external aids of any kind. Another was that distinctions too fine to be detected by casual observation could not be incorporated into the conditions of application of a predicate like 'bald' or 'heap' whose utility depends on being applicable on the basis of casual observation. A third was the thought that the conditions of application of an expression which, like 'adult', is associated with sub-stantial moral and social significance, cannot incorporate distinctions too refined to sustain that significance. A fourth was that the conditions of application of any ostensively teachable expression cannot incorporate

232

distinctions—in particular the kind of distinction that may obtain between observationally indiscriminable items—not amenable to ostensive display. A fifth was that no distinction can mark a watershed in the conditions of application of an expression if treating it as such would standardly be taken to display a misunderstanding of that expression.

Typically, the major premises in plausible Sorites paradoxes can be supported by considerations of more than one of these kinds. The paradox for 'red', for example, may draw on any, or all, of the first, second, fourth and fifth. A paradox for 'heartbeat of childhood' may draw on either the second or third. And so on. The exact manner in which the major premises are supposed to follow from such considerations will bear rather more careful scrutiny than I gave it before and we shall return to the question in section VII. But what seems beyond dispute is that if we accept that linguistic competence is constituted by sensitivity to the dictates of inter-nalized rules, it is overwhelmingly natural to suppose that features of these rules may be discerned by reflection on our practical limitations, the ways the application of an expression would be defended if challenged, the interest which we attach to the classification effected by it, the way it is standardly taught, and the criteria for someone's misunderstanding it.

How we should respond to the situation depends, of course, on whether the specific arguments are truly watertight. But there is little at least in the literature with which I am familiar which tries to meet the problem head-on, to show that it is simply mistaken to think that relevant major premises can be validated in this way. For the present, let us continue to suppose that they can.

On that assumption, the position which we arrive at is as follows. Two separable claims have been made (together, they constitute the "governing view"). There is the general claim that our language mastery is to be seen as the product of (mostly) implicit knowledge of rules; and there is the more specific claim that the enterprise of attempting to arrive at an explicit statement of these rules — or to enlarge on features of them which need not be reflected in such a statement, if we are doing semantics in the 'homophonic' mould — is rightly seen as constrained by considerations of the seemingly platitudinous sort outlined. If both these claims are accepted, we are faced with the realization that Sorites paradoxes (or at least some of them) merely serve to unravel certain features of the semantics of the expressions with which they deal, which must accordingly be viewed as at least *de facto* incoherent; that is, the rules for the use of those expressions, taken in conjunction with undisputed features of the world,

233

enable flawless cases to be made for simultaneously withholding them from and applying them to certain situations.

There are then only three possible modes of response. The first is acceptance of both claims, with their consequence that our use of a large class of expressions is governed by incoherent rules. Second, we might essay to retain the first claim while jettisoning the second: the result would be to hold that while the implicit-knowledge conception of our language mastery is fundamentally correct, the platitudinous-seeming constraints on the explicatory enterprise are not. Third, there is the apparently radical step of rejecting the first claim. How deep such a rejection would have to go would be a matter for determination, but, at the very least, it would be necessary to construct some alternative conception of our mastery of the use of Sorites-susceptible expressions, to supplant the idea that it is comprised by implicit knowledge of the requirements of rules.

It is important to be clear that there cannot be any further alternative. (There is, of course, no possibility of retaining the second claim at the expense of the first.) Or rather: the only alternative is to disclose error in the supposition that the two claims, in conjunction with further undisputed premises (e.g., that indiscriminability is a non-transitive relation), inescapably generate Sorites paradoxes. A fully satisfactory treatment of the problem must therefore accomplish one of two things: either it must demonstrate that the link between the two ingredient claims of the governing view and the major premises of Sorites paradoxes can be broken — in which case it will need, in addition, to explain how exactly those premises fail of truth —, or it must explain why one of the three resonses outlined is superior to the others and intuitively satisfying when properly fleshed out.[6]

II
Accepting the Paradox

Towards the end of his [1975] Dummett summarizes some conclusions as follows:

'(1) Where non-discriminable difference is non-transitive, observational predicates are necessarily vague.

(2) Moreover, in this case, the use of such predicates is intrinsically inconsistent.

234

(3) Wang's paradox merely reflects this inconsistency. What is in error is not the principles of reasoning involved, nor, as on our earlier diagnosis, the induction step. The induction step is correct, according to the rules of use governing vague predicates such as 'small': but these rules are themselves inconsistent, and hence the paradox....'[7]

This is a clear enough example of the first kind of response. Why not this response? Why should it not be the case that the 'language-game' is governed by (de facto) inconsistent rules? Clearly, appeal to our knowledge of such rules couldn't provide a completely satisfactory explanation of our actual linguistic practice — some supplementary account would be wanted of why we do not fall into confusion. But the idea of making some sort of successful enterprise out of an incoherently codified practice is not in itself incoherent.[8] So what, if any, is the real objection to it here?

There are, it seems to me, three related objections. First it is open to question whether the idea of inconsistency can actually get any grip unless we are concerned with rules which have been *explicitly* codified. Suppose that we play an inconsistently formulated game, but never notice the inconsistency, sometimes going by one rule in a particular kind of situation and sometimes going by another which conflicts with it. If a third party, trying to find out the exact character of the rules of the game, was restricted to observation of play and *never allowed to see a written version of the rules,* what reason could they possibly have to suspect that the game was inconsistent rather than that at least two different courses of action were permissible in the situation in question? The analogy is not, indeed, accurate, since it is not as if we sometimes go along with the results of Sorites reasoning but sometimes prefer the response which such reasoning can be made to contradict. But just for that reason a radical interpreter of our use of color vocabulary, for instance, would have even *less* reason to suspect an inconsistency than the observer of the game. If we say that there is an inconsistency nevertheless, with what right do we claim an insight into the character of the rules governing this vocabulary from which an observer of the way we use it is barred? The answer will be, presumably, that the methodology implicit in the second claim of the governing view already transcends the resources available to radical interpretation if the latter must prescind from consideration of our intellectual and sensory limitations, the kind of training which we receive, and so on. But then is not deliverance of so peculiar a conclusion some cause for disquiet with the richer methodology?

235

Second, what is actually responsible, on this view, for the large degree of coherence and communicative success which our use of color vocabulary enjoys? Indeed, what is the justification for continuing to think of the use of such expressions as governed by rule? Knowledge of appropriate rules was supposed to constitute linguistic competence. But it cannot do so if competent usage essentially has a coherence which, in Dummett's view, the rules lack. Dummett's response needs supplementing with an explanation of our communicative success with such vocabulary in which the idea of knowledge of inconsistent rules has an ineliminable part to play. For either such knowledge is still to be a basic ingredient in competence or we should drop the idea. But it is quite unclear, to me at least, how such an explanation should proceed.

That brings us to the third, and, I think, a decisive objection to Dummett's response. I do not see how we can rest content with the idea that certain implicitly known semantic rules are incoherent when *nobody's* reaction, on being presented with the purported demonstration of the inconsistency, i.e., the paradox — even if they can find no fault with it — is to lose confidence in the unique propriety of the response — e.g., "That's orange' — which the demonstration seems to confound. Think of your reaction when, having received as explanation of the notion of class only the usual informal patter plus the axioms of naive set theory, you first confronted Russell's Paradox. If, which is unlikely, you held any intuitive conviction about whether Russell's class was a member of itself, you will have been forced to recognize an exact parity in the opposing case; and the effect should have been to cause you to realize that there is just no view to take on the matter before some refinement of the notion of class has been made. But that is exactly *not* our response to the Sorites paradox for 'red'. Our conviction of the correctness of the non-inferential ingredient in the contradiction is left *totally undisturbed*. So far from bringing us to recognize that, pending some refinement in the meaning of 'red', there is just no such thing as justifiably describing something as 'red' or not, our conviction is that no-one *ought* to be disturbed by the paradox — and this conviction is not based on certainty that we shall be able to disclose some simple fallacy. If the rules for the use of 'red' really do sanction the paradox, why do we have absolutely no sense of disturbance, no sense that that a *real case* has been made for the inferential ingredient at all? Are we so abjectly irrational that we cannot recognize our confusion even when it is completely explicit? A different account is called for.

236

III
Resolving the Paradox by Jettison of the First Claim of the Governing View

In Wright [1975] and [1976] I came down fairly strongly in favor of the third approach: that of rejecting the claim that understanding the relevant expressions should be seen as consisting in the implicit knowledge of semantic rules. This response need not involve total banishment of the concept of semantic rule from any satisfactory account of what it is to understand such an expression. Indeed, it would be a serious objection if banishment were involved since *some* sort of notion of rule would appear to be prerequisite for the very idea that such expressions have *meaning,* which we wish, presumably, to retain. So I am now inclined to acknowledge little real contrast between the third approach and the second — that of retaining the first part of the governing view while rejecting the patchwork epistemology of semantic rules mooted in its second part. What I had in mind in my earlier proposal was that we should, in any case, repudiate the idea that knowledge of the character of the rules governing Sorites-affected expressions is to be construed as *propositional* knowledge, acquired by immersion, and largely implicit. That is the idea which sets up the project of rendering the purported content of this knowledge explicit, and which makes the second ingredient seem so natural and harmless. If we insist that, in these cases at least, there are *no* 'contents' implicitly known — so, in only that sense, no rules — there is no space for an argument that the rules are incoherent. However, if it continues to be appropriate to think of correct use of the relevant expressions as subject to rule, in whatever appropriate sense, then it ought also to be harmless to continue to think of competent speakers as *knowing* the rules in question. Something like the letter of the first part of the governing view may accordingly yet be sustained. But, if we follow my former proposal, the knowledge which competent speakers have of the rules relating to the relevant expressions will not be implicit, propositional knowledge; it will be knowledge which — like the knowledge *how* which constitutes practical skills like ice-skating and balancing — calls for no content as its object.

This proposed response to the Sorites may still seem frustratingly indefinite. If it can be worked into a satisfactory treatment of the problem, two matters, in particular, need extensive clarification. We need an account of the restricted class of cases, if any, in which it remains appropriate to think of understanding as constituted by implicit knowledge of the contents

237

of rules. And we need a much fuller account of what positive conception of understanding should supplant the implicit knowledge conception in the problematic cases. If understanding 'red', for instance, is to be thought of on the model of a practical skill, comparable to the ability to hit a good crosscourt backhand or ride a bicycle, we need both a sharp theoretical account of the distinction between those abilities in whose explanation it is legitimate and needful to invoke the idea of the agent's implicit propositional knowledge and the abilities, like those two examples, where we feel it is not; and also, ideally, independent reason to regard our understanding of Sorites-affected expressions as constituted by abilities of the latter sort. Pending clarification of these matters, my earlier proposal offers, at most, a strategy for resolving the Sorites paradox, and a pointer only to the general tenor of the lesson which it has to teach us.

Christopher Peacocke [1981] is unsanguine about the strategy, believing that it can be shown in advance that certain versions of the paradox will survive even if we abandon the conception that understanding is informed by (implicit) propositional knowledge of semantic rules. Peacocke invites consideration of a predicate, C, which is to apply to an object just in case the community will agree in calling that object 'red'. Suppose that some small difference, d, in the wavelength of light is not visually discriminable by any member of the community. And let light of wavelength k be definitely red. Then, according to Peacocke, we still have this paradox:

> If a reflects light of wavelength k, then C(a).
>
> If an object differs in the wavelength of light when it reflects by just d from something that is C, it too is C.
>
> ∴. All visible objects (reflecting pure light) are C.

Hence, "the paradox seems to arise even if we do not suppose that the use of these expressions is governed by rules."[9]

This may seem a rather puzzling thought. The example ought to be one, it seems, where a paradox arises for a predicate understanding which does not involve propositional knowledge of a rule. So how can a predicate *expressly defined* for the purpose, understanding which is thereby explicitly associated with knowing the content of a semantic rule, fit the bill?

Well, there are two different points to whose service Peacocke's example might be put. First, if C really is Sorites-susceptible, there seems, for just the reason given — viz., that it is introduced by explicit statement of a rule — simply to be no space for the response recommended in my earlier papers. How can we possibly avoid bringing in the idea of propositional

238

knowledge of the content of a rule if pressed to give an account of what it is to understand C? What Peacocke should claim, it seems, is not that the paradox arises in certain cases even after we have made the recommended response but that it arises in cases where the recommended response is not a possible option. But second, and more interesting, there is, in any case, reason to think that C is Sorites-susceptible only if there is reason to think that communal consensus with 'red' will be stable under indiscriminably small variations in light wavelength. Now of course, *one* reason for that thought would be if one supposed that 'red' was governed by incoherent rules and that the community would stick by them. But if that were the only reason, dissolving the 'red' paradox in the way I recommended would dissolve the C-paradox as well. So that cannot be Peacocke's reason. Rather, his thought is best interpreted as follows. Think, if you will, of competence with 'red' as a practical skill, uninformed by propositional knowledge. The operation of the skill involves differential sensitivity to varying visual stimuli, but if — as is so — the stimuli permit variation too slight to be detected by our visual apparatus, the skill cannot involve discrimination exercised over differences of that order of magnitude. So if there is communal consensus that a patch of color is red, and if all that the participants in the consensus are responding to is the visual stimulus, it may seem quite unintelligible how the response can vary if the situation is changed only by altering the stimulus by an amount insufficient to be picked up by our visual systems.

That this way of looking at the matter really does prescind from the propositional knowledge conception is testified to by the fact that it now seems incidental to the example that it concerns subjects who are capable of grasping contents at all. Think instead of a digital tachometer on a car. All such a device goes on, all it is designed to respond to, are variations in electronic impulses. There will be limitations to its sensitivity — so sufficiently slight variations in the incoming impulse will not, presumably, provoke any variation in the reading. And now it is neither more nor less plausible than before to conclude that, provided we are careful at each stage not to vary the impulse too greatly, the reading will *never* vary over a series of steps, no matter how many.

It would, of course, be quite unphilosophical to take comfort in the thought that actual tachometers do not behave this way — it is only *because* they do not so behave that we have the appearance of paradox. So Peacocke should be granted that some apparent paradoxes of the Sorites family are not amenable to the response I suggested. That, of course, completely

239

disposes of the value of that suggestion only if we assume that the entire family must admit of a uniform solution. What, I suggest, Peacocke's example teaches us is that that is not so. In fact, when Sorites paradoxes are individuated by reference to the type of ground adduced for their major premises, they fall, I believe, into at least three separate groups. One group involves support from the governing view; the C-paradox and the Tachometer paradox exemplify another group; and we shall be concerned with the third in section VI. (The groups are not, of course, individuated by the expressions which they concern: there is no reason a priori why one and the same expression should not be argued to be Sorites-susceptible by reasoning of each of the three kinds.)

If it would be unphilosophical to take comfort in the fact, it remains that actual functioning tachometers cannot be stabilized through any number of uni-directional but sufficiently marginal changes in the incoming impulse: sooner or later they jump — and a good one will do so often enough, no matter how marginal the changes, to continue to serve as a practically reliable instrument. Therefore the premises of the Tachometer paradox are not true of actual tachometers; and the paradox must be resolved by explaining how exactly that is so. The obvious and presumably correct thought is nothing very exciting: the major premise, to the effect that if the instrument gives reading R in response to impulse i, it will also give R to any i' differing from i by no more than some specified amount, will turn out to be a misinterpretation of what is entailed by possession of a sensitivity threshold. Quite what the correct account will be will depend on the design of the particular instrument concerned. But one would expect it to be fairly typical that, within the continuous interval, $\langle i_1,...,i_n \rangle$, there will be a finite subset of points, $\{b_1,...,b_k\}$, such that in response to an impulse of any of these values, b_j, or lying within some margin, d, of b_j, the instrument will always respond with the reading f_j; while responses to other impulses with values in the interval, $\langle i_1,...,i_n \rangle$, will depend additionally upon the prior state of the instrument.[10] One consequence, of course, is that such an instrument will, on different occasions, give different readings in response to the same impulse, depending on how that impulse is 'approached'; but perhaps that is just what such instruments do.

The salient point is that it has to be consistent with the claims that a tachometer has a sensitivity threshold, and that its response is entirely to stimuli of a certain sort, to suppose that the major premise in the paradox is not everywhere true. Presumably, then, since communal consensus in the use of 'red' does not extend all the way down to orange objects, the

240

major premise in the C-paradox will be false on similar grounds. And the explanation of its falsity will be consistent with supposing that we all possess visual sensitivity thresholds and respond only to visual stimuli in our use of 'red'.

Peacocke is aware of the possibility of this kind of response to the C-paradox, and is dissatisfied with it. He writes

> '... any model for an application of vague observational predicates must provide analogues of three things involved in such application: there must be states which are the analogues of having experiences, there must be something analogous to the ... non-transitivity of non-discriminable difference, and there must be some analogue of the application of an observational predicate upon a particular occasion.'[11]

Peacocke believes that it is impossible to provide for all three conditions in the sort of model illustrated by the tachometer. Suppose we assimilate the state of the instrument, induced by the reception of a particular impulse, to the having of an experience; and the reading which it issues to the application of the predicate. What about the non-transitivity of indiscriminability? In the case illustrated, for instance, we can, by choosing a pair of values which are respectively within and just outside the d-margin of some b_j elicit — at least sometimes — differential responses from the instrument *no matter how small* the difference between the two chosen values may be. How then can it be claimed that sufficiently small differences are indiscriminable for the instrument, so that — since larger ones are not — the relation, '... is not discriminable by the instrument from ...', behaves non-transitively?

But Peacocke is wrong about this. Or rather: he is right to think it essential that some analogue of non-transitive indiscriminability should be a feature of the model but wrong to suppose that one cannot be provided. The impression to the contrary requires that the instrument's *sometimes* issuing different responses to a pair of stimuli should be regarded as sufficient for its being able to discriminate between them. And the error in that supposition is easily demonstrated if, reverting once again to a case in which human beings are the 'instruments', we consider Michael Dummett's example of the slowly moving pointer.[12] You observe a pointer which is initially at rest but then begins to move, too slowly however for you to see that it is moving. You are asked to give a signal, — to raise your right hand, say — as soon as it seems to you to occupy a different position to the initial position; and — let us add to Dummett's example

241

— to raise both hands as soon as it seems to you to have moved into another position again. Suppose that you raise one hand after four seconds, and both after eight. Are we to conclude that the position which the pointer seemed to you to occupy after the third second looked different to the position which it seemed to you to occupy after the fourth? The answer, of course, is that you cannot detect any difference between those positions — what you meant to indicate when you first raised one hand was only that the position now seemed to you to be different to the starting point.

As Dummett brings out, there are ways of describing the phenomenology of the situation which are threatened with incoherence — and perhaps none which are not. But it is legitimate to suppose, for the purposes of the analogy with the tachometer, that your permissible 'signals' are restricted to the raising and lowering of hands. Suppose, then, that the actual positions of the pointer after each second are determined by some appropriately finely calibrated instrument. The situation will be that you sometimes will and sometimes won't respond to the sorts of changes involved from one second to the next by varying your signal; but that *asked* about any single such change, you will report the positions involved seem to you to be the same. More: if we imagine that, once you have raised both hands, the movement of the pointer is reversed, we would expect to find that when it regained the position originally occupied after three seconds, you would still have one hand raised. If we — as is permissible — suppose further that your signals for the positions occupied initially, after four seconds, and after eight seconds, are by-and-large respectively uniform, then your performance is now in all relevant respects assimilable to that of the hypothetical tachometer. And, crucially, the fact that you always (more or less) have a hand raised when the four-second position is presented, but only sometimes have a hand raised when the three-second position is presented is *no indication that those positions seem different to you*. On the contrary, if they did seem different, you would *always* give a different signal. So far, then, from its being a sufficient reason for regarding the tachometer as able to discriminate between a pair of impulses that it some-times issues different readings in response to them, the fact that it does not *always* do so should be regarded as decisive for regarding the relation in which they stand to it as an analogue of *indiscriminability*. And the modelling of non-transitivity is then provided by the reflection that among three distinct impulses, the tachometer may *only sometimes* respond dif-ferentially with respect to the first and second, and *only sometimes* respond differentially with respect to the second and third, but *always* respond

242

differentially with respect to the first and third.[13]

I conclude that this class of Sorites paradoxes need not detain us. Their major premises are false, and false in a way which classical two-valued semantics is quite adequate to describe. The details of why they are false will, of course, depend upon the character of the particular 'instruments' concerned; and I am not seriously suggesting that the tachometer provides a satisfactory model of all aspects of human competence with color vocabulary. My point is only that, once the first claim of the governing view is abandoned, the residual case which Peacocke thinks he sees for supposing C to be Sorites-susceptible is perfectly matched by the Tachometer paradox; and the model solution sketched for the latter depends on no feature of the tachometer without an analogue in the 'instrument' constituted by a human being responding to color. Accordingly, while different things might need to be said about why precisely the major premise in the C-paradox is false, we can at least understand why we shall not be committed to its truth simply by supposing that our descriptions of color are nothing but responses to visual stimuli among which our discriminations are not everywhere transitive.

IV

Tolerance and Observationality

Peacocke concludes his objection to the tachometer analogy with the complaint that, if it were apt,

'... there would be no difficulty in the idea of an observational predicate ... definitely applying to one but not to the other of two objects which produce experiences which are not discriminably different in quality. But in fact we can make no sense of this idea.'[14]

Undoubtedly, we can make no sense of it if observational predicates are governed by the *rule* that they apply to both, if to either, of any pair of observationally indistinguishable items. But that, I want to suggest, is just the misunderstanding of what an observational predicate is which the governing view encourages. The question is therefore, what should we replace it with?

Of course, the question whether *any* vocabulary is worth distinguishing as "observational" has been intensively debated by twentieth century philosophy of science; and the received wisdom has been in favor of a negative answer. Still, it is natural to feel, there *is* a class of predicates

243

— those expressing colors, places, musical pitch, or perhaps any Lockean secondary quality — for which Peacocke's thought has prima facie force. Whether we style these predicates as observational or not, how exactly are they equipped to effect the sort of distinction, among qualitatively indistinguishable items, whose intelligibility Peacocke is questioning? We shall get an inkling about how an answer to this question might go by reflecting on what may seem a fairly obvious lacuna in the response to the Sorites which I have been canvassing. The response claims, in essentials, that the *motivation* for the major premises in (a large class of) Sorites paradoxes is misguided; that the pressure to assent to the major premises is generated only by the mistaken assumptions of the governing view. But can it suffice for a solution just to undercut the *motivation* for the major premise in, e.g., the Sorites paradox for 'red'. Do we not need, in addition, to disclose some *specific* fault with the major premise? If there is really no paradox, then that conditional *cannot be true*; and we cannot rest content until we have explained exactly why not, and, if necessary, have devised a semantic apparatus in terms of which the manner of its untruth can be properly described.

We are now in a position to see how at least the outline of such an account might run. Let us essay to see ourselves, in accordance with rejection (or, at least, severe qualification) of the first claim of the governing view, on the model of *signalling instruments* for qualities like red — albeit voluntary signallers. Then there ought to be the same kind of connection between an item's being red and the description to which we are prepared to assent, when appropriately placed and functioning normally in normal circumstances, as obtains between, say, an engine's operating at 4500 rpm and the reading issued by a number of correctly connected tachometers which are functioning properly and are free from external interference. This connection is actually such as to generate a *truth condition*: for a particular reading to be correct *is* for it to be the case that the mean value of the readings issued by sufficiently many appropriate instruments, which are appropriately situated, functioning properly, and free from external interference, will coincide with that reading. So, too, for red, or any quality of the kind which Peacocke's quoted thought seems plausible. For an object to be (definitely) red *is* for it to be the case that the opinion of each of a sufficient number of competent and attentive subjects who are appropriately situated to command a clear perception of the object, functioning normally, and free from interfering background beliefs — for instance, some doubt about their situational competence — would be that it was red.

244

What follows? Well, one thing we know — and, having disposed of the Tachometer paradox, can bring to bear with a clear conscience — is that if a Sorites-series of indistinguishable color patches is run past a group of such subjects under the specified conditions, and they are agreed about the redness of the initial elements, there will be, as the series moves towards orange, a specific *first* patch over which the consensus fractures. The occurrence of the fracture need not raise any doubt about the continuing competence of the subjects or about the satisfaction of the specified conditions at that point; nor need any of the subjects be able to re-identify that patch if it is subsequently re-presented with its immediate predecessor in a different setting. So the upshot is simple. If observational qualities, whatever should be their proper characterization, essentially sustain the sort of equation which the tachometer analogy suggests — if their application is a function of what competent observers, under suitable conditions, etc., will assent to — then a predication of red, or any observational quality, fails of definite truth as soon as the right hand side of the equation fails of truth. We have therefore to acknowledge, surprising as it may seem, that a Sorites series of indistinguishable color patches *can* contain a last patch which is definitely red: it will be a patch about whose redness there is a consensus meeting the described conditions, and its immediate successor will be a patch about which the consensus breaks down in a way that still observes those conditions.

It may be difficult or impossible to *identify* such a patch in a particular case, either because it is difficult to verify that consensus has broken down in a way which still observes the conditions, or because it is known that one or more of the conditions is unsatisfied. But it remains that there can be such a patch; that it *can* be identified under favorable circumstances; and that, even in circumstances which are not favorable, when the consensus has broken down for irrelevant reasons, it is legitimate to suppose that there is such a patch which could have been identified had the conditions been different. No doubt, it would usually be of absolutely no interest to effect such an identification. And it should be noted that there is, a priori, no reason to suppose that 'the last definitely red patch' would turn out to have a stable reference; if it did not, that would disclose an element of context-relativity in the concept of red which we normally do not suspect.

Provided, then, that observational qualities indeed sustain the relevant sort of equation, we can, more than undercut the motivation for believing them to be true, explain how specifically the major premises in Sorites

245

paradoxes for observational predicates fail of truth. (The question remains how such explanations could proceed for other kinds of Sorites — I shall return to it below.) We can also make sense of the idea of an observational predicate definitely applying only to one of a pair of phenomenally indiscriminable objects, which Peacocke claimed was unintelligible. What *is* unintelligible is how someone might tell just by inspection of a pair of indiscriminable items that they in fact mark the watershed, in a particular context, between what is definitely red and what is not. But Peacocke's claim follows from that only if we suppose that any observational predicate worth the title must be such that this distinction can be drawn purely observationally. And that is the mistake: the raising of an arm, in the example of the slowly moving pointer, is a gesture made purely in response to how things look — it is explicitly instructed that this is how it should be made — and so may be regarded as analogous to the application of an expression which is observational in the sense which Peacocke intends. The fact remains that if we suppose that having a single arm in the air is *definitely* an appropropriate response only in situations when — supposing the satisfaction of all the relevant conditions — the subject *always* has a single arm raised, the observationality of the response is consistent with the existence of a sharp boundary to the cases where it is definitely appropriate.

On the account of observationality which figures in Dummett's [1975] and in my earlier papers, a predicate is observational if whether it applies to an object can invariably be determined just by (unaided) observation of that object. It seems to follow immediately that any such predicate must be applicable to both, if to either, of any pair of objects which (unaided) observation cannot tell apart; so that the account renders all observational predicates tolerant with respect to indiscriminable change and thus takes us straight over the precipice. But there is an unclarity here: is it intended that the distinction between definite cases of application and *all others* be observationally decidable,[15] or merely that between definite cases of application and those where the predicate definitely *fails* to apply? A little reflection suggests that what we really want is the second, weaker version. The observationality of 'red', for instance, should entail only that something is red only if any suitable subject may, in appropriate circumstances, observe it to be so; and that something is not red only if any such subject may, likewise, in appropriate circumstances, observe it not to be red. A predicate's being observational in this sense need not entail that a distinction can always be drawn, just by observation of the objects concerned, between

246

cases to which it definitely applies and cases of any other sort. Indeed there is something absurd about the idea that *that* distinction might be observationally decidable, since borderline cases of 'red', for instance, are exactly cases about whose classification observation, even in appropriate circumstances, leaves at least some suitable subjects essentially *unsure*.

Let me pause to review the two principal suggestions of this section so far. The first is that it is not, contrary to Peacocke's claim, unintelligible how anything worth regarding as an observational predicate might definitely apply only to one of a pair of observationally indistinguishable items; on the contrary, the possibility is implicit in the consideration (to which we led by consideration of the analogy with the tachometer but which stands independently) that it is a necessary condition for an observational predicate to apply to a particular item that any suitable subject who observes the item under appropriate circumstances will judge that it does so. The second is that it is an error to think that Peacocke's idea follows if we agree that anything worth regarding as an observational predicate must be such that whether or not it applies to an object can be determined, in appropriate circumstances, etc., just by observation of it. What that commits us to is only the second, weaker principle sketched in the preceding paragraph. To be sure, the governing view arguably provides grounds for thinking that 'red' is observational in a sense which subserves the stronger, paradoxical principle. We shall review those grounds in section VII.

There is much more that needs to be said about the notion of an observational predicate. Here though, I can only briefly canvass two further suggestions. First, on the relation between the two thoughts just summarized. There is no objection, one would imagine, to interpreting the reference to 'appropriate circumstances' in the weaker principle along exactly the lines featured in the biconditional for qualities like red to which consideration of the tachometer analogy led us. 'Appropriate circumstances' are those where the subject is appropriately placed to observe the object concerned; 'suitable subjects' are competent, attentive, perceptually normal subjects who are free of interfering background beliefs. So the weaker principle, so interpreted, has just the effect in its own right which the biconditional had: reason to doubt that competent, etc., subjects would, in appropriate circumstances, judge an object to be red — reason afforded, perhaps, by such subjects actually failing to do so — is reason to doubt that the object should be classified as red. Thus the weaker principle, so far from motivating acceptance of the major premises for Sorites paradoxes involving observational predicates, actually subserves the form of rejection

247

of such premises suggested by the analogy with the tachometer.

Note, therefore, the perspective in which the stronger principle is now placed. The stronger principle differs from the weaker only in its second clause. But the effect of the weaker principle just described is wholly owing to its first clause. So the stronger principle likewise has that effect: to endorse the principle is to be committed to treating a lack of consensus among competent, etc., subjects in appropriate circumstances as a sufficient reason for doubting that something is definitely red, even if the object in question is indistinguishable from one about which a consensus obtains. If therefore, as it seems, the stronger principle also entails that 'definitely red' tolerates indiscriminable difference, the conclusion is forced that the ingredient clauses of the stronger principle are actually incoherent: collectively they entail that a difference must always be apparent to observation between anything which is definitely red and anything of any other sort; but the first clause also entails that different responses by competent, etc., subjects may suffice for such a difference even if the items in question are observationally indistinguishable. This is a quite different conclusion to anything that might be inferred from the role of the stronger principle in the generation of one kind of Sorites paradox for 'red'. For one thing, that paradox depends, in addition, on assumption of the non-transitivity of indiscriminability; for another, the kind of semantic incoherence which the paradox has been thought — by Dummett, for instance — to disclose, must not be confused with the kind of incoherence just bruited. It is not inconsistent to suppose that there might be semantically incoherent predicates; but no predicate can be observational in a sense which subserves the stronger principle — at least if the foregoing is correct — since to be so would be to meet an inconsistent specification.[16]

Finally, let me sketch a possible connection with a proposal for the characterization of observational expressions which I have entertained elsewhere.[17] Observational expressions are those which connect most intimately and immediately with our experience. The account of observationality given by Dummett, (and Peacocke's purported refinement — see note 16) both locate this intimate, immediate connection in the *conditions of application* of the expression. Observational expressions are to apply, or not, in virtue of how things appear. But an alternative would be to seek an account by exploration, rather, of the *conditions of understanding*. To be sure, items qualified by an observational predicate ought, under appropriate circumstances, to present a distinctive appearance; and understanding the predicate ought to involve attaching a proper significance to

248

the appearance, or range of appearances, associated in this way with it. But there are some expressions — a good example is a natural kind term like 'banana' — which we would not wish to regard as observational because, e.g., they carry implications about constitution and origin, even though the things they qualify have distinctive enough appearances. And it is striking that somebody's use of an expression could be quite properly responsive to the kind of appearances associated with it and yet they might not understand it precisely because ignorant of such additional implications. A natural suggestion, then, is that observational expressions are marked off from this wider class precisely by the circumstance that, while anyone who understands such an expression must display an appropriate sensitivity to a certain distinctive range of appearances with which its application is standardly associated, there must in addition be no space for the possibility of their doing so yet failing to understand the expression. 'Gene', 'torque converter', and 'pulsar' fail to qualify as observational because there is no particular kind of appearance possessed by the things to which they apply; 'banana' and 'human being' fail to qualify because someone who had no grasp of the kinds of thing which bananas and human beings respectively are might still be quite clear about the way they tend to look, and might show as much in their employment of 'banana' and 'human being'.

This indicates, at most, merely the rough shape which a satisfactory account of observationality might assume.[18] What is salient is that any expression which conforms to it is presumably going to sustain something very close to the clause which is common to the weaker and stronger principles which we have been discussing. If 'φ' is such an expression, for instance, then it follows from the sketched account that anything which it qualifies will present, under appropriate circumstances, a distinctive kind of appearance to which anyone suitable who understands 'φ' will be appropriately responsive; i.e., they will be prepared to judge that the item is so qualified provided other appropriate conditions are met.

Natural kind concepts, like 'banana' and 'human being', will likewise sustain such clauses, provided it really is essential that their instances have a certain sort of appearance. (If it is not, then they should be classified with 'pulsar' and 'torque converter' after all.) The question arises, therefore, whether the addition of a stipulation that such sensitivity is to suffice for understanding 'φ', though not 'banana',[19] somehow imports commitment to the tolerance of 'φ' with respect to indiscriminable change. It is clear that the stipulation preempts the kind of explanation which can be

249

given, in the case of natural kind predicates, of how they may distinguish between observationally indiscriminable items. An artefact — a 'fool's banana', for instance — may be indistinguishable from an instance of natural kind; understanding 'banana' involves grasping this possibility. Needless to say, preempting one kind of explanation need not be to foreclose on all, but the question is still outstanding: how might the corresponding explanation go in the case of predicates which conform to the proposed account? How could the 'appropriate kind of responsiveness' operate selectively among indiscriminabilia? If I am right that the kind of account of observationality mooted will entail that observational expressions comply with the first clause of the stronger (and weaker) principle, then what is asked for is, in effect, an explanation of why the mooted account does not entail their compliance with the second clause of the stronger principle — the clause whose inclusion makes it impossibly strong. I shall revert to the matter in section VII.

V
Peacocke's Positive Proposals — Degree Theoretic Approaches

Vague expressions tend to be associated with a comparative: red things can be more or less red. It is natural to suppose that if something can be more or less φ, then that gives sense to the idea that the simple positive predication '... is φ', can be more or less true of it. Peacocke's idea is that once we take degrees of truth seriously, we can diagnose the Sorites paradox as arising from the attribution of contradictory properties to the conditional in the major premise. Specifically, the paradox will result from assuming simultaneously that Modus Ponens inferences are valid without restriction for the conditional and that it suffices for the truth of a conditional that its antecedent and consequent enjoy degrees of truth which, if not identical, are at least extremely close. In order to apply this proposal to the case of color predicates, and other observational expressions, we shall require, of course, that the degrees of truth of 'a is red' and 'b is red', for instance, may be different even though a and b are indiscriminable in point of color. Peacocke is thus committed to utilizing a notion akin to Goodman's idea of a *quale*.[20]

Appealing as it may seem, Peacocke's treatment is open to a number of objections which are, I think, collectively decisive against any similar degree-theoretic account. More specifically, it is only the third and fifth of the six objections which follow which, so far as I can see, relate to

250

features of Peacocke's paper which connect inessentially with his overall approach.

(i) **Can a degree-theoretic account explain the plausibility of the major premises?** There is no difficulty, of course, with the usual, quantified conditional form of premise. The explanation will claim that each instance, Fa $\rightarrow$ Fa', of ($\forall$ x) (Fx $\rightarrow$ Fx') is *almost* true: that its consequent enjoys a degree of truth ever so nearly but not quite as great as that of its antecedent. And this claim will then be followed, presumably, by a stipulation that the degree of truth of any universally quantified statement is the minimum of the degrees of truth enjoyed by its instances.

There could be some debate about this stipulation for the universal quantifier — why should it not be, for example, that ($\forall$ x) (Fx $\rightarrow$ Fx') takes on the *sum* of the errors of its instances instead? But let us now forego that debate. More important is the reflection that the major premise doesn't need to be conditional at all. In the case of the Sorites-series of indiscriminable color patches for instance, we could just as well take it in the form

$$(\forall x) - [\text{red}(x) \ \& \ -\text{red}(x')].$$

All the ways of making the conditional form of major premise seem intuitively plausible would be applicable to this conjunctive form.[21] Now, we have already seen, in effect, how *if* we assume that each instance of the conjunctive major premise is merely almost true, it follows quite naturally that successive inferences from red(x) to red(x'), or from $-$red(x') to $-$red(x) will involve marginally diminishing degrees of truth.[22] But Peacocke needs to explain, in addition, with what right such a conjunctive major premise may be regarded as *almost true*; otherwise he cannot explain its plausibility, or duly acknowledge the force of the arguments which seem to sustain it. However the attempt to extend his account in the most natural way, fails to produce the right results. Plainly we need degree-theoretic matrices for negation and conjunction of such a sort that $-$(P & $-$Q) turns out to be almost true whenever the truth value of Q is almost as high as that of P. Assuming that the matrix for negation yields that $-$A is almost true just in case A is almost false, it follows that P & $-$Q has to be almost false in that same case. Classically, a conjunction is false just in case at least one conjunct is false; so a conjunction which marginally fails to meet the minimum standard for falsity ought to be one in which at least one of the conjuncts is almost false while the degree of falsity of the other is no greater. The result — if I may be forgiven for spelling it

251

out in detail — is that P & $-$Q cannot be almost false if P has a middling sort of degree of truth and Q, as hypothesized, a marginally smaller degree. For in that case P is not almost false and neither, by the sort of matrix for negation envisaged, is $-$Q. Accordingly, although Peacocke displays the materials sufficient to explain how, *if* it is agreed that the instances of a quantified conjunctive premise are all almost true, reiterated applications of modus ponendo tollens will lead to successive conclusions of a deteriorating degree of truth, he has provided absolutely no explanation of why the close proximity in degree of truth of successive Fx and Fx$'$ should make for the almost-truth of a quantified conjunctive major premise, even granted the stipulation for the universal quantifier noted above. And it is unclear what direction a degree-theoretic account might take which would fare better on the point.

(ii) **Is the notion of difference of degree which the account demands fully intelligible?** Since it is essential to Peacocke's purposes that the degrees of redness of a pair of indiscriminable color patches can be different, an account of identity and distinction among degrees is called for which sustains that possibility and, indeed, enables us to regard the degree of redness as diminishing step by step as we move along the relevant sort of Sorites series. Peacocke's offering, inspired by Goodman, is

(a) x and y are red to the same degree if and only if any color matching the color of x matches the color of y and vice versa;

and

(b) The degree to which x is red is greater than the degree to which y is red if and only if some color matching the color of x is redder than the color of y.[23]

One evident problem with clause (b) is that it will deliver what is wanted only if it is true to say of some pairs of color patches which are *just* discriminable, i.e., form the first and third members of a non-transitively matching trio, that one is redder than the other. I do not know why Peacocke thinks that this is so. It seems certain that, with shades of color as close as that, no-one would feel able to say that either in particular was redder than the other when presented with them in isolation; and presentation in the context of a Sorites series in which one appeared later than the other would be of no help unless one was assured that the series was moving *uniformly* away from red (which, in view of the variability of color in several parameters saturation, intensity, hue need not be so.) But that

252

assurance is precisely the assurance that of each successive just-discriminable pair of patches, the first is redder than the second — and it is the *content* of that claim in such cases that is at issue.

Peacocke offers nothing by way of an account of the comparative, 'is redder than' utilized in clause (b). He is, however, making a substantial assumption about it. Let us say a relation R is *counter-transitive* with respect to a domain of entities just in case whenever Rxy holds good for x and y in that domain, x also bears the ancestral of the contrary of R to y, the intermediate links in the ancestral chain being supplied from the domain in question. Thus, with respect to a suitable domain of color patches, the relation of discriminability is counter-transitive, since any pair of discriminable color patches can be linked by a chain of pairwise indiscriminable ones. Peacocke's assumption, then, is that the relation 'is redder than' is *not* counter-transitive with respect to any domain of color patches any pair of which are at least just-discriminable. I cannot see that our intuitive conception of the relation validates this assumption; it seems evident that we make no use of the comparative except in cases involving differences much greater than just-discriminability. In any case, Peacocke needs a demonstration that that is not so, which would presumably have to draw on the analysis of this and similar comparatives which he does not give. Unless the lacuna can be filled, the degree-theoretic treatment simply cannot proceed as far as Sorites paradoxes involving indiscriminability are concerned.

(iii) **Can the notion of a color which features in Peacocke's clauses (a) and (b) be satisfactorily accounted for?** It seems doubtful whether the idea can reasonably be taken as primitive in the present context. But we strike trouble pretty speedily as soon as we attempt to explain what sort of things the colors are which constitute the range of the quantifiers in Peacocke's two clauses. A natural proposal would be to treat colors as Goodmanian *qualia* so that

Color c_1 is the same as c_2 just in case any color matching c_1 matches c_2 and vice versa.

But this has the drawback that it quantifies over colors in attempting to explain the criterion of identity distinctive of the sort. That, admittedly, need not vitiate an explanation whose purpose is simply to characterize a feature of a concept rather than to introduce it. Indeed it is arguable that even certain introductory explanations can afford this kind of apparent circularity.[24] But it is doubtful whether anything mitigates the charge of

253

circularity in the present case. We *do* need an introductory explanation of qualia in general and of colors, so understood, in particular. And introductory explanations which involve quantification over instances of the concept we are trying to explain can get away with that only when it is possible for a trainee, so far innocent of the concept, to understand statements involving such quantifications. If a case can be made for saying that that is so, it will be because the quantification is involved in the formulation of conditions which can be recognized to be true or false of a domain of objects quite independently of full knowledge of the sortal distinctions exemplified in that domain. (A trivial example would be $(\forall x)(x = x)$. The proposed account of color identity, and indeed Goodman's account of identity for qualia generally, do not come into that category.

The obvious response is to take the matching relation as obtaining between colored objects rather than colors themselves and modify the account to read something like

> The color of x is the same as the color of y just in case it is not possible that there should be a z which matches one but not the other,

where x, y and z are all variables for individuals. But of course it is *always* possible that there should be such a z — if x and y are allowed to shift their color! If we then stipulate that the colors are to remain the same across the relevant range of possible circumstances, we once again use the notion we are trying to explain. And if we leave the modal operator out, the precise color of an object will turn out to be relative to what other objects happen to exist: objects which are currently the same in color may suddenly become distinct in virtue of the creation of some third object which sets up non-transitive matching, and may once again come to coincide in color if that object is destroyed — all this without themselves undergoing any change.[25]

(iv) **Can degree-theoretic approaches fulfill their intuitive promise?** Let us be clearer what makes such approaches seem promising. The motivation is not merely the thought that if the relevant expressions admit degrees of applicability, the major premises can fail of strict truth. It goes deeper. Consider the series of conditionals that make up the intervening links in a Sorites. The dilemma, basically, is that if we say that they are all true, we are bereft of the means to explain how the transition from, e.g., redness to non-redness is effected; and if we say that one of them is false — or, anyway, other than true — we secure an explanation only at

254

the cost of imputing a determinate sharp boundary when none seems to exist. The preceding section tried to explain how it might be that our scruples about the second horn of this dilemma are ill-founded in an important class of cases. But, whether in that way or another, the apparent problem is to explain *how* the change from redness to non-redness is effected across these conditionals. Redness disappears as we work our way along the series — if there is not a first point at which the patch has ceased to be definitely red, how is the trick pulled? What we would most *like* to be able to say is exactly that the change takes place *at no particular point*. And if that idea can be saved from incoherence at all, something close to Peacocke's suggestion seems to be the only way of doing so: *each* of the conditionals contributes towards the change, but contributes too small an amount to make it worth saying, in any practical context, that its consequent is less acceptable than its antecedent. It then needs to be recognized merely that this notion of the acceptability of a conditional will not sustain indefinite reiteration of modus ponens, and the paradox is solved.

The real attraction, then, of the degree-theoretic approach is that it promises simultaneously to explain how the transition from Fs to non-Fs is effected in a Sorites-series and so to block the paradox, to explain the plausibility of the premises which generate it, yet to avoid postulation of any determinate boundary of the sort that is apt to strike us as fictitious. So it is important to realize that, when matters are inspected closely, this promise cannot be redeemed by Peacocke's account, (nor, so far as I can see, by any degree-theoretic approach) irrespective of whatever other objections there may be to it. A major difficulty emerges when we reflect that there is intuitively no *pinnacle*, as we might put it, to the scale of degrees to which something can be red or exemplify any other vague concept of degree. There is, e.g., no more-or-less-red among crimson, scarlet, and vermilion.[26] In general, any vague concept F admits of quite a wide variety of discernible cases all of which are *definitely* and *absolutely* F. Thus although scarlet is more like orange than vermilion is, it is not less red: both are paradigms of what is absolutely red. The result is that if we start the red-orange Sorites-series of matching color patches with a vermilion one, we have to say that the conditionals which correspond to the early transitions in the series involve *no diminution of degree of redness at all*. And now we can scarcely forebear to ask how the transition is effected from conditionals of this kind to conditionals where such a diminution *is* involved. If it takes place at some specific nth conditional, that is as much as to say that there is a first patch whose degree of redness is

255

not absolute; and if we are willing to admit that, it is not clear why we should scruple over the idea that there is a first patch which is not definitely red — and so rescind the consideration which gave the degree-theoretic approach its special appeal.

Again, what we would *like* to be able to say is that the transition from the one kind of conditional to the other takes place at no particular point but still takes place. The obvious ploy would be to attempt to activate the degree-theoretic account at a higher level, considering, e.g., the series of conditionals of the form:

If C_n is absolutely acceptable, then C_{n+1} is absolutely acceptable, where C_k is the kth conditional called on in the Sorites paradox for 'red', and such a conditional is absolutely acceptable just in case the degree of truth of its consequent is no smaller than the degree of truth of its antecedent. But the trouble with this is two-fold. First, the notion of a degree of *absolute* acceptability, in contrast with a degree of redness, seems to make no sense — it is like Orwell's joke about political equality. Second, even if it can be made to make sense, the problem recurs; for sufficiently small values of n, such conditionals will themselves be absolutely acceptable, so if those taking larger values are not, when and how is the transition between the two kinds of cases effected? I suggest the conclusion not merely that Peacocke's account fails to solve the original dilemma in the attractive-seeming way promised, but that there is no solution to the dilemma of that sort.

(v) **C is not a predicate of degree.** Peacocke's predicate, C, you will recall, applies to an object just in case the community will agree in calling it red. Consensus, of course, admits of degrees; but as the notion figures in Peacocke's definition of C, it is intended to be complete and absolute. But it was, in Peacocke's view, the Sorites paradox for C which showed the error of my [1975] and [1976] diagnosis and called for a different account. Yet Peacocke grants that C is not a predicate of degree, and indeed allows that it is precise, determinately either applying or failing to apply to any particular object. Ought he not, then, to admit that his own proposal can cope no better with the C-paradox than he believes mine did?

His response is perplexing. He writes

'... [the degree-theoretic approach] covers the metalinguistic paradox indirectly. Sometimes we have to state a theory about the extension of a sharp predicate by using a vague predicate. In particular is this

256

true of the sharp predicate C and the vague predicate 'red'. We have to say that:

> Any object which is red the community will agree in calling 'red';

> any object which is not red the community will not agree in calling 'red'.

I suggest that the only reason that we feel tempted to accept the major premise of the metalinguistic paradox is that we employ reasoning using conditions with properties (1) and (2) —

[viz., unrestricted iterativity of modus ponens, and 'if Fa, then Fb' true whenever F is observational and a and b are indistinguishable] —

> and which contain 'red', and then go on to apply these two general principles to draw conclusions which contain the predicate C.'[27]

Otherwise, Peacocke claims, there is no reason to believe the major premise of the C-paradox. Rather, since C is sharp

> '... some instance of this premise must be false if the principles displayed earlier — [in the immediately preceding quotation] — are true. Since C itself is not an observational predicate, there is no pressure against this conclusion ...'[28]

What is happening here? The thought, I take it, is that in the presence of the two cited principles determining the extension of C, we can negotiate the transition, for arbitrary indistinguishable a and b, from Ca to Cb provided we take it that the indistinguishability of a and b verifies 'if a is red, b is red' and this conditional sustains the modus ponens inference. The point is immediate if we assume that Peacocke intends the two principles cited to have the effect that 'red' and C have the same extension — but that he cites two principles, rather than just the biconditional of the first, suggests that he may intend, rather, a relation like that between the extension of 'red' and the extension of an admissible sharpening of it (in the sense of Fine's [1975]).[29] Even so, the sharpness of C, and the consequent validity of double negation elimination, affords us $Cx \rightarrow --red(x)$ by contraposition on the second principle, and $--red(x) \rightarrow Cx$ by double contraposition on the first. Since we have $--red(a) \rightarrow --red(b)$ by double contraposition on the conditional for 'red', the reasoning thus goes through.

Peacocke's claim, then, is presumably that, in this precise sense, the major premise for the C-paradox stands or falls with that for the 'red'

257

paradox. But, of course, if that is true, the C-paradox cannot be a *special* problem for anyone! In particular, it cannot be a special problem for the view that the major premise for 'red' derives a specious plausibility from the governing view. By attempting to treat the C-paradox in this way, Peacocke undercuts his case for thinking that a different approach to the Sorites was called for in the first place.

The wood gets further tangled when we ask whether Peacocke himself accepts the two principles (hedged about, as he presumably intends them to be understood, by appropriate constraints on the normality and competence of the members of the community and the appropriateness of the conditions of observation). The argument of section IV would have it that, so hedged about, their acceptability is of a piece with the most promising account of the sense in which 'red' is observational.[30] But Peacocke cannot accept them, at least as stated. To do so would be inconsistent with his whole approach. Consider the situation at the cut-off point for C which Peacocke allows to exist. Let a be the last C element, and b the first non-C one. Then contraposing on the first principle, b is not red. Whereas, contraposing on the second and eliminating the double negation on $--Ca$ we have that a is not-not-red. So there is after all a sharp boundary which we could describe in terms of 'red': if we run the paradox from right-to-left, as it were, and take 'not-red' as its subject, then there is a determinate last case to which the predicate applies and a determinate first case to which it does not. But 'not-red' is, in this context, as much an observational predicate as 'red' is; and the entire *motif* of Peacocke's approach was to find a way of avoiding the paradox while acknowledging what he considered essential, that observational expressions lack sharp boundaries. Needless to say, if Peacocke were to respond to this by rejecting the two principles, then he would either have to admit that the C-paradox is no paradox — if it has to depend on the two principles[31] — or allow that the degree-theoretic approach provides no means for coping with it — if it does not. Actually, I think that Peacocke's diagnosis of the plausibility of the major premise in the C-paradox is farfetched and that its appeal derives rather from the train of thought outlined in presenting the Tachometer paradox above. But that is a minor criticism at this stage.

(vi) **Can degree-theoretic approaches engage Sorites paradoxes which involve phenomenal predicates?** I argued in Wright [1975] and [1976] that there are certain examples of the Sorites paradox for which no degree-theoretic approach can be adequate.[32] To grant, for instance, that

258

'x is red' admits of indefinitely many degrees of *truth* is of no immediate help when it comes to addressing the paradox that arises for the *assertibility* of 'x is red'. Let us pursue the matter a little.

In section IV I tried to give cause to doubt that a proper account of the observationality of 'red' would sustain the principle, which leads directly to paradox, that of any pair of indistinguishable items, both must be red if either is. It is notable, in addition to the considerations advanced earlier, that any account with that consequence is going to have the effect that there is no difference between looking and being red; at least, it is going to have that effect unless it is believed that whatever factors can bring it about that the look and the reality come apart could not apply selectively to one of a pair of items which are indiscriminable in a single sensory presentation. But there is no reason to endorse such a belief. It certainly isn't *a priori*, for instance, that a pair of objects only one of which is red cannot both assume indistinguishably red appearances when bathed in red light of the right intensity and hue. But if 'red' is not observational in any such sense, that does nothing to mitigate the threat of a paradox for 'looks red'. It is not entirely clear whether a proponent of the degree-theoretic approach may legitimately assume that whenever 'is φ' admits of degrees of truth, so will 'looks φ'. Certainly, 'looks φ-er than' will then exist; but is that the right kind of comparative — if a looks φ-er than b, is 'looks φ' *truer* of a than of b? I don't know, but let's suppose it is. Even so, there is surely no sense in the idea that, when a and b are indistinguishable, one might nevertheless *look redder* than the other in the context in which they are indistinguishable. *If* 'looks red' is a predicate of degree, it is still utterly unclear how it might be true of indistinguishable objects to differing degrees. What settles the degree to which something looks red — if there is indeed any question of 'degree' — is the way it looks; so how could things that look absolutely similar look red to different degrees?

These simple reflections seem to dispose of all prospect of a degree-theoretic treatment of the Sorites paradoxes for the family of predicates which 'looks red' exemplifies. There is no discussion of the matter in Peacocke's paper. But there is a footnote at least on the paradox of warranted assertibility to which we are now rapidly led. For the thought is a compelling one that the conditions of warranted assertion of a statement of the form 'x is red' will coincide, ceteris paribus, with the *truth-conditions* of 'x looks red'. More specifically, if you have no reason to suspect the conditions of observation, or your own powers of observation, or your competence with 'red', then you are warranted in asserting that something

259

is red if it looks red; under the same provisos, it follows — granted the reflections of the preceding paragraph — that the *degree* of warranted assertibility cannot be different between two statements of that form concerning indistinguishable items. Accordingly, every φ apt for the formulation of statements whose assertion conditions can be characterized using an appropriate 'looks φ' — or 'sounds φ', etc., — is apt for the generation of Sorites paradoxes to which the degree-theoretic approach would seem to have no response.

The objection presented in my [1975] and [1976] was formulated slightly differently. I argued that the degree-theoretic approach could provide no satisfactory treatment of predicates of the type, 'x is such that its description as "red" is on balance justified'. Peacocke's response (in the footnote) — hardly calculated to inspire confidence after our reflections in subsection (v) — is that he would treat this predicate as he treated C if it is not a predicate of degree, otherwise as he treats 'red' itself.[33] Plainly it is no more a predicate of degree than is 'is heavy enough to tip the scales'. So Peacocke's response would be, in effect, to propose an extension-determining theory via two principles differing from those discussed on page 257 above only in that their consequents would be formulated in terms of this predicate rather than C, and then to suggest that there is no reason to accept the major premise of the relevant paradox except the defective one that goes via our acceptance of the major premise of the paradox for 'red'. But even if two such principles would have sufficient plausibility to subserve the suggestion, it should be clear that it is wide of the mark anyway: the reason for thinking that the major premise holds for 'x is such that its description as "red" is on balance justified' is exactly the role of *looking red* in the conditions of warranted assertion of 'x is red', and this reason is entirely independent of the status of the major premise in the 'red' paradox.

The stark problem posed by Sorites paradoxes involving predicates like 'looks red' is that the claim that '... is indiscriminable from ...' is a congruence for them seems *utterly* convincing. The function of such predicates is purely to record how things appear — so what space can they make for any kind of selective response among things whose appearances are absolutely similar? I believe that this kind of thought is convincing if, but only if, we model our understanding of such predicates in accordance with the first part of the governing view. *Rules* for the use of such predicates — predicates geared just for the description of how things appear — would indeed have to assign exactly similar appearances exactly similar worth. But if, in contrast, competence with such predicates consists not in the

260

internalization of rules but in habituation in certain patterns of response, then the thought that the responses must be the same to a pair of indiscriminable stimuli is only another version of the reasoning behind the Tachometer paradox, and may be treated accordingly. Since the apparent Sorites-susceptibility of 'it is warrantedly assertible that ... is red' depends entirely upon the possibility of characterizing the assertibility conditions of 'x is red' in terms of 'looks red' the same treatment will also serve for the Sorites paradoxes of assertibility.

It is still on our agenda to try to get much clearer about what this proposal comes to. But the foregoing reflections ought to be persuasive that a degree-theoretic approach, at any rate, is no alternative.

VI
Higher-order Vagueness and the No-Sharp-Boundaries Paradox

It was remarked earlier that in addition to the lines of reasoning which the governing view encourages, major premises for Sorites paradoxes may be motivated in (at least) two further, quite different ways. One such is the train of thought that generates the Tachometer and C- paradoxes. But it is another which seems to make the connection between vagueness and Sorites-susceptibility most explicit and most intimate. Surely, this train of thought proceeds, the very vagueness of φ must entail that in a series of appropriately gradually changing objects, φ at one end but not at the other, there will be no nth element which is φ while the $n + 1^{st}$ is not; for if there were, the cut-off between φ and not-φ would be sharp, contrary to hypothesis. Accordingly, the vagueness of φ over such a series must always be reflected in a truth of the form:

(i) $-(\exists x) (\varphi x \ \& \ -\varphi x')$

That, of course, is a classical equivalent of the universally quantified conditional which is the major premise in standard formulations of the Sorites paradox — a thought which has prompted Putnam[34] to suggest that a shift to intuitionistic logic might be of value in the treatment of the paradox. So indeed it might. But it will not be enough; for intuitionistic logic will yield a paradox direct from the negative existential premise (as will any logic with the standard $\exists$- and &-Introduction rules + *reductio ad absurdum*.)[35] This form of the paradox — the *No Sharp Boundaries*

261

paradox — thus appears to constitute a proof that (one kind of) vagueness is *eo ipso* a form of semantic incoherence.

Only 'appears' though. What the No Sharp Boundaries paradox brings out is that, when dealing with vague expressions, it is essential to have the expressive resources afforded by an operator expressing *definiteness* or *determinacy*. I have heard it argued that the introduction of such an operator can serve no point since there is no apparent way whereby a statement could be true without being definitely so. That is undeniable, but it is only to say that — in terms of Dummett's distinction[36] — the *content senses* of 'p' and 'Definitely P' coincide; whereas the important thing, for our purposes, is that their *ingredient senses* differ, the vital difference concerning the behavior of the two statement forms when embedded in negation. Equipped with an appropriate such operator, we can see that a proper expression of the vagueness of φ with respect to the relevant sort of series of objects is not provided by the above negative existential but rather requires a statement to the effect that no definitely φ element is immediately succeeded by one which is definitely not φ; formally,

(ii) $-(\exists x) [Def(\varphi x) \& Def(-\varphi x')]$

This principle generates no paradox. The worst we can get from it, with or without classical logic, is the means for proving, for successive x', that

(iii) $Def(-\varphi x') \rightarrow -Def(\varphi x)$.

And nothing untoward follows from that.

A believer in higher order vagueness[37] may want to reply that this merely postpones the difficulty. If, for example, the distinction between things which are φ and borderline cases of φ is itself vague, then assent to

(iv) $-(\exists x) [Def(\varphi x) \& -Def(\varphi x')]$

would seem to be compelled even if assent to (i) is not. Someone who sympathized with the drift of section IV will want to dispute that predicates like 'red' do indeed possess higher order vagueness — the truth is merely that the distinction between the last definitely φ item and the first case where some measure of indefiniteness enters is not to be placed just by looking. For present purposes, though, let us suppose that the phenomenon is real. Then once again the materials for paradox seem to be at hand, each ingredient move taking the form, for instance, of a transition from $-Def(\varphi k')$ to $-Def(\varphi k)$.

I believe that this thought, that higher order vagueness is *per se* a source of paradox, may quite possibly be correct. But some complication is

262

needed. For the following is the immediate reply. Of *any* pair of concepts which share a blurred boundary, we shall want to affirm

(v) $-(\exists x)$ [$Def(\varphi x)$ & $Def\,(\psi x')$]

when x ranges over the elements of an appropriate series in which the blurred boundary between φ and ψ is crossed. The original problem occurred when, with $-\varphi$ in place of ψ, we overlooked the necessary role of the definiteness operator. And now we are guilty of the same oversight again; it is merely that this time ψ has been replaced by $-Def(\varphi)$. As soon as the inclusion of the definiteness operator is insisted on, all that emerges is

(vi) $-(\exists x)$ [$Def(Def(\varphi x))$ & $Def(-Def(\varphi x'))$]

which yields nothing more than the harmless

(vii) $Def(-Def(\varphi k')) \rightarrow -Def(Def(\varphi k))$

Evidently the strategy will generalize; we need never, it seems, be at a loss for a way of formulating φ's possession of vagueness, of whatever order, in a way that avoids paradox.

But this is too quick. We are able to be confident that the sort of formulation illustrated avoids paradox only because we have so far no semantics for the definiteness operator, and are treating it as logically inert. Without considering what form a semantics for it might take, the crucial question is whether it would be correct to require validation for this principle:

(DEF) $\dfrac{\Gamma \vdash P}{\Gamma \vdash Def(P)}$; provided Γ consists of propositions, all of which are prefixed by '*Def*'

For, in the presence of DEF, and assuming that the corrected formulation, (vi) above, of what it is for the borderline between φ and its first-order borderline cases to be itself blurred, is itself *definitely* correct, the harmless (vii) gives way to

(viii) $Def(-Def(\varphi k')) \rightarrow Def\,(-Def(\varphi k))$,

a generalization of which will enable us to prove that φ has no definite instances if it has definite borderline cases of the first order.[38]

DEF has this effect because it sanctions the inference from $Def(\varphi k)$ to $Def(Def(\varphi k))$. More generally, assuming that what is definitely true is true, DEF yields the biconditional theorem.

(IT) $\vdash Def(A) \leftrightarrow Def(Def(A))$.

263

If, as was supposed, A and *Def*(A) may have differing ingredient senses, this may seem obviously unacceptable. But matters are not so straight-forward. The question is whether the difference is one to which the context, '*Def*( )', is, like 'Not( )', itself sensitive. Suppose, for instance, we had a semantics which accounted *Def*(A) *false* when A was anything other than definitely true, and Not-A as borderline when A was borderline. Not-*Def*(A) would then, presumably, be true when A was borderline — di-verging, as it intuitively should, from Not-A; but *Def*(A) and *Def*(*Def*(A)) would both be false. An approach having these features would not, I suppose, be a non-starter — the idea has some appeal that if A is on any sort of borderline, the claim that it is anything else is false. But unless we found more to say, there would be no evident objection to DEF.

There *is* more to say, of course. To take higher-order vagueness seriously is just to allow that cases may arise where it is indeterminate whether a statement is true or borderline. To say that its definitization was *false* in such a case would be, in effect, to rule that the original was borderline — to ignore its leanings, as it were, towards truth. So the sort of semantics adumbrated which promises to validate DEF is anyway guilty of failing to take higher-order vagueness seriously: to repeat, taking higher-order vagueness seriously involves allowing that *Def*(A) may itself, on occasion, be borderline.

So to allow, however, will make no difference as far as IT is concerned unless, when *Def*(A) is borderline, *Def*(*Def*(A)) may either be false or, at any rate, of an inferior degree of truth to that of *Def*(A). But the idea that *Def*(*Def*(A)) may be false when *Def*(A) is borderline can hardly be separated from the corresponding claim about *Def*(A) and A respectively — the claim just accused of failing to take higher-order vagueness seriously. So a defender of higher-order vagueness should prefer the second type of pro-posal: when a statement is borderline, so should its definitization be, but not in such a way as to sustain IT, left-to-right.

How might this proposal be elaborated? Let us pretend for a moment that we really do understand the idea of an indefinite hierarchy of orders of vagueness, along the following lines:

Level $-\omega$: statements which are false

$\cdot$

$\cdot$

$\cdot$

Level -2 : statements which are neither definitely false nor def-initely at level -1

264

Level -1 : statements which are neither definitely false nor definitely at level 0

Level 0 : statements which are neither definitely true nor definitely false

Level 1 : statements which are neither definitely true nor definitely at level 0

Level 2 : statements which are neither definitely true nor definitely at level 1

.
.
.

Level ω : statements which are true

where it is understood that each statement of finite positive level is closer to truth than any statement which is definitely of level 0; and each statement of finite negative level is closer to falsity. Then a relatively straightforward proposal would be that, for any statement of level k, its definitization is of level k-1, if k is finite or 0, but otherwise is also of level k. So if A is definitely true, or false, so is *Def*(A); but if A is on some sort of borderline, *Def*(A) lies on the immediately inferior borderline. And A $\leftrightarrow$ B will hold only if A and B are of the same level.

But the problem with this is evident enough. Any statement, A, of finite level k, > 0, is part-characterized as one which is not definitely true. So the statement of which that part-characterization is the negation ought to be *false* — and that statement ought to be *Def*(A). It is accordingly impossible to marry the sense given to '*Def*' by the proposal with our intuitive understanding of 'definitely'.

The point is a general one. We cannot intelligibly characterize higher-orders of vagueness in terms which invoke statements' failure to be definitely true, yet simultaneously require definitization to generate falsity only when applied to false statements. We could, of course, drop the latter requirement without reverting to the original idea that *Def*(A) *always* polarizes to truth or falsity. A (somewhat messy) compromise would be that *Def*(A) continues to drop a level on A just in case A occupies some finite negative level, but polarizes to falsity if A occupies any finite positive level or 0. But, as will be apparent if we now construct an explicit derivation, this compromise does nothing to obstruct the relevant version of the No Sharp Boundaries paradox:

1   (1) *Def*$(-[\exists x] [Def(Def(\varphi x)) \& Def(-Def(\varphi x'))])$; Ass.

265

2   (2) *Def*(−*Def*(φx′)) ; Ass.

3   (3) *Def*(φx) ; Ass.

3   (4) *Def*(*Def*(φx)) ; IT.

2,3 (5) (∃x) [(*Def*(*Def*(φx)) & *Def*(−*Def*(φx′)))] : 2,4, ∃ -intro.

1   (6) −(∃x) [*Def*(*Def*(φx)) & *Def*(−*Def*(φx′)))] ; 1, *Def*-elim.

1,2 (7) −*Def*(φx) ; 3,5,6, RAA.

1,2 (8) *Def*(−*Def*(φx)) ; 7, DEF.

1   (9) *Def*(−*Def*(φx′)) → *Def*(−*Def*(φx)) ; 2,8 CP

Clearly IT and DEF do not survive the mooted compromise without re-strictions: *Def*(A) will fail of equivalence to *Def*(*Def*(A)) for any A of finite negative level; and *Def*(A) will be of level lower than A for A of any finite level, negative or positive. But remember that each φx or φx′ with which we are concerned in this version of the paradox is of level 0 or greater — we start with an x′ which is a definite borderline case of φ and 'work left', so to speak. So the compromise poses no obstacle to the application of IT at line (4): *Def*(φx) and *Def*(*Def*(φx)) will both be false for every germane φx. Similarly, the compromise offers no objection to the application of DEF at line (8): if φx is of level 0 or greater finite level, −*Def*(φx) and *Def*−*Def*(φx)) will both be true, and if φx is of level ω (in which case why doesn't it mark a sharp boundary?), −*Def*(φx) and *Def* −(*Def*(φx)) will both be false. And, to stress: every φx at which we arrive by 'working left' is in one of those two cases.

I am under no illusion that these sketchy remarks constitute a treatment of the topic. But they do suggest that the No Sharp Boundaries paradox may have something to teach us, not about vagueness generally but about the idea of an indefinite hierarchy of orders of vagueness. The idea, which has smitten some writers,[39] that lack of sharp boundaries is *per se* para-doxical is merely retribution for working with too crude a formulation of what lack of sharp boundaries is. It is essential to have the expressive resources of a definiteness operator. But the case for thinking that higher-order vagueness — always a difficult and vertiginous-seeming idea — may be *per se* paradoxical still needs an answer.

There is a further, very important point to which the introduction of a definiteness operator into our formulations gives rise. Earlier, in section IV, the complaint was briefly considered that it could not suffice to resolve a Sorites paradox simply to undermine any motivation for believing the major premise. For there is a paradox in any case unless the major premise,

266

motivated or not, is *untrue* — the paradox is, indeed, a *reductio* of the major premise — and an account is therefore owing of *how*, specifically, it is untrue. Such an account is provided, for expressions like 'red', by the considerations — if sound — which were advanced in section IV to support the idea that such expressions can have determinate thresholds of application in the appropriate kind of series. But, correct or not, would nothing less than such an account, or, e.g., an account along Peacocke's lines, suffice to discharge the obligation?

The doubt whether such an account is necessary arises because — obviously enough — the presence of vague expressions may divest complex sentences of determinate truth value no less than simple predications. So it is plausible to insist that we have other options open besides affirmation or denial of the major premise in a Sorites — the premise may itself lack any determinate status. Once that is recognized, the point about the insufficiency of responses which do no more than undercut the motivation for the major premises is likely to seem less compelling. Perhaps we will have said enough about the 'manner of untruth' of the major premises when we have shown that they may reasonably be regarded as lacking any determinate truth status. And showing that there is no good reason to regard them as true would accomplish just that — provided that there is also no good reason to regard them as false.

Well, the evident problem with this is that a Sorites, as remarked, itself appears to constitute a *reductio* of its major premise, and so to exclude the response that the premise may be viewed as indeterminate. As a result, the challenge to explain how specifically the major premise is false — does one of the ingredient conditionals, e.g., cross a threshold, or are they all marginally false, or what? — is apt to seem compelling. Perhaps it could be resisted — it is a suggestive thought, for instance, that a conjunction of indeterminate conjuncts may have determinately false consequences. But with definiteness operators in our weaponry, we can now see, once again, that the challenge need not arise. What, we should claim, a Sorites establishes as false is the claim that the major premise is *definitely* true. The attempt to express the major premise without recourse to the definiteness operator is a source of needless difficulty no less than the corresponding attempt with the expression of lack of sharp boundaries. The paradox establishes only the negation of the definitization of the major premise — a conclusion quite consistent with its indeterminacy. In contrast, the challenge presupposes, illegitimately, that what is established is the definitization of the negation.[40]

267

The suggestion demands, of course, sufficient of a semantical account for the definiteness operator to sustain the intuitively plausible idea that what is really up for *reductio* is the definitization of the major premise — that it is somehow improper to put it forward undefinitized. But it does not seem farfetched to suppose that such an account can be provided. If it can, then undercutting the motivation for the major premises can be enough, provided it is acceptable methodology to regard compound statements involving vague expressions as indeterminate in truth-status if no evidence is available to the contrary. If so, one corollary is that the way adopted with observational expressions in section IV involved biting off more than it was strictly necessary to chew. (That is not to say that the conclusions there drawn are not good in any case.) Another, more general, is that the problem of explaining *how* the transition from F-ness to non-F-ness is effected in a Sorites series can be dismissed as spurious. It *may* be that there is, after all, some kind of sharp boundary, and it *may* be, in particular cases, that diminutions of degree are uniformly involved. But we are under no obligation to provide an account — there may be nothing determinately correct to say about the matter.

VII
The Governing View and the Major Premises

I now revert to the various types of ground, briefly canvassed in section I, which the governing view provides for the major premises in Sorites paradoxes. Such grounds involve reference to a number of factors including (i) standard criteria for understanding and misunderstanding an expression, (ii) the way the use of the expression is standardly explained, (iii) what is known about relevant cognitive limitations of ours — of visual acuity, or memory, for instance —, (iv) the way the applicability of the expression can standardly be determined — by casual observation, for instance —, and (v) our understanding of the point or significance — the practical consequences — of its applying. These are all considerations of a kind which, once it is accepted that our use of an expression is governed by implicitly known semantic rules, must at least be *relevant* to determining the character of the content which those rules have. But it's a different question how convincing in detail are the applications which were made of them in my [1975] and [1976].

There is, to begin with, some doubt about the use there is made of the first. It is one thing to claim that standard criteria for misunderstanding

268

'red' should have implications for the character of the rules which govern the use of that predicate; quite another to argue, as I did, that a subject who was prepared to describe one but not the other of a pair of indiscriminable color patches as 'red' would invariably give cause to think that he misunderstood the predicate.[41] Certainly, that would very *often* be so; it would, I suggest, always be so if the pair was presented to the subject in isolation and in a context in which he had no reason to suspect any abnormality of lighting, etc. But it is not obviously so precisely in the context that interests us, namely that of a Sorites series: here we should want to recognize the right of the subject to 'switch off' at some point — and since that *is* our response to the case, an argument for saying that his doing so would nevertheless be in contravention of the rules for 'red' would better draw on considerations of a different sort.

There are also reservations about the way the line of thought, which appealed to the moral and explanatory significance of terms like 'adult' and 'child', was supposed to support the major premise for the 'heartbeats of childhood' paradox.[42] It is of course true that giving these concepts mutually sharp boundaries would bring it about that important differences concerning rights and responsibilities would come to be associated, in certain cases, with no substantial change in the subject, and useful lay-psychological and lay-sociological generalizations would be placed in jeopardy. But the fact is that, in order for a certain distinction — say that between 'adolescent' and 'adult' — to serve such purposes and carry such consequences, it is not necessary that absolutely *every* instance of the distinction be able to do so. If we troubled to have an absolutely sharp distinction between adolescence and adulthood, it would still be true that most adolescents would differ sufficiently from most adults to justify the sort of differential treatment and expectations visited upon them. So the absence of a sharp boundary cannot be imposed on us by the wish to have these concepts associated with their standard moral and explanatory significance. Only if the association had to be unfailing would there seem to be such an imposition. But it would be quite good enough if it were merely usual; or so it could be plausibly argued.

At this point, however, it becomes urgent to inquire how *exactly* the kinds of consideration canvassed are supposed to yield the major premises in Sorites paradoxes. The point just considered, for example, would seem to have been whether *mutual precision* in the 'child', 'adolescent', 'adult' family of predicates would be consistent with certain important features of their use. But surely, even if we decided that it would not, the most

269

that could follow — on the standpoint of the governing view — would be that the semantic rules for these expressions prescribed that there should be *no sharp cut-offs*; and that conclusion would foster paradox only if lack of sharp boundaries was thought to be tantamount to Sorites-susceptibility — a thought which, unqualified, is just an endorsement of the No Sharp Boundaries paradox, to which an outline solution was offered in the previous section. Presenting considerations which, under the aegis of the governing view, deepen our understanding of the vagueness of certain expressions is absolutely not to the point unless it is supposed (that one who endorses the governing view must hold) that to be vague *is* to be Sorites-prone.

I was, I think, alive to this consideration in my previous discussions.[43] But how exactly do the arguments advanced negotiate the pitfall? The second, third and fourth types of consideration — those concerning standard explanation, cognitive limitations of users, and methods for assessing application of an expression — can each, it seems, be made to furnish a case for the quantified conditional

(∀ x) (it is not the case that the rules for 'red' prescribe its application to x′ → it is not the case that the rules for 'red' prescribe its application to x).

where x, as usual, ranges over the members of an appropriate series of color patches and x′ is the successor of x therein. We proceed by putting forward for *reductio* the suggestion that there is a last case in the series to which the rules prescribe application of 'red'. That is distinct, of course, from the suggestion that a case to which the rules prescribe application of 'red' is immediately followed by one to which they prescribe that 'red' does not apply. The argument, whether we bring the second, third or fourth kind of consideration to bear, is very straightforward. Mastery of a set of rules involves knowledge not just of what they require, permit and prohibit, but also of their limits: if the rules for 'red' do not prescribe application of the predicate to certain kinds of case to which they also do not prescribe its non-application, then one who has mastered those rules ought to know this and have the capacity, other things being equal, to recognize such a case if presented. So much will be part of understanding 'red'. Suppose, then, that all the distinctions which such an understanding empowers one to draw can be drawn by someone with normal limitations of memory without reliance on external aids; that nearby surveyable objects which are red can, in normal circumstances, be recognized to be so just by casual

270

observation; and that a full understanding of red can be bestowed by ostensive training. Each of these suppositions is plausible and, apparently, inconsistent with the hypothesis. If the hypothesis were true, a k to which the rules prescribed application of 'red' would be followed by a k' to which they did not. Someone who fully understood 'red' ought therefore to be able to recognize the pair as such. But doing so would require reliance upon a distinction — that exemplified by k and its successor — which was unmemorably fine, undetectable by casual observation, and — in the case where adjacent patches are indiscriminable — incapable of ostensive display. So, on the stated suppositions, the hypothesis is false for arbitrary choice of k'; and the quantified conditional is thereby established.

One response to these considerations would be to wonder whether appropriate interpolation within the argument of occurrences of the definiteness operator might make some material difference. Someone who fully understands the rules for the use of 'red' ought indeed to be empowered to recognize any k and k' such as the rules definitely prescribe applications of 'red' to k and definitely do not prescribe its application to k'. But that is to say that understanding bestows grasp of the distinction between things which are definitely red and things which are definitely something else — either definitely not-red or definitely borderline. Can anything weaker be so much as coherently formulated? In Wright [1975] I spoke of the distinction between things which are definitely red and things of "any other sort."[44] But it seems reasonable to insist that any distinction to which understanding demands sensitivity must be one between items of *determinate* status — otherwise it cannot be determinate that there *is* any relevant distinction between them, and of no response in particular can it therefore be correct to say that it is what understanding requires.

Well and good. What follows is that all we may legitimately put up for reductio is the supposition of a sharp threshold between the definite reds and the definite borderline cases. And no problem will be presented if this is expressed by straightforward definitization of what we had above, 'The rules prescribe applications of 'red' to k' and 'The rules do not prescribe application of 'red' to k'' respectively. For the worst that will then be forthcoming is

(Definitely: it is not the case that: the rules prescribe application of 'red' to k') → (It is not the case that: definitely: the rules prescribe application of 'red' to k.)

However, the supposition of sharp threshold would be as well captured

271

by conjoint assumption of

> Definitely: it is not the case that: definitely: the rules prescribe application of 'red' to k';

and

> Definitely: the rules prescribe application of 'red' to k.

And a *reductio* of this assumption does afford a conditional of paradoxical potential — one of uniform antecedent and consequent — provided, as we know, the definiteness operator may be manipulated in accordance with DEF.

The position, then, is that, in the presence of DEF and assuming the (definite) correctness of the suppositions about memorability, casual observability and ostensive teaching which were made, there is a good case for thinking that 'red' and its ilk are subject to the higher-order version of the No Sharp Boundaries paradox. That is not, indeed, the conclusion drawn in my [1975] and [1976], that the governing view constrains us to regard such predicates as tolerant with respect to indiscriminable (or unmemorable) degrees of change.[45] But it is no more comfortable a conclusion for one who wants to hold the governing view.[46]

What room is there for maneuver? DEF can be challenged, of course, though the prospects do not seem encouraging. A more plausible response, at least at first sight, is to claim that the background suppositions are at fault. The claim, for instance, that the whole meaning of 'red' may be taught by ostensive means is open to the same reproach as that earlier directed against Dummett's account of observationality: if it were right, there would be no space for a distinction between looking red and really being so. For that distinction cannot, presumably, be ostensively explained — (or, if it can, it is no longer evident why ostensive teaching could not encompass an explanation of how only one of a pair of indiscriminable items could be red.) The same consideration, indeed, seems to disarm the other suppositions as well. If we allow that a k which is definitely red can be indiscriminable, or only just barely discriminable from a k' which is definitely borderline, the memorability is thereby jeopardized of the distinction between things which are definitely red and things which are not only if that distinction has to be drawn by reference solely to appearances. But to recognize the contrast between looking red and being red — and a corresponding contrast, presumably, between looking a borderline case of red and being one — is to recognize that appearance need not be the whole

272

basis of that distinction. Likewise, it is possible to tell that something is red by casually observing it only insofar as its appearance is, in the circumstances, a good guide to its color; for the appearance is all that, to a casual observer, is apparent. There is therefore no requirement that every distinction in real color status be one to which casual observation can be sensitive.

This response does not, however, take us very far. Even if 'red' could be saved by such appeals to the 'is red' — 'looks red' contrast, no parallel hope seems to be proffered for the other contrasted term. Surely the distinction between what definitely looks red and what does not is one which it must be possible to make salient by ostensive means;[47] and surely it is a distinction which the competent can draw just by casual observation and without reliance upon external aids. These suppositions remain vividly plausible for 'looks red' even if they do not hold for 'red', and the problem remains, even if restricted in scope. Moreover, it is not even clear that the scope of the problem *has* been restricted. For to be red *is* to look red in circumstances of observation which leave nothing to be desired — however exactly that condition should be explicated. So if there is still a Sorites paradox to solve for 'looks red', then, relative to the assumption that such are the prevailing circumstances of observation, there is still a Sorites paradox for 'red'. Looking red suffices for being red when other things are equal, and, as noted towards the conclusion of section V, for being justifiably assertible to be red if there is no reason to doubt that other things are equal. It is as well, of course, to pay proper heed to the distinction between looking and being red; but doing so gets us no further forward in the present context.

Whether it is correct to suppose that 'looks red' is ostensively definable or not (see note 47), it is certain that competence with that predicate is practised by subjects who have only quite ordinary powers of observation and rely on nothing but casual exercise of those powers, eschewing in particular any use of the kind of external aids — color charts, or whatever — which could compensate for limitations of memory. So the escape route *has* to involve finding a way of avoiding drawing paradoxical conclusions from this undeniable fact. But how — if we say that competence consists in knowing certain rules and their limits? For it then seems inescapable that both every distinction prescribed by the rules, and the distinction between cases where the rules have something to say and cases where they do not, can be based only on contrasts which may be detected by casual exercise of ordinary powers of observation, without reliance on external

273

aids; and, hence, that no such distinction can be exemplified by items which conjointly exhibit no such contrast. So the conclusion still looks good: competence with expressions of the class which 'looks red' typifies *cannot* consist in knowledge of the requirements of certain rules and of their limits.

There is still the possibility of making trouble for DEF, without which the particular version of the Sorites which threatens is not, apparently, obtainable. But once we concentrate on 'looks red' and its ilk, the route to paradox outlined is anyway needlessly indirect. The *best* argument afforded by the governing view for a major premise for 'looks red' is the one bruited towards the end of section V. If items are visually indiscriminable, they look the same. So visual indiscriminability is necessarily a congruence relation for any predicate whose rules of application take account only of how things look. Our conception of the role and purpose of 'looks red', 'looks orange', etc., is, unquestionably, that we use them purely to record public appearances. So if, as the governing view permits, we may legitimately base upon this conception conclusions which concern the character of the rules governing such predicates' application, those rules must have the feature, it seems, of relating only to appearance, of prescribing application of such predicates only and purely on the basis of appearance. But then they must prescribe application of such a predicate to both members, if either, of any pair of items whose appearances are the same.

Now there is terribly little room for maneuver. It would suffice if we could somehow drive indistinguishability and sameness of appearance apart — if, while granting that the rules for 'looks red' relate only to appearance, we could provide some principled basis for denying that indistinguishability suffices for sameness of appearance. But I do not see any hopeful direction for such an attempt to take. I mention it only by way of noting one formal alternative to the response which I want to recommend. This, as the reader will anticipate, is to drop the idea that the harmless truism that we use predicates like 'looks red' to record how things appear to us has any bearing on the character of the putative semantic rules which govern their use. Since — so it seems to me — the truism could hardly fail to have the very direct bearing illustrated so long as there *are* governing semantic rules for such expressions at all, the recommendation is that we drop that assumption, and adopt, as the matter was expressed in my [1976], a "more purely behavioristic" conception of what competence with such expressions involves.

274

But what does that mean? A way to provide at least the beginnings of a proper account of the contrast is to enquire what connection there is between the obtaining of an instance of the type of state of affairs which confers truth on a token of the sentence 'x looks red', and our willingness to assent to that token. It should not be controversial, I think, that each is necessary and sufficient for the other so long as certain provisos are met which are distinctive of this class of expression; that is, for no other class of expressions do exactly these provisos subserve such a biconditional dependence. The provisos are that the judging subject understands 'looks red', that his perceptual faculties are functioning normally, that he is otherwise in good cognitive order, that x is presented to him in clear view, and that he is attentive to x.[48] Subject to these provisos, assent and truth necessarily coincide. We have, that is,

> Necessarily: if the subject and the circumstances are as required by the provisos, then 'x looks red' is true if and only if the subject assents to 'x looks red'.

Now, two quite different broad perspectives on this principle — the *proviso biconditional* — are possible. One — required by the governing view — will hold that the circumstance that 'looks red' applies to x is settled, independently of any subject's response to x, by the semantic rules which govern the use of 'looks red' in English and by how, objectively, x appears. The role of the provisos, on this view, is to ensure the subject's ability to 'track', or detect this independent state of affairs: thus, his understanding of 'looks red' and his being in "good cognitive order" ensure his sensitivity to the requirements of the relevant semantic rules; and his normal perceptual function and attentiveness to x, and the clear presentation of x to him, ensure his sensitivity to x's objective appearance. On this view, then, appeal to the idea of how a subject will or would respond to the utterance when the provisos are satisfied, should play no essential part in an account of its truth-conditions. Such an account need consider nothing but the semantics of the utterance and the characteristics of x. What the provisos do is to foreclose on every possible explanation of fracture between the fact of the matter and a subject's response.

The reason why this perspective is required by the governing view is not that the latter *platonizes* semantic rules, as it were — writes us out of the story altogether — but rather that we are allowed to feature only in a limited way. To be sure, it is by reference to *our* cognitive limitations, what *we* can casually observe, *our* conception of the function of a certain

275

class of predicates, etc., that conclusions are to be drawn about the content of semantic rules. But conclusions are, apparently, *not* to be drawn by reference to the character of our response to a predication of 'looks red' when, *by ordinary criteria*, all the provisos are fulfilled. And in order for this exclusion to be legitimate, the dictates of the semantic rules for 'looks red' have to be thought of as constituted independently of such responses. The point is, indeed, quite general and simple. In order for there to be a Sorites paradox of this kind, our actual classificatory responses have to be out of accord with the requirements of the relevant semantic rules. So to think you have such a paradox, you have to be working with an episte-mology of semantic rules which allows firm conclusions to be drawn about their character independently — or anyway, in a manner sufficiently in-attentive to — the shape which those responses are disposed to assume. Such a view need nevertheless pose no barrier to recognizing the proviso-biconditional. Whatever epistemology is utilized, it will contain the re-sources for a declaration that where responses are not in accord with the rules as it contrues them, subjects are, for instance, muddled or inconstant in their apprehension of the requirements of those rules and so, to that extent, out of good cognitive order.

That is the first broad perspective. But the alternative perspective turns it all around. Now the proviso-biconditional is seen, instead, as itself supplying the canonical form of a statement of the truth-conditions of 'x looks red'. The provisos are no longer seen as serving to describe the conditions under which a subject succeeds in tracking an independent fact; rather, for x to look red just *is* for subjects to be willing to assent to that judgment when the provisos are met. Whatever else we want to say about the meaning of such expressions, and about the epistemology of their meaning, it is all answerable to this point. Hence, there is no possibility of such an epistemology teaching us that the provisos are in fact not satisfied in circumstances where, by normal criteria, we should have been satisfied that they are. It is the other way about: if we are tempted to opinions about the meanings of such expressions which force us to draw such a conclusion, it is those opinions which are at fault. There is no ulterior fact which meeting the provisos ensures that we can detect, so no possibility — by reference to independent criteria for the existence or non-existence of such a fact — of surprising conclusions about when the provisos really are not met, notwithstanding the satisfaction of ordinary criteria for saying that they are.

What does this distinction do for us?[49] Well, suppose a group of subjects

276

are agreed that 'looks red' applies to the first color patch in a Sorites-series in circumstances when the provisos are met. As the series is run from apparently red to apparently orange patches, a point will be reached where, despite its indiscriminability from the immediately preceding patch, a consensus in subjects' responses breaks down for the first time. That is undeniable. It is also undeniable that, when that happens, no single subject in the group need, by ordinary criteria, have fallen out of accord with any of the provisos. (Borderline cases are exactly cases about which competent subjects are allowed to differ.) The second perspective on the proviso-biconditional gives us the right to treat such a circumstance as raising a doubt about the predicability of 'looks red'.[50] The first perspective, by contrast, leads to the cancellation of that right.

Someone who succumbs to the arresting, apparently simple thought that 'looks red' must be applicable to both, if to either, of any pair of indis-criminable items, is likely to be taking it in one of two ways. Taken one way, it is the central move in the Tachometer paradox: how can a signalling device, even a properly functioning signalling device, discriminate among stimuli whose difference is smaller than its sensitivity threshold? We now know, I hope, how to respond to that question. But the second and, I think, more natural way of taking the thought conceals, in effect, a pre-supposition of the first perspective on the proviso-biconditional. You have to forget that we do not, or would not so apply the predicate in every case and fall in, instead, with the idea that something can be discerned about its *proper* conditions of application just by intuitive reflection upon the kind of content which it overtly seems to have. If there were facts about the proper application of such predicates which were constituted inde-pendently of our best responses — the responses we would have when, by ordinary criteria, all the provisos were met — what else could they be but the offspring of rules which correlated their proper use with *appear-ances*? And how, when following such rules, could it ever be justifiable, in consequence, to assign to identical appearances distinct responses? The reply should be that competence with such predicates is nothing to do with the capacity to fit one's usage to the dictates of rules of that kind. Indeed, it is not a matter of compliance with rules at all, if that is taken to imply the propriety of a 'detective' direction of interpretation of the proviso-biconditional. What it is correct to say using such a predicate is a function only of what we are actually inclined to say when there is no reason to doubt that the provisos are met.

The distinction, between these two perspectives on the proviso-bicon-

277

ditional would undoubtedly benefit from further work. But it is, I hope, tolerably clear. It unravels the most important strand in what my [1975] and [1976] were attempting to say. The 'behavioristic' conception of meaning advocated towards the conclusion of the latter is just the non-detective reading of the proviso-biconditional which lets us take 'best behavior' seriously. It is, indeed, the belief in the conception of meaning which sustains a detective reading of proviso-biconditionals that constitutes the real essence of the governing view, of which its involvement with the idea of *implicit knowledge* of semantic rules is merely a corollary. For, as noted, the separation, effected by the detective reading, between on the one hand what *constitutes* the correctness of a particular response with 'looks red', for instance, and on the other the fact that the response observes the provisos, enjoins viewing the provisos as collectively ensuring a subject's ability to keep cognitive track of the independent, constitutive fact. So there has to be substantial knowledge of the ingredients of that fact, including the requirements of the relevant semantic rules. Since the subject's response will usually come to him involuntarily and unreflectively, it would appear that the relevant cognitive processes must usually be implicit. One does not need to be wary of the notion of implicit knowledge in general to reject this way of involving it in the theory of linguistic understanding. The rules to which competent use of 'looks red' conforms — if indeed we wish to continue to think in terms of semantic rules for such expressions at all — are not rules knowledge of whose content, at some deep level, makes competence possible; rather their content is given by the course assumed by competent use.

It remains to respond to a question outstanding from the end of section IV. An observational expression, it was there suggested, should be characterized as, inter alia, one which a subject could not fail to understand if his use of it displayed an appropriate sensitivity to a certain distinctive range of appearances. The question was whether the presence of this condition in an account of observationality could be seen to spare observational expressions the tolerance which Dummett's account, for instance, would lumber them with. "How could the *appropriate* kind of sensitivity operate selectively among indiscriminabilia."[51] If what I have been saying has any force, it should now seem as if this question has rather disappeared. 'Looks red' ought certainly to qualify as observational on this count. But the kind of sensitivity to appearance which someone who understands 'looks red' (and is normally-sighted) must have just *does* operate selectively among items which, in respect of apparent color, cannot be told apart. If,

278

for instance, we present a subject with each of the elements of a Sorites-series of indistinguishable color patches but *out of sequence*, such selectivity will have to be manifest in the collation of his or her responses to the question, 'Does this look red?', so long as we are to be satisfied that the provisos have been met in every case. The suggestion that there is some kind of tension between such selectivity of response and its being the sort of response appropriate in the case of an observational predicate, depends on the thought that it cannot then be purely in response to appearance — either it is unprincipled or else it is a principled response to *more* than the appearances. But 'unprincipled' here just means: not guided by rules correlating responses with appearances. So we should embrace the first alternative: such responses are indeed unprincipled, and no less appropriate or less purely 'to' appearances on that account.

VIII

Conclusion

We have covered a lot of ground, and a substantial proportion of our findings — most especially about degree-theoretic responses to the Sorites — have been negative. But a number of positive points have emerged which, if I am right about them, are worthy of emphasis. First, it is a mistake to think of the Sorites as a *single* paradox, which admits of a large variety of illustrations. Rather, we have here a family of different paradoxes, formally similar but distinct when identified by reference to the type of ground which supports their major premises. By this criterion, the Tachometer paradox, the No-Sharp-Boundaries paradox, and the various Sorites paradoxes which are the progeny of the governing view, all need to be distinguished and treated differently. A further important group are constituted by the Sorites paradoxes which affect predicates of practical intellectual possibility — like 'intelligible' when applied to expressions, 'memorable' when applied to patterns, 'surveyable' when applied to proofs, and so on. The proper response to paradoxes in this last group is a matter of great consequence for finitism in the philosophy of mathematics, but I have no space to engage the issues here.[52]

Second, the distinctions among the Sorites family correspond to a variety of lessons which the broad lines of solution to these paradoxes contain for us. The Tachometer paradox is perhaps the least interesting in this respect.

279

The positive lesson from it is only the very local point that what it is for an instrument or human subject to have a threshold of sensitivity is something which needs careful description if incoherence is to be avoided. That is a point which it is as well to know, but it hardly trembles the foundations of the subject. Still, as recently noted, the paradox represents one way of being seduced by the thought that phenomenal predicates like 'looks red' apply to both, if to either, of any pair of indiscriminable items, so it needs a proper resolution. And pursuit of the signalling instrument analogy which arose in the course of our discussion of the paradox led us indirectly towards what, I would contend, is the germ of a good account of the distinguishing features of *observational* vocabulary. The essence of the No-Sharp-Boundaries paradox, on the other hand, is the thought that vagueness is *per se* Sorites-generating, and hence that the paradox afflicts almost the entirety of our language. This thought is defective, or so I argued, but points up the need for definiteness-operators in any coherent philosophical treatment of vagueness, and for a proper account, in turn, of the semantics and logical behavior of such operators. Even so, the question was left unresolved whether paradox might not be inherent in higher-order vagueness at least. Since, as noted, at least two important writers on vagueness have taken the view that higher-order vagueness is essential to a large class of vague expressions, this is an issue on which further work is much needed.

Third, we have found ourselves taking a somewhat unexpected route to one of the central concerns of Wittgenstein's later philosophy of language, mathematics and mind: the issues to do with what I have elsewhere called the *objectivity of meaning*[53] which seem to me to be the principal focus of his discussions of rule-following in the *Investigations* and *Remarks on the Foundations of Mathematics*.[54] One who accepts the objectivity of meaning thinks of understanding an expression in quasi-contractual terms: there is, for instance, when one confronts the question whether or not to classify a particular object in a particular way, a response to which one is *obligated* if one is to keep faith with the meaning of the relevant expressions as they were taught to one (and pay due heed to the relevant wordly facts).[55] Less metaphorically, finite, over-and-done-with episodes of explanation and other forms of linguistic behavior may succeed in fully determining the meaning of a so far unmade utterance in a fashion independent of any reaction which we will or would have to it: there is only the question of whether our reaction, if one is forthcoming, is in line with the meaning so independently constituted. Wittgenstein's discussion has tempted many commentators to conclude that he is advocating scepticism about the very

280

notions of meaning and of correct or incorrect linguistic practice.[56] But that is as much a misreading as the opposite tendency which finds in Wittgenstein's discussion only a corrective to a crudely phenomenalistic conception of understanding, intention, and other cognate notions.[57] What is at issue is the conception of our relationship with the meaning of a novel utterance as being *purely cognitive*, of its meaning as something which, with any necessary help from the world, can determine the utterance's truth value quite independently of any contributive reaction from us. There is and can be no such independence. We are ceaselessly actively involved in the determination of meaning. This is not to say that a reference to 'us' or to the 'community' is somehow implicit in a full account of the truth-conditions of any sentence, — still less that truth is what we take to be the truth, or anything of that sort.[58] Wittgenstein, like anyone else, was aware that whole communities can be in error. The point is rather that failure to keep track of independently constituted meanings is not intelligible as a separate source of such error, alongside factual ignorance, misperception, illusion, prejudice, forgetfulness and the like. The platitude, that the truth-value of an utterance is a function only of its meaning and of the context and prevailing states of affairs, is not in question either. What is in question is the idea that the meaning of the utterance is something which is as much constituted independently of our response to it as are the germane features of the context and prevailing states of affairs, and no less something which, in appraising the utterance's truth-value, we have to cognize.

As Wittgenstein saw, the conception of meaning as objective in this way is prerequisite for the sort of platonism in mathematics which views mathematical activity as the exploration of our conceptual constructs. It is also prerequisite for the thought — essential to the 'private linguist' — that just by a personal ostentive definition, or by otherwise going through the motions of forming an intention to use an expression in a certain way, it is possible to generate facts about the correct use of that expression on subsequent occasions which are independent of our reaction on reaching them, and supply, at least in principle, the standard which those reactions have to meet.[59] It is the objectivity of meaning which is at the heart of the governing view and which sustains the detective reading of the sort of proviso-biconditional discussed in the last section. The principal lesson of that discussion is thus that, more than being deeply problematical for the reasons which Wittgenstein's deliberations bring out, the objectivity of meaning is actually a source of paradox.

281

The third species of Sorites paradoxes — those issuing from the governing view — thus bring us up against absolutely basic questions in the philosophy of language. But how general is the challenge to the objectivity of meaning which they uncover? It is not confined to 'looks red and its ilk since, as noted above, being red *is* looking red when other things are equal. So both secondary quality predicates and, when they are distinct, what we might term their phenomenalizations — the corresponding 'looks', 'sounds', etc., predicates — come within range. Also in range are all predicates of the sort, 'is justifiably characterizable as ...' for whose correct application looking, or sounding, etc., so-and-so suffices, *ceteris paribus*. This is a wide class but hardly a comprehensive one. While it is possible to argue for a global rejection of the objectivity of meaning on the basis of its failure for this class,[60] I do not know of any such argument which does not draw on considerations quite different to those with which we have been occupied here. In my earlier papers I rather tended to write as though the governing view paradoxes show us a route to rejection of the objectivity of meaning which is quite different to anything to be found in Wittgenstein's writings but of no less generality. But that now seems to be an over-statement. The best paradoxes of the kind — par excellence that for 'looks red' on which we have concentrated — depend on quite special considerations, and no global extension seems to be in prospect.[61]

NOTES AND REFERENCES

1. Wright, C. [1975] "On the Coherence of Vague Predicates," *Synthese* 30, 325-65.
          [1976] "Language Mastery and the Sorites Paradox," in *Truth and Meaning: Essays in Semantics*, eds. G. Evans and J. McDowell, Oxford, 223-47.
2. Though other forms of major premise will serve. Overlooking this has caused problems for some writers — see the remarks about Peacocke and Putnam below, on pp. 251-2 and 261-2 respectively.
3. Dummett, M. [1975] "Wang's Paradox," *Synthese* 30, 301-24.
4. Peacocke, C. [1981] "Are Vague Predicates Incoherent?" *Synthese* 46, 121-41.
5. For some reflections on it see Wright, C. [1986] "Theories of Meaning and Speakers Knowledge" in *Philosophy in Britain Today*, ed. S.G. Shanker, Croom Helm, 267-307; reprinted as chapter 6 of Wright, C. [1986a] *Realism, Meaning and Truth*, Blackwell, 204-38.
6. In comparison with what is needed, it is not unduly harsh to say that much of the literature on this topic which has mushroomed over the last decade or so has amounted to little more than tinkering. Writers have been content to devise more or less *ad hoc* semantics in whose terms the major premises fail of strict truth without doing anything to disclose *why* their plausibility is specious, still less confront the intuitively powerful arguments on their behalf. It should go without saying that the solution of a paradox requires more than designing wallpaper.

282

7. Dummett [1975], 319-20.
8. A point Wittgenstein stresses repeatedly in the observations on consistency in the *Remarks on the Foundations of Mathematics*, 3rd edition, Blackwell 1978.
9. Peacocke [1986], 122.
10. Pictorially:
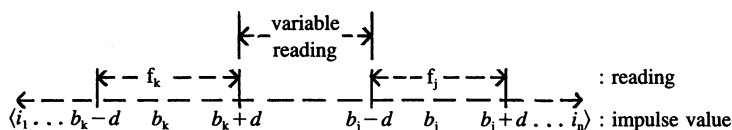
11. Peacocke [1981], 123.
12. Dummett [1975], 316.
13. It may occur to the reader that, as I have presented the example, the raising and lowering of hands will involve judgments in which an observational *relation* — 'seems to be differently positioned from' — rather than an observational predicate is an ingredient. I do not think this qualifies the force of the example, however. First, the holding aloft of zero, one, or two hands — while undoubtedly 'observational' responses in the sense which currently occupies us — would have to be encashed in terms somewhat like, 'looks to be in the original position', 'looks to be in the four second position', and 'looks to be in the eight second position' respectively. It is not obvious that these are not properly regarded as observational predicates. Second, the point that the example is being used to make — that a tendency sometimes to respond differentially to a pair of stimuli need imply no capacity to tell them apart — could be illustrated in lots of other ways in any case.
14. Peacocke [1981], 124.
15. Note that observational decidability, in the intended sense, does not preclude the intrusion of considerations from scientific theory into our conception of *when* 'unaided observation' is to be able to deliver the correct verdict. So 'red' can count as observationally decidable consistently with the occurrence of red illusions associated, e.g., with red shift and the Doppler effect on light traversing galactic distances. Observational decidability, like all concepts of possibility, is a concept of what is possible *in appropriate circumstances*; but our idea of which circumstances are appropriate may well be theoretically informed.
16. Peacocke's discussion takes a rather strange turn just after the passage quoted on p. 243, on which he proceeds to base a refinement of the idea of an observational predicate. (See Peacocke [1981], 127-8) According to the refinement, observational predicates are distinguished by their satisfaction of the condition that, with respect to any pair of indiscriminable items x and y, it cannot be the case that such a predicate definitely applies to x but does not definitely apply to y; formally,

$$Ixy \rightarrow -(Def(Fx) \& -Def(Fy))$$

How exactly does this proposal prevent observational predicates from being generically Sorites-susceptible? Peacocke's answer, in a familiar tradition, is that such predicates are predicates of degree. His own positive response to the paradox, to be reviewed in the sequel, is that the consequents of the conditionals which are instances of the major premise will typically be *slightly less* true, accordingly, than the antecedents. He is thus able to maintain that the consequent of the above conditional may have a lower degree of truth than its antecedent consistently with the conditional's retaining its distinctive acceptability for observational expressions: it will be so acceptable just in case whenever the antecedent is true, the consequent is *almost* true. Now, the consequent will, presumably, be almost true just in case the conjunction, $Def(Fx) \& -Def(Fy)$, is almost false; which, intuitively, should be the situation when one, or both of the conjuncts are

283

almost false, although neither is actually false. Hence if *Def*(Fx) is true, −*Def*(Fy) will be almost false, so *Def*(Fy) will be almost true. But that, on Peacocke's view, will be the most that can be inferred from the acceptability of the original conditional, the indiscriminability of x and y, and the truth of *Def*(Fx). So the paradox is obstructed. (The reader will be able to see how similar remarks would apply to the attempt to work the paradox from right-to-left, as it were, taking as premise −*Def*)Fy).)

What immediately strikes one, however, is that, once Peacocke has a conditional which allows of almost-truth, there is no need for this reformulated account of obser-vationality. For the mere *almost-truth* of the conditional

Ixy → (*Def*(Fx) → *Def*(Fy))

would not subserve the Sorites paradox either. The way in which Peacocke makes his preferred account of observationality avoid paradox is equally open to the original, seemingly inevitably paradoxical version. Still, that is just a quibble. The substantive point remains that if we wish to avoid trafficking in degrees of truth, we owe a different kind of account of what observationality is — if not the sort outlined in the text, than another.

17. Wright, C. [1986b] "Rule-Following, Meaning and Constructivism" in *Meaning and Interpretation*, ed. C. Travis, Blackwell, 271-97, especially 276-80.

18. This proposal is actually very close to the account of observationality offered by Pea-cocke himself in his [1983] *Sense and Content*, Oxford, chapter 4. Peacocke is there concerned not with the Sorites but with the prevailing scepticism among philosophers of science concerning the very idea of an observational concept/statement. I doubt if I should have conceived the present suggestion without the benefit of conversations and correspondence with Peacocke prior to the publication of his [1983]. But the reader should be wary of assuming that Peacocke would endorse my suggestion; it contains, in particular, no explicit analogue of the play with *experience* which Peacocke's treatment involves.

19. The proposal that if someone's use of 'red' manifests an appropriate sensitivity to color-appearances, that suffices for their understanding 'red', leaves us free, of course, to ponder just what kind of sensitivity *is* appropriate. In particular, we are free to insist that the sensitivity be informed by the conception that red is a property of the object of predication, and so to require, e.g., that the subject acknowledge certain a priori principles which evince this conception — for instance, that red things retain their color when unperceived. Analogues of all such principles may be expected to hold for natural kind terms also, so there is no threat here of compromise of the contrast drawn in the text.

20. *Locus classicus*: Goodman, N. [1951] *The Structure of Appearance*, Cambridge, Mass., ch. IX, section 2.

21. And it would be unwise to rule out the possibility of disjunctive formulations as well, although (∀ x)(− red(x) v red(x')) — the simplest disjunctive analogue of the conditional and conjunctive formulations — would intuitively offend against the fact that 'red' possesses borderline cases.

22. See note 16 above.

23. Peacocke [1983], 126.

24. See my [1983] *Frege's Conception of Numbers as Objects*, Aberdeen University Press/ Humanities Press, ch. 4, note 8, 180-4.

25. Peacocke is fully aware of the difficulties which beset attempts to elucidate the kind of notion of identity of color which he wants, but he conceals their force from himself with a rather curious thought, viz.

... Since these difficulties equally effect sharp observational notions, .. [they] .. cannot be the source of the Sorites Paradox ... [Peacocke, 1981, p. 126]

284

I am not sure what the sharp observational notions are which he has in mind. But the excuse would be appropriate, anyway, only if it was necessary to have recourse to Goodmanian ideas in order to develop the Sorites paradox in the first place. To someone who then thought that the *solution* to the paradox might be to make trouble for such ideas, Peacocke's remarks could possibly be appropriate. But the fact is that the ideas in question are not involved in the *paradox* in that sort of way at all: what they are involved in is Peacocke's attempted solution, and any difficulties which they encounter are difficulties for it.

26. No doubt many people's grasp of these distinctions is somewhat shaky; the words derive from the various animal, vegetable and mineral origins of the strikingly red dyes to which they were once applied. 'Crimson', e.g., was the name of a dye derived from the pregnant female *Kermes* beetle, which in turn took its name from the Kermes oak, native to Southern Europe and North Africa, on which it feeds.

27. Peacocke [1981], 128.

28. Peacocke [1981], 129.

29. Fine, K. [1975] ''Vagueness, Truth and Logic,'' *Synthese* 30, 265-300, especially 271 and following. Cf. Dummett [1975], 310-11. A *sharpening* of a vague predicate is a stipulation which renders it (more) precise, and is *admissible* just in case it respects uncontroversial cases, i.e., does not reclassify any definite Fs or definite non-Fs. Then if S is such a sharpening of 'red', and 'red$_S$' expresses the result, it seems we can affirm both

>   If a is red, a is red$_S$,

and

>   If a is not red, a is not red$_S$,

although the converse of either may fail since both red$_S$ items and non-red$_S$ items may be borderline cases of 'red'. Likewise things that are not C may —since even an ideal community may fail to agree about the redness of borderline cases — nevertheless not be non-red, so the converse of Peacocke's second principle fails in similar manner. Not so the converse of his first principle, presumably, but that is not really a point of disanalogy: C corresponds to the, as it were, left-most admissible sharpening of 'red', which classifies all but the definitely red things as falling under the contrary of the sharp predicate.

I have unfortunately no space in this paper to discuss the bearing on the Sorites paradox of Fine's approach to vagueness. The basic idea, as is familiar, is that sentences containing vague expressions are true only if they remain so under *every* admissible sharpening of those expressions. Hence the negations of major premises for Sorites paradoxes will generally turn out true — without implication of a 'fictitious boundary' — because, for instance, no matter how we sharpen 'red', there will be in any appropriate series — if the sharpening is both admissible and complete — a determinate last item to which the sharpened predicate applies immediately followed by a first item to which it does not. I regard the approach as unsatisfactory principally because there is no evidence that our use of sentences containing vague expressions is sensitive to such truth-conditions, and plenty of evidence that it is not — for instance, a widespread squeamishness about Excluded Middle; and also because the notion of an *admissible* sharpening manifestly inherits Sorites-susceptibility from any predicate which is so susceptible — if 'red' tolerates indiscriminable change, for instance, so does 'would be classified as ''red'' under any acceptable sharpening of that expression' — and so cannot contribute towards a satisfactory solution of the paradox until independently sanitized.

30. The reflections of section IV (p. 246 above) would actually support a stronger second principle than Peacocke's stated version, viz.

285

Any object which is not red the community will agree in not calling 'red'.

But there are ambiguities here — see ensuing note 31.

31. It is worth considering — to anticipate somewhat the business of the next main section — how the connection demonstrated between the Sorites paradoxes for 'red' and C would fare if 'definitely' was appropriately interpolated into the argument. The first of Peacocke's principles then becomes

Any object which is definitely red the community will agree in calling 'red'.

But what is the second? Is it

Any object which is definitely not red the community will not agree in calling 'red'?

Or the stronger

Any object which is not definitely red the community will not agree in calling 'red'?

It is the former which is supported by the reflections of section IV (via something with a stronger consequent — see note 30) but the latter which, as the reader may verify, is required to link the major premises for C and '(definitely) red' in the manner illustrated. And given the stronger version of the second principle, it is still true that if C has a sharp cut-off in a Sorites-series for 'red', then, where k is the last C patch, it will follow that it is not-not-definitely-red; whereas its successor, qua not-C, will — by Peacocke's first principle — be not-definitely-red. So the dilemma remains: either Peacocke can't link the 'red' and C-paradoxes in the way he — ill-advisedly — wants, or the relevant kind of series must always contain a determinate sharp boundary describable in terms of 'red'.

To revert just for a moment to the 'supervaluational' approach of Fine and others, briefly considered in note 29: clearly — if a definiteness-operator is to occur in its antecedent — the principle for 'red$_S$' corresponding to Peacocke's second principle has to be

If a is definitely not red, a is not red$_S$,

rather than

If a is not definitely red, a is not red$_S$.

So there is no prospect of a similar argument that 'red$_S$' has to be Sorites-susceptible if 'red' is. But that does nothing to mitigate the corresponding charge about 'would be classified as "red" under any acceptable sharpening of that predicate'. (Note, by the way, that the two 'definitely'-free conditionals for 'red$_S$' cited in note 29 are both supervaluationally unacceptable, since there may well be ways of sharpening 'red' so as to make it false that, e.g.,

If a is red, a is red$_S$,

viz. sharpenings that place the red/not-red cut-off further across towards the non-reds than the cut-off for red$_S$. Thus, from the supervaluational point of view, what those two conditionals try to say can only be said coherently by recourse to the definiteness-operator.)

32. Wright [1975] 348-51 and [1976] 237-40.

33. Peacocke [1981] note 11, page 140.

34. Putnam, H. [1983] *Realism and Reason, Philosophical Papers Volume 3*, Cambridge, ch. 15, "Vagueness and Alternative Logic," especially 284-6. Intuitionistically, $-(\exists x)(\varphi x \,\& -\varphi x')$ is equivalent to $(\forall x)\,\varphi x \rightarrow --\varphi x')$, but not to $(\forall x)(\varphi x \rightarrow \varphi x')$ unless φ is effectively decidable; for non- effectively-decidable φ, the last may consistently be denied without commitment to $(\exists x)(\varphi x \,\& -\varphi x')$.

35. See Read, S. and Wright, C. [1985] "Hairier than Putnam Thought," *Analysis*, 45, 56-8. Putnam has since protested — [1985] "A Quick Read is a Wrong Wright," *Analysis*

286

45, 203 — that he never intended to recommend *denial* of (∃ x) (φx & −φx'); his point was only that, intuitionistically, we could accept −(∀ x) (φx → φx') — as the paradox seems to force us to — without commitment to the fictitious boundary imported by asserting (∃ x) (φx & −φx'). But Read and I said nothing to the contrary. Our point was that bringing the distinctions of intuitionist logic to bear on vague statements — for which, Putnam acknowledges, the semantic motivation must be quite unlike that appealed to in the mathematical case — won't stop the Sorites paradox. We have, independently, to find something wrong with the impression that −(∃ x) (φx & −φx') is a satisfactory expression of the vagueness of φ in the relevant series of objects.

In Read and Wright [1985] we remarked that there would still be a place for at least one intuitionistic ploy — rejection of the inference from − −(∃ x) (φx & −φx') to (∃ x) (φx & −φx'). Putnam acknowledges this, but in fact the argument to 'unmotivate' −(∃ x (φx & −φx') about to be developed in the text shows that even this is wrong: − −(∃ x) (φx & φx') is as bad an expression of the proposition of which the paradox is a legitimate proof by *reductio* as −(∃x) (φx & φx') is of φ's vagueness in the relevant series.

36. Dummett, M. [1981] *Frege: Philosophy of Language*, Second edition, Duckworth, 446-7.

37. A nice example of succumbing to the allure of the idea is provided by Dummett [1978] *Truth and Other Enigmas*, Duckworth, chapter 11, "Wittgenstein's Philosophy of Mathematics, page 182:

> ... the vagueness of a vague predicate is ineradicable. Thus 'hill' is a vague predicate, in that there is no definite line between hills and mountains. But we could not eliminate this vagueness by introducing a new predicate, say 'eminence', to apply to those things which are neither definitely hills nor definitely mountains, since there would still remain things which were neither definitely hills nor definitely eminences, and so *ad infinitum* [Sic].

A sophisticated discussion of the topic is Fine's [1975] section 5.

38. A derivation is given on pp. 265-6 below.

39. The most notorious case, I suppose, is Peter Unger. A typical example of his use of the Sorites is his [1979] "I do not exist" in G. Macdonald, ed., *Perception and Identity: Essays presented to A.J. Ayer*, MacMillan, 235-51.

40. Refer to note 35, concluding remarks.

41. Wright [1976], 234.

42. Wright [1976], 231-2.

43. Wright [1975], 334-5; [1976], 229-30.

44. Wright [1975], 342.

45. Wright [1975], 333-4; [1976], 229.

46. It is not clear, however, that the argument about 'child', 'adult', etc., based on considerations about moral and explanatory role, can be similarly reconstructed. Even if — contrary to the suggestion in the text above — *every* instance of the adolescent/adult distinction has to involve substantial difference if the moral and explanatory role of these expressions is not to be compromised, the corresponding claim about the adolescent/definite-borderline-case-of-adolescence distinction is rather less plausible. The trouble is that whereas each of

> If x is an adolescent, there are certain moral demands which it would not be reasonable to make on x but which it would be reasonable to make, ceteris paribus, on an adult;

and

287

> If x is an adult, there are no moral demands which it would not be reasonable to make on x but which it would be reasonable to make, ceteris paribus, on an adult;

is unexceptionable, no such conditional beginning

> If x is definitely on the adult/adolescent borderline, ...

is clearly correct. And some such conditional is going to be needed in order to reconstruct the argument, since it will have to be argued that a sharp distinction between adolescence and definite-borderline status would be associated with disproportionate moral/explanatory significance. But I leave the reader to ponder the matter.

47. That is not the same as saying that 'looks red' is ostensively definable. The claim in the text would still hold if 'looks more than six feet tall' were substituted for 'looks red'. So the question of ostensive definability turns on whether 'looks red' is genuinely semantically structured, with 'red' a significant component in that structure. A bad reason for an affirmative answer would be based on confusing 'looks red' with, 'looks as if it is red' or 'looks the way red things look', since in special circumstances it may be necessary, in order to look the way red things look, to look, e.g., brown (if, say, bathed in green light). To be sure, it is a priori that looking red is looking the way red things look in good lighting, etc. But that does not entail that understanding 'looks red' involves grasp of this a priori truth. For a sophisticated discussion of such matters, see Peacocke, C. [1984], "Colour Concepts and Colour Experience," *Synthese* 58, 365-81.

48. The corresponding provisos for 'x is red' would include not merely that x be presented in clear view but that the circumstances of presentation be normal, or appropriate.

49. This very important form of distinction was utilized by Mark Johnston in classes on ethics in Princeton in the spring of 1986. It is further discussed in Wright, C. [1987], "Realism, Anti-realism, Irrealism, Quasi-Realism," in *Midwest Studies in Philosophy*, volume XII, eds., P. French, H. Wettstein and T. Uehling, University of Minnesota Press, 000-000; and put to work in Wright, C. 1987a, "On Making Up One's Mind: Wittgenstein on Intention," in *Logic, Philosophy of Science and Epistemology*, edited by P. Weingartner and G. Schultz, Proceedings of the Eleventh International Wittgenstein Symposium, held at Kirchberg-am-Wechsel, Austria, August 1986, 266-79.

50. Because A $\rightarrow$ (B $\leftrightarrow$ C) entailes B $\rightarrow$ (A $\rightarrow$ C).

51. See above, page 250.

52. Some preliminary discussion may be found in my [1982] "Strict Finitism," *Synthese* 51, 203-82, especially 262-9. The paper is reprinted in Wright [1986a].

53. Wright [1986a], *Introduction*, 3-8 and 26-9, and [1986b].

54. One is more wary of quoting Wittgenstein in support of a particular interpretation of his writings than virtually any other philosopher. But a sceptical reader should be reminded of these passages from *Philosophical Investigations*:

> 188. Here I should first of all like to say: your idea was that that act of meaning the order had in its own way already traversed all those steps: that when you meant it your mind as it were flew ahead and took all the steps before you physically arrived at this or that one.
>
> Thus you were inclined to use such expressions as: "The steps are *really* already taken, even before I take them in writing or orally or in thought." And it seemed as if they were in some *unique* way pre-determined, anticipated — as only the act of meaning can anticipate reality.
>
> 218. Whence comes the idea that the beginning of a series is a visible section of rails invisibly laid to infinity? Well, we might imagine rails instead of a rule. And infinitely long rails correspond to the unlimited application of a rule.

288

219. "All the steps are really already taken" means: I no longer have any choice. The rule, once stamped with a particular meaning, traces the lines along which it is to be followed through the whole of space. — But if something of this sort really were the case, how would it help?

No; my description only made sense if it was to be understood symbolically. — I should have said: *This is how it strikes me*.

When I obey a rule, I do not choose.

I obey the rule *blindly*.

Likewise consider these passages from the *Remarks on the Foundations of Mathematics*:

IV. 47. Might I not say: if you do a multiplication, in any case you do not find the mathematical fact, but you do find the mathematical proposition? For what you *find* is the non-mathematical fact, and in this way the mathematical proposition. For a mathematical proposition is the determination of a concept following upon a discovery.

You *find* a new physiognomy. Now you can, e.g., memorize or copy it. A *new* form has been found, constructed. But it is used to give a new concept together with the old one. The concept is altered so that this *had* to be the result. I find, not the result, but that I reach it. And it is not this route's beginning here and ending here that is an empirical fact, but my having gone this road, or some road to this end.

48. But might it not be said that the *rules* lead this way, even if no-one went it? For that is what one would like to say — and here we see the mathematical machine, which, driven by the rules themselves, obeys only mathematical laws and not physical ones.

I want to say: the working of the mathematical machine is only the *picture* of the working of a machine. The rule does not do work, for whatever happens according to the rule is an interpretation of the rule.

VII. 42. When I said that the propositions of mathematics determine concepts, that is *vague*; for '2 + 2 = 4' forms a concept in a different sense from '$p \supset$ p', '$(x) . fx \supset fa$', or Dedekind's Theorem. The point is, there is a family of cases.

The concept of the rule for the formation of an infinite decimal is — of course — not a specifically mathematical one. It is a concept connected with a rigidly determined *activity* in human life. The concept of this rule is not more mathematical than that of: following the rule. Or again: this matter is not less sharply defined than the concept of such a rule itself. — For the expression of the rule and its sense is only a part of the language-game: following the rule.

One has the *same* right to speak of such rules in general, as of the activities of following them.

Of course, we say: "all this is involved in the concept itself," of the rule for example — but what that means is that we incline to *these* determinations of the concept. For what have we in our heads, which of itself contains all these determinations?

289

55. The point of Wittgenstein's concentration upon simple arithmetical series is that it enables him to drop this part of the proviso.
56. See most notably S. Kripke's [1982] *Wittgenstein on Rules and Private Language*, Blackwell.
57. A criticism which could be levelled at the otherwise useful contributions of M. Budd [1984] "Wittgenstein on Meaning, Interpretation and Rules" *Synthese* 58, 303-23, and C. McGinn [1984], *Wittgenstein on Meaning*, Blackwell.
58. The matter is so widely misunderstood that it is still worth emphasizing Wittgenstein's explicit denials at *Philosophical Investigations* 241, *Remarks on the Foundations of Mathematics* VII, 40, and elsewhere.
59. That is not to say that 'private language' falls with objective meaning. It has still to be shown that there is no other way of establishing content for the 'seems right'/'is right' contrast which the private language has to have. For an attempt to show just that (and to explain why the contrast is necessary), see my [1986c] "*Does Philosophical Investigations I* 258-60 suggest a cogent argument against Private Language" in P. Pettit and J. McDowell, eds., *Subject, Thought and Context*, Oxford, 209-66.
60. See Wright [1986b], 292-4.
61. Versions of parts of this material were presented at colloquia at M.I.T., Johns Hopkins University, and the University of Michigan at Ann Arbor in Spring 1986. I am grateful to those who participated for a number of useful comments and questions.

290