IX

Language-Mastery and the Sorites Paradox*

CRISPIN WRIGHT

I

Throughout Frege's writings are scattered expressions of the conception that the vagueness of ordinary language, and especially the occurrence of predicates for which it is not always determinate whether or not they may truly be applied to an object, is a defect. His reason for such a view seems to have been that orthodox logical principles fail when applied to sentences containing expressions whose range of application has been defined only partially.[1] Thus Frege seems not to have considered, or not to have thought worth considering, the possibility that vague terms might require a *special* logic. Vagueness is rather something which can and should be expurgated from language, if it is to be suitable for 'scientific purposes'. The same conception is to be found in Russell's Introduction to the *Tractatus*. Ordinary language is always more or less vague, but a logically perfect language would not be vague at all; so the degree of vagueness of a natural language is a direct measure of its distance from being everything which it 'logically' ought to be.

Of course, we have since learned a greater respect for language as we find it; we no longer regard the vagueness of ordinary language as a defect. But a higher-order analogue of the Frege-Russell view continues to figure in our thinking about language: even if many predicates in natural language are vague, there can still be a precise semantics for such expressions and indeed for the whole language, i.e. a theoretical model of the information assimilated in learning it as

* This discussion is a synopsis of, or, better, a series of excerpts from my 'On the Coherence of Vague Predicates', forthcoming in *Synthese*.

[1] Excluded Middle is the obvious example. But, as Frege points out, contraposition also fails: *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*, Band ii (Hermann Pohle, Jena, 1903, and Olms, Hildesheim, 1962), p. 65.

a first language or, equivalently, of the conceptual apparatus possession of which constitutes mastery of the language. There need be no imprecision, it seems, in such a model; at any rate, none occasioned purely by the vagueness of the expressions of mastery of whose senses it is to provide an account.

We tend to picture our use of language as something essentially *regular*. We tend to think of language-learning as ingestion of a set of rules for the combination and application of expressions. Thus the task of a philosophical theory of meaning, in one natural sense of that phrase, would be to give a systematic account of the contribution made by the constituents of a semantically complex expression to its overall sense; and the theory would be concerned especially with the epistemology of the transition from understanding of subsentential components of a new sentence to recognition of the sense of the whole. Such a philosophical theory will normally only be concerned with *types* of contribution made by constituent expressions; in just this connection arise the familiar questions concerning the nature of the distinction between proper names and other singular terms, between singular terms generally and predicative expressions, whether the notion of reference may illuminatingly be extended to predicative expressions, etc. So the completion of such a theory would only be a preliminary to what we think of as a full semantic description of a natural language; for it is not just the type of contribution but the *specific* contribution which a constituent expression makes to complex expressions containing it which we think of as determined by rule.

It is worth emphasizing that no obstacle to such a conception is posed by the fact that we cannot in general state such rules in such a way as to explain the sense of an expression to someone previously unfamiliar with it. Consider a schematic rule for a one-place predicate, F:

> F may truly be applied to an individual, a, if and only if a satisfies the condition of being ϕ.

How should we specify ϕ if F is 'red'? Clearly the only completion of the rule which is actually constitutive of our understanding of 'red', rather than a mere extensional parallel, is to take ϕ as the condition of being red. In general we cannot expect instances of such a schematic rule to be of explanatory use if they are stated in a given language for a predicate of the same language; in consequence, it will not generally

be possible to appeal to such rules to settle questions about the applicability of an expression. Nevertheless we may still legitimately regard such a rule as an exact expression of (part of) what is understood by someone who understands e.g. 'red', for it states conditions recognition of which is sufficient to justify him in describing an object as 'red'; the statement of the rule is uninformative only in the sense that such a capacity of recognition cannot be imparted just by stating it.

So our picture is that correct use of language is essentially nothing other than use of it which conforms with a set of instructions, a set of semantic rules, which we have learned. Of course we handle language in general in a quite automatic way. But a chess player's recognition of the moves allowed for a piece in a certain position can be similarly automatic; it remains true that an account of his knowledge is to be given by reference to the rules of chess.

If language-mastery is thought of in such terms, the question arises, what means are allowable in the attempt to discover general features of the *substantial* rules for expressions in our language, the rules which determine specific senses? The view of the matter on which this paper centres is that here we may legitimately approach our use of language from within, i.e. reflectively as self-conscious masters of it, rather than externally, equipped only with behavioural notions. Thus it is legitimate to appeal to our conception of what justifies the application of a particular expression; to our conception of what we should count as an adequate explanation of the sense of a particular expression; to the limitations imposed by our senses and memories on the kind of instruction which we can actually carry out in practice; and to the kind of consequence which we associate with the application of a particular predicate, to what we think of as the point or interest of the distinction which the predicate implements. The primary concern of this paper is with the idea, henceforward referred to as the *governing view*, that from such considerations can be derived a reflective awareness of how we understand expressions in our language, and so of the nature of the rules which determine their correct use. The governing view, then, is a conjunction of two claims: that our use of language is rightly seen, like a game, as a practice in which the admissibility of a move is determined by rule, and that general properties of the rules may be discovered by means of the sorts of consideration just described. What I am going to argue is that these theses are mutually incoherent.

The difficulty has to do with the fact that the second thesis of the governing view, concerning the means whereby general features may be discovered of the semantic rules which we actually follow, forces us to recognize *semantic incoherence* in our understanding of a whole class of predicates—elements whose full exploitation would force the application of these expressions to situations in which we should otherwise regard them as not applying. The second thesis requires us to recognize rules which, when considered in conjunction with certain general features of the situations among which their associated expressions are to be applied, issue in contradictory instructions. Nevertheless we succeed in using these expressions informatively; and it seems that to use language informatively depends on using it, in large measure, consistently. It follows that our use of these expressions cannot correctly be pictured purely as the implementation of rules of the character which the second thesis yields for them; these rules cannot be implemented by any consistent pattern of behaviour. The governing view is therefore incoherent; for if its second thesis is true, the semantic rules governing certain predicates are capable by consistent beings only of selective implementation and thus, contrary to the first thesis, are not *constitutive* of what we count as the correct use of these expressions.

The predicates in question are all vague; but their vagueness is not just a matter of the existence of situations to which it is indeterminate whether or not they apply. Rather it is something which Frege, under the guise of a favourite metaphor, constantly runs together with possession of borderline-cases, viz., the idea of lacking 'sharp boundaries', of dividing logical space as a blurred shadow divides the background on which it is reflected. The conflation is natural because the figure equally exemplifies the idea of the borderline-case, a region falling neither in light nor shadow. But there is no clear reason why possession of borderline-cases should entail possession of blurred boundaries. If, following Frege, we assimilate a predicate to a *function* taking objects as arguments and yielding a truth value as value, then a predicate with borderline-cases may be seen simply as a partial such function—which is consistent with the existence of a perfectly sharp distinction between cases for which it is defined and cases for which it is not. Borderline-case vagueness of this straightforward kind presents no difficulty for the governing view; it is merely that there are situations to which no response in terms of a certain range of predicates is determined by their associated semantic rules as

correct. In contrast, if the second thesis of the governing view is correct, then predicates with 'blurred boundaries' are, in typical cases, rightly regarded as semantically incoherent.

This incoherence is implicit in the very nature of their vagueness. Vagueness is hardly ever, as Frege and Russell thought, merely a reflection of our not having bothered to make a predicate precise. Rather, the utility and point of the classifications expressed by many vague predicates would be frustrated if they were supplied with sharp boundaries. The sorts of argument allowed by the second thesis of the governing view will transpire to yield support for the idea that such predicates are essentially vague. The thesis equips us to argue that lack of sharp boundaries is not in general merely a superficial phenomenon, a reflection of a mere hiatus in some underlying set of semantic rules. In almost all the examples one comes across lack of sharp boundaries is not the consequence of an omission, but e.g. a product of the kind of task to which an expression is put, the kind of consequences which we attach to its application or, more deeply, the continuity of a world which we wish to describe in purely observational terms. Lack of sharp boundaries is a phenomenon of semantic depth. It is not usually a matter simply of our lacking an instruction where to 'draw the line'; rather the instructions we already have determine that the line is *not* to be drawn.

This conclusion might seem a welcome contribution to our understanding of the nature of vagueness, even from the standpoint of the governing view, were it not that it comes out in the form that no sharp distinction may be drawn between cases where it is definitely correct to apply such a predicate and cases of *any* other sort. But that is obviously a paradoxical concept. Thus it is that someone who espouses the governing view simply has no coherent approach to the Frege-Russell view of vagueness. His second thesis furnishes him with conclusive reasons to reject the suggestion that vagueness is a superficial, eliminable aspect of natural language with no real impact upon its informative use. But it does so in such a way that he is constrained to regard many vague predicates as semantically incoherent—specifically, as prone to the reasoning of the Sorites paradox—so that, unless the Frege-Russell view is right, he cannot maintain his first thesis with respect to such expressions. Only if their vagueness were an incidental feature could he maintain that the *essential* semantics of such expressions conformed to his first thesis.

II

Let us then consider some examples of the Sorites paradox in order to be clear how the governing view cuts off traditional lines of solution, indeed, all lines of solution. To begin with the classical case: if a pile of salt is large enough to be fairly described as a heap, the subtraction of a single grain of salt cannot make a relevant difference; if $n+1$ grains of salt constitute a heap, so do n grains. Thus one grain, and, indeed, zero grains constitute a heap. To block the paradox, it seems we have to be able to insist that, for some particular value of n, $n+1$ grains of salt would amount to a heap while n grains would not. But that is simply not the sense of 'heap'. Exact boundaries for the concept of a heap, either in terms of the precise number of grains contained or, indeed, in terms of any other precise measure, simply have not been fixed. But without such boundaries, a transition from $n+1$ grains to n grains can never be recognized as transforming a case where 'heap' applies into a case where it does not. Here we gravitate towards the idea that lack of exact boundaries is, as such, an essentially incoherent semantic feature.

A second example is given by Essenin-Volpin.[2] Consider the typical span of time between one human heartbeat and its successor. Then the concept of childhood—the sense of 'child'—is such that one does not, within a single heartbeat, pass from childhood to adolescence. Not that we are children for ever; but at least childhood does not evaporate between one pulse and the next. Similarly for the transition from infancy to childhood, and from adolescence to adulthood. 'Infant', 'child', 'adolescent', 'adult' are thus all semantically incoherent expressions; for the sense of each of these predicates is such that, in a typical process of growing-up, their correct application will always survive the transition from one heartbeat to its successor or to its predecessor. So, by appropriately many steps of *modus ponens*, we may force the application of each of these predicates to cases we should otherwise regard as falling within the domain of one of the others.

As a third example, consider a series of homogeneously coloured patches, ranging from a first, red patch to a final, orange one, such that each patch is *just* discriminable in colour from those immediately next to it, and is more similar in respect of colour to its immediate

[2] A. S. Essenin-Volpin, 'Le programme ultra-intuitioniste des fondements des mathématiques', in *Infinitistic Methods* (*Proceedings of the Symposium on Foundations of Mathematics, Warsaw, 2–9 September* 1959) (Pergamon Press, Oxford, 1961), p. 203.

neighbours than to any other patches in the series. That is, marginal changes of shade are involved in every transition from a patch to its successor, and each such transition carries us further from red and closer to orange. Now, the sense of colour predicates is such that their application always survives a very small change in shade. If one is content to call something 'red', one will still be so content if its colour changes by some just discriminable amount. There is a notion of a degree of change in respect of colour too small to amount to a change *of* colour. Only if a substantial difference comes between two patches of colour shall we consider ourselves justified in ascribing to them incompatible colour predicates.

This, obviously, is to attribute semantic incoherence to colour predicates. We have an easy proof that all the patches in the example are red, or that they are all orange, or that they are all doubtfully either. Moreover any two colours can be linked by such a series of samples; so any colour predicate can be exported into the domain of application of one of its rivals.

What is involved in treating these examples as genuinely paradoxical is a certain *tolerance* in the concepts which they respectively involve, a notion of a degree of change too small to make any difference, as it were. The paradoxical interpretations postulate degrees of change in point of size, maturity and colour which are insufficient to alter the justice with which some specific predicate of size, maturity or colour is applied. This is quite palpably an incoherent feature since, granted that any case to which such a predicate applies may be linked by a series of 'sufficiently small' changes with a case where it does not, it is inconsistent with there being any cases to which the predicate does not apply. More exactly, suppose ϕ to be a concept related to a predicate, F, as follows: that any object which F characterizes may be changed into one which it does not simply by sufficient change in respect of ϕ. Colour, for example, is such a concept for 'red', size for 'heap', degree of maturity for 'child', number of hairs for 'bald', etc. Then F is *tolerant* with respect to ϕ if there is also some positive degree of change in respect of ϕ insufficient ever to affect the justice with which F applies to a particular case.

In essentials, then, the Sorites paradox interprets certain vague predicates as tolerant. But this might seem a tendentious interpretation. Not that there is any doubt that the predicates in question do lack sharp boundaries; and the antiquity of the paradox bears witness to how easy it is to interpret this as involving the possession by these

predicates of a principle of re-application through marginal change. But is this a correct interpretation? Because 'heap' lacks sharp boundaries, it is plain that we are not entitled to single out any particular transition from n to $n-1$ grains of salt as being the decisive step in changing a heap into a non-heap; no one such step is decisive. That, however, is not to say that such a step always *preserves* application of the predicate. Would it not be better to assimilate the situation to that in which bordering states fail to agree upon a common frontier? Their failure to reach agreement does not vindicate the notion that e.g. a single pace in the direction of the other country always keeps one in the original country. For they have at least agreed that there is to be a border, that *some* such step is to be a decisive one; what they have not agreed is where. If we regard the predicates in the example in the terms of this model, we shall conclude that their vagueness is purely a reflection of our intellectual laziness. We have, as it were, decided that a disjunction is to be true—at some stage, n grains will be a heap where $n-1$ grains will not—without following up with a decision about *which* disjunct is true. On this view, the notion that these predicates are tolerant confuses a lack of instruction to count it the case that a proposition is false with the presence of an instruction to count it true. This conflation would be permissible only if the semantic rules for our language were in a certain sense complete, that is, if we possessed instructions for every conceivable situation. But for there to be vague expressions in our language is, on this view, precisely for this not to be so.

Someone who holds the governing view is bound to reject this suggestion as a deep misapprehension of the nature of the vagueness of these predicates. The lack of sharp boundaries possessed by these examples is correctly interpreted as tolerance, provided that we may discover elements of their senses in accordance with the second thesis. It would be inconsistent with elements already present in the semantics of these predicates so to refine their senses that the Sorites reasoning was blocked. How is this?

'Heap' is essentially a coarse predicate, whose application is a matter of rough and ready judgement. We should have no use for a precisely demarcated analogue in contexts in which the word is typically used. It would, for example, be ridiculous to force the question of obedience to the command, 'pour out a heap of sand here', to turn on a count of the grains. Our conception of the conditions which justify calling something a heap is such that the appropriateness

of the description will be unaffected by any change which cannot be detected by *casual observation.*

A different argument is available for supposing colour predicates tolerant with respect to marginal changes in shade. We learn and teach our basic colour vocabulary ostensively. Evidently it is a precondition of the feasibility of so doing that we can reasonably accurately remember how things look. Imagine someone who can recognize whether simultaneously presented objects match in colour, so that he is able to use a colour-chart, but who cannot in general remember shades of colour sufficiently well to be able to handle without a chart colour predicates for which we are able to dispense with charts. Such a person might, for example, be quite unable to judge whether something yellow, which he was shown earlier, would match the orange object now before him. Thus, for such a man, an ostensive definition of 'yellow' would be useless; in order to apply 'yellow' as we apply it, he would have to employ a chart. We, in contrast, are able to dispense with charts for the purpose of making distinctions of colour of the degree of refinement of 'yellow'. Any object to which a colour predicate of this degree of refinement definitely correctly applies may be recognized as such just on the basis of our ostensive training. Plainly, then, it has to be a feature of the senses thereby bestowed upon these predicates that changes too slight for us to remember—that is, a change such that exposure to an object both before the change is undergone and afterwards leaves us uncertain whether the object *has* changed, because we cannot remember sufficiently accurately how it was before—never transform a case to which such a predicate applies into one where such is not definitely the right description. The character of our basic colour training presupposes the *total memorability* of the distinctions expressed by our basic colour predicates; only if single, unmemorable changes of shade never affect the justice of a particular basic description can the senses of these predicates be explained entirely by methods reliant upon our capacity to remember how things look.

For the tolerance of 'child', etc., the governing view affords a third type of argument. The distinctions expressed by these predicates are of substantial social importance in terms of what we may appropriately expect from, and of, persons who exemplify them. Infants, for example, have rights but not duties, whereas of a child outside infancy we demand at least a rudimentary moral sense; we explain the anti-social behaviour of some adolescents in terms of their being adolescents; and

we make moral and other demands of character on adults which we would not impose on the immature. Plausibly, these predicates could not endure such treatment, were they not tolerant with respect to marginal changes in degree of maturity—certainly with respect to the changes involved in the transition from one heartbeat to the next. It is *ceteris paribus* irrational and unfair to base substantial distinctions of right and duty on marginal differences; if we are forced to do so, e.g. with electoral qualifications, it is with a sense of injustice. Moreover it is only if a *substantial* change is involved in the transition from childhood to adolescence that we can appeal to this transition to explain substantial alterations in patterns of behaviour. That predicates of degree of maturity should possess tolerance is a direct consequence of their social role; very small differences cannot be permitted to generate doubt about their application without correspondingly coming to be associated with a burden of moral and explanatory distinctions which they are too slight to carry.

On the second thesis of the governing view, then, our embarrassment about where to 'draw the line' with these examples is to be viewed as a consequence not of any hiatus in our semantic programme but of the tolerance of the predicates in question. If casual observation alone is to determine whether a predicate applies, then items not distinguished by casual observation must receive the same verdict.[3] So single changes too slight to be detected by casual observation cannot be permitted to generate doubt about the application of such a predicate. Similarly, if the conditions under which a predicate applies are to be generally memorable, it cannot be unseated by single changes too slight to be remembered. Finally, very slight changes cannot be permitted to generate doubt about the application of predicates of maturity without contravening their moral and explanatory role. The utility of 'heap', the memorability of the conditions under which something is 'red', the point of 'child' impose upon the semantics of these predicates tolerance with respect to marginal change in the various relevant respects.

To allow these considerations is to concede that the vagueness of these examples is a phenomenon of semantic depth—that it is sacrificed at much more than the cost of the intellectual labour of the stipulation —and that it is a structurally incoherent feature. Two things follow. First, there is no special logic for predicates of this sort, crystallizing

[3] Not that it has to be the case that a definitely correct verdict can always be reached; but if it cannot, that in turn must be the situation with respect to each item in question.

what is distinctive in their semantics in contrast with those of exact predicates; for what is so distinctive is their inconsistency. Second, the fashion in which we typically use these expressions needs some other model than the simple implementation of rules, if these rules are to incorporate all the features of their senses which we should wish to recognize on the basis of the second thesis.

III

There is a fourth, and more profound way in which tolerance, according to the second thesis, would seem to arise. Colour predicates will again serve as an illustration. Plausibly, these predicates are in the following sense purely *observational*: if it is possible to tell at all what colour something is, it can be told just by looking. The look of an object decides its colour, as the feel of an object decides its texture, or the sound of a note its pitch. The information of one or more senses is decisive of the applicability of an observational predicate; so a distinction exemplified by a pair of sensorily-equivalent items cannot be expressed solely by means of such predicates. What is about to be illustrated is a feature of any predicate whose sense is purely observational in the fashion just adumbrated.

If colour predicates are observational, any pair of patches indistinguishable in colour must satisfy the condition that any colour predicate applicable to either is applicable to both. Suppose, then, that we build up the series of colour patches of the third example, interposing new patches to the point where every patch in the resultant series is indiscriminable in colour from those immediately adjacent to it. The possibility of doing so, of course, depends upon the non-transitivity of our colour discriminations. The observationality of 'red' requires it to be tolerant with respect to the kind of change involved in passing from any patch in this series to an immediate neighbour, so we have a Sorites paradox. If 'red' is observational, its sense must be such that from the premisses, that x is red and that x looks just like y, it follows that y is red, no matter what objects x and y may be. Thus we are equipped to conclude that each successive patch in the series is red, given only the true premiss that the first patch is red.

The memorability, then, of the conditions of application of 'red' requires that it be tolerant with respect to changes which, under favourable circumstances, we can actually directly discern. Now, however, it appears that even if our memories were to be as finely discriminating as our senses, colour predicates and others would still

possess tolerance; only the changes which their application tolerated would not be changes which we could directly discern in objects which underwent them. This tolerance has nothing to do with the limitations of our memories; it is a consequence of the observationality of these predicates.

These considerations are broadly analogous to what was said of 'heap': if we so fix the sense of a predicate that whether it applies has to do with nothing other than how an object seems when casually observed, then changes other than such as can be determined by casual observation cannot transform a case to which the predicate applies into one to which there is some question whether it applies. The point remains good if we omit the word 'casual'. But this fourth example is prima facie deeper-reaching, at any rate for someone who, like Frege, believes that language should be purified of vague expressions. The cost of eliminating predicates of casual observation would be no more than convenience; to require, however, that language should contain no expression tolerant in the manner of the fourth example would be to require that it contained no expressions of strictly observational sense. If we stipulated away the tolerance of colour predicates, we should have to forgo our whole present idea of what justifies the application of these predicates, viz. the *look* of a thing. In general, there would be no predicate whose application to an object could be decided just on the basis of how it looked, felt, sounded, etc. Might there not then be a higher price to pay, namely the jeopardizing of contact between language and empirical reality?

We shall return to the last thought. First we require to see how the governing view sustains the idea that there is a large class of predicates whose senses are purely observational. If we are to understand the scope of the fourth example, we also require to know under what circumstances we may expect our sensory discriminations to be non-transitive.

That we do intuitively regard the semantics of colour predicates as purely observational is beyond doubt; and simply illustrated by the fact that we should regard it as a criterion of lack of understanding of such a predicate if someone was doubtful whether both of a pair of objects which he could not tell apart should receive the same description in terms of it. We regard it as a criterion of understanding such a predicate that someone, presented under suitable conditions with an object to which it applies, can tell that it does so just on the basis of the object's appearance. Certainly, then, our ordinary conception of

how to tell that a particular colour predicate applies, of what justifies its application, would involve that these predicates are purely observational.

In addition, it is plausible to suppose that any *ostensively definable* predicate must be observational. If an expression can be ostensively defined, it must be possible to draw to someone's attention those features in his experience which warrant its application; and if this is possible, there can be no question of the expression applying to some but not others among situations which he cannot distinguish in experience. It would be a poor joke on the recipient of an ostensive definition if the defined expression applied selectively among situations indistinguishable from one which was originally displayed to him as a paradigm.

In general the connection between an expression's being observational—its applying to both, if to either, of any pair of observationally indistinguishable situations—and its being ostensively definable is as follows. The picture of acquiring concepts by experience of cases where they do apply and cases where they do not—a picture which surely has *some* part to play in a philosophically adequate conception of the learning of a first language—cannot be wholly adequate for concepts which differentiate among situations which look, feel, taste, sound and smell exactly alike. So if that picture is wholly adequate for any concepts, they must be concepts whose range of application does not include situations which experience cannot distinguish from situations which may not definitely correctly be regarded as falling within that range. To master the sense of a predicate is, at least, to learn to differentiate cases to which it is right to apply it from cases of any other sort. If such mastery can be bestowed ostensively, a comparison of two such cases must always reveal a difference which sense-experience can detect. The notion that the whole range of application of a predicate can be made intelligible by ostensive means presupposes that it is never the case that only one of a pair of objects, which the senses cannot tell apart, is characterized by it.

This is a clear, absolutely general connection. If there is in the conditions of the correct application of a predicate nothing which is incapable of ostensive communication, then the predicate must apply to both, if to either, of any pair of indistinguishable objects. But it seems manifest that adjectives of colour, and many others, do precisely not involve any such further condition of correct application; on the contrary, ostensive training would appear fully determinant of

their meaning—or, if it is not, it is the only training which we get. The governing view thus vindicates the observationality of colour predicates twice over: as a consequence both of our general conception of what justifies their application, and of the character of the training in their use which we receive.

The other question was to do with the scope of the phenomenon of non-transitive indiscriminability. Suppose that we are to construct a series of colour patches, ranging from red through to orange, among which indiscriminability is to behave transitively. We are given a supply of appropriate patches from which to make selections, an initial red patch, and the instruction that each successive patch must either match its predecessor or be more like it than is any other patch not matching it which we later use. Under these conditions it is plain that we cannot generate any change in colour by selecting successive matching patches; if indiscriminability is to be transitive, then if each patch in the first n selections matches its predecessor, the nth selection must match the first patch. The only way to generate a change in colour will be to select a non-matching patch.

When the series is complete, how will it look in comparison with the series of the fourth example? It is clear that we shall have lost what was distinctive of that series: the appearance of *continuous* change from red to orange. In the new series the shades are exemplified in discrete bands, containing perhaps no more than one patch, and all the changes take place abruptly in a transition from a patch to its successor. It thus appears that, were our judgements of indiscriminability to be universally transitive among samples of homogeneous colour, no field of colour patches could be ordered in the distinctive fashion now possible: i.e., so as to give the impression of a perfectly smooth change of colour. If matching generally behaved transitively among shades, no series of colour patches could give the impression of continuous transformation of colour; by contraposition, then, for matching to function non-transitively among a finite set of colour patches, it is sufficient that they may be arranged so as to strike us as forming a phenomenal continuum. This reasoning may obviously be generalized. Any finite series of objects, none of which involves any apparent change in respect of ϕ, may give an overall impression of continuous change in respect of ϕ only if indiscriminability functions non-transitively among its members.

The reasoning may in fact be generalized further. It can be shown (cf. 'On the Coherence of Vague Predicates') that the non-transitivity

of our discriminations may be seen as a consequence of the continuity of change, viewed as a pervasive structural feature of our sense-experience. The general lesson of the fourth example is thus as follows. If we attempt to mark off regions of a seemingly continuous process of change in terms of predicates which are purely observational—predicates of which it is understood that ostensive definition gives their whole meaning—these expressions are bound to display tolerance in a suitable series of stages selected from the process. An analogue of the fourth example may thus in principle be constructed for any ostensively defined predicate; for absolutely anything which it characterizes might undergo seemingly continuous change to a point where it could be so characterized no longer. The fourth example indicates a basic fault, as it were, lying deep in the relation between the nature of our experience and those parts of language by means of which we attempt to give the most direct, non-theoretical expression to it.

This conclusion rests upon two assumptions: that it is right to regard the senses of colour predicates, etc., as purely observational; and that this is a very fundamental fact about their senses, whose sacrifice would be possible only at great cost. The governing view, as we have seen, yields the first assumption. For the second, however, no argument has so far been presented; I merely voiced concern that 'contact' between language and the empirical world might be attenuated if the use of purely observational predicates was abandoned. Before this concern is evaluated, and the general implications assessed of stipulating away the tolerance, and so the observationality, of the relevant predicates, we must consider a general objection to the way in which all four examples have been treated.

IV

If it is conceded that the vagueness of these examples is correctly interpreted as tolerance, then plainly no consistent logic does justice to the semantics of such predicates. It is natural to suggest, however, that the argument for this interpretation may have overlooked an essential feature of this sort of predicate: that they typically express distinctions of *degree*. There are degrees of redness, of childishness, and, if a smaller heap is regarded as less of a heap, of heaphood.

What is it for the distinction between being *F* and not being *F* to be one of degree? Typically, it is required that the comparatives, 'is less/more *F* than', are in use and that iteration of one of these relations

may transform something F into something not F, or vice versa. In addition, the semantic relations between the comparatives and the simple descriptions, 'is F' and 'is not F', are such that if a is less/more F than b, then the degree of justice with which a can be described simply as F is correlatively smaller or larger than that with which b can be so described. That is, a twofold classification of possible states of affairs into those which would justify the judgment, 'a is F', and those which would not, misses what is distinctive about the predicate whose application is a matter of degree. For that to be so is exactly for there to be *degrees* of such justice.

It is thus plausible to suppose that a logic for distinctions of this sort cannot be based upon simple bivalence. With such predicates there are, as it were, degrees of truth, whose collective structure is that of the set of degrees of being F. In this sense it is arguable that the examples do require a special, non-classical logic. But how did the earlier arguments for the tolerance of these predicates overlook that they expressed distinctions of degree?

The suggestion is that the paradoxical reasoning essentially depends upon the constraints of bivalence. Consider a pair of objects one of which, a, we are happy to describe as F, while b is slightly less F than a. How is b to be described? If our admissible descriptions are restricted to 'F' and 'not F', if we *have* to say one or the other, then presumably we shall describe b as F. For if something is more like something F than something not F, to describe it as F is the less misleading of the two alternatives. But the justification with which 'F' is applied in successive such cases successively decreases. We have no principle of the form: if a is F, and b differs sufficiently marginally from a, then b is F; with distinctions of degree there are no 'small changes insufficient to affect the justice with which a predicate applies'; they are, on the contrary, small changes *in* the degree of justice with which the predicate may be applied. Of course, we do have the principle: if the judgment that a is F is justified to some large degree, and b is marginally less F than a, then the description of b as F will be better justified than its description as not F. But that is not a paradoxical principle.

Anyone who thinks he here feels the cool wind of sanity fanning his brow would do well to be clear why we do not still have *this* principle: if b is marginally less F than a, then if the less misleading description of a is 'F', the less misleading description of b is 'F'. Yet if this principle is false, there must in any Sorites-type series be a last case of

which we are prepared to say that if we *had* to describe it either as *F* or as not *F*, the better description would be '*F*'. Why, then, is it usually embarrassing to be asked to identify such a case without any sense of arbitrariness?

Let us assign to '*a* is *F*' a *designated* value just in case '*F*' is a less misleading description of *a* than 'not *F*'. Then our embarrassment is exactly to identify a last object to which the application of '*F*' would receive a designated value. But now the suspicion arises that tolerance is with us still; only it is no longer the *truth* of the application of '*F*' that would survive small changes but its designatedness.

Is this suspicion justified? One thing is clearly correct about the assumption of bivalence: faced with a situation and a predicate, we have only two choices—to apply or to withhold. There is not a series of distinct linguistic acts in which we can reflect every degree of justification with which a predicate may be applied. The crucial practical notion to be mastered for a predicate associated with the distinction of degree is thus that of a situation to which the application of the predicate is *on balance* justified. Without mastery of this notion no amount of information about the structure of variations in the degree with which '*F*' applies entails how the predicate is to be used. Now of this notion may it not still be a feature that it always survives sufficiently small changes?—that if *a* and *b* are dissimilar only to some very small extent, then if describing *a* as *F* is on balance justified, so is thus describing *b*?

It is clear that all the previous considerations will apply, and that the introduction of a complex structure of degrees of justification will get us no farther. For among these degrees we have still to distinguish those with which for practical purposes the application of the predicate is to be associated; otherwise we have not in repudiating bivalence done anything to replace the old connection between justified assertion and truth. But plainly, once we attempt to make such a distinction, the arguments afforded by the governing view sweep aside this proposed solution to the Sorites paradox as an irrelevance. To rehearse the reasons: if we are to be able to *remember* how to apply '*F*', then differences too slight to be remembered cannot transform a situation to which its application is on balance justified into one which is not so; if we are to be able to apply '*F*' just on the basis of *casual observation*, the same applies to differences too subtle to be detected by casual observation; if the distinction between cases to which the application of '*F*' is on balance justified and others is to be made just on the basis of

how things look, or sound, etc., then any pair of indistinguishable situations must receive the same verdict; finally, if '*F*' is associated with moral or explanatory distinctions which we are unwilling to tie to very small changes, we shall likewise be unwilling to allow such changes to generate doubt about the status of a situation previously regarded as on balance justifying description as *F*. Of course the use here being made of the notion of a situation to which the application of '*F*' is 'on balance' justified is quite uncritical. But this is legitimate. For, as remarked, there must be *some* such notion if a many-valued logic for distinctions of degree is to have any practical linguistic application.

V

Let us turn then to the question whether we could not eliminate, at not too heavy a cost, the tolerance of observational predicates. The resulting predicates would no longer be strictly observational; hence the initial doubt whether such a purified language could engage with the observational world at all. On reflection, though, it is clear that the dislocation of language and the world of appearance generated by such a purification would not have to be as radical as that. When three situations collectively provide a counter-example to the transitivity of indiscriminability, there is nothing occult in the circumstance that they do so. It is an observationally detectable difference between indiscriminable situations that one is distinguishable from a third situation from which the other is not; the relation, '*a* matches *b* matches *c* does not match *a*', is an *observational relation*, i.e. one whose application to a trio of objects can be determined just by looking at them, listening to them, etc.

Observational concepts evidently require narrower criteria of re-application than indistinguishability, if they are to be purified of tolerance. But we should not jump to the conclusion that to provide such criteria will require surrender of observationality altogether, for the phenomenon which is causing the trouble is itself observational. Indeed, the *only* kind of observationally detectable difference which there can be between indiscriminable items is that one should be distinguishable from some third item from which the other is not. So if the class of expressions in question is to remain in contact with observation, we have to look for some form of stipulation which *exploits* the non-transitivity of indistinguishability to provide a basis for describing indiscriminable situations differently. No other

explanation can correspond to a distinction which sense-experience can determine to obtain, a distinction which we can simply be shown.

After such a stipulation, the question whether a pair of indiscriminable colour patches should receive the same colour description may turn on their respective relations of indiscriminability/discriminability with respect to some third patch. But now we have to take note of a striking aspect of the philosophical psychology of non-transitive matching: it does not seem to be possible to conduct experiments with non-transitively matching triads in *memory*. For suppose that a predicate, *F*, is defined ostensively by reference to some individual, *a*, which, it is noted at the time, perfectly matches another individual, *c*; it is understood that *F* is not to be applied to individuals which match *a* unless they also match *c*. Later the trainee comes across *b* which, so far as he can determine, matches *a* perfectly; the question is, does *b* match *c*? It is evident that the issue is only resoluble by direct comparison, and especially that it cannot be settled by memory, however accurate. For the most perfect memory of *c* can give no further information than that it looked just like *a*; which, when non-transitive matching is a possibility, is simply insufficient to determine whether it would match *b*. This, it must be emphasized, in contrast with our conclusions concerning the third example, is *not* a limitation imposed by the feebleness of our memories; it is a limitation of principle.

It thus appears that if we are to be able to exercise expressions whose application to matching individuals depends upon their behaviour in relation to a third, possibly differentiating individual, then we have to be able to ensure the *availability* of the third individual. Expressions of this species will be practicably applicable only in relation to a system of paradigms. So we can see, even in advance of attempting a specific stipulation to remove the tolerance of 'red' as displayed in the fourth example, that the kind of semantic construction it will have to be is going to tie the application of expressions of colour to the use of a colour-chart.

Let us then consider, as a test case, how we might go about the construction of such a chart. What we require of the chart is that it should enable us to identify a last red patch in any series of the type of the fourth example. There is one obvious way to achieve this, namely to devise a single ad hoc paradigm. It is plausible to suppose that we could complete a colour-chart for the red/orange region at least in the sense that anything which we should wish to regard as falling within

that region would match something on the chart. Consider, then, an arrangement of colour patches which form in this sense a complete colour-chart for the red/orange region and which are simply ordered by similarity, i.e. every patch on the chart more closely resembles its immediate neighbours than any other patches on the chart. Then a sharp red/orange distinction can be generated as follows. Select some patch towards the middle of the chart; then any colour patch matching something on the chart either matches the selected patch or it does not; if it does, it is red; if it does not, but matches a sample to the left of the selected patch, it is again red; otherwise, it is orange.

Naturally it could not be guaranteed that duplicates of this chart would always deliver the same verdict. Charts could look absolutely similar, and even satisfy the condition that the *n*th sample on either matched and was distinguishable from exactly the same samples on the other chart as its own *n*th sample, yet deliver discrepant results. But they would not often do so. Besides, the situation is not novel. Rulers, for example, sometimes give different results. A final criterion for one system is deposited in Paris; and we could do the same with a colour-chart.

Generalized, this proposal might seem quite ludicrous in practical terms. We are confronted with the spectacle of a people quite lost without their individual wheelbarrow loads of charts, tape-recordings, smell- and taste-samples and assorted sample surfaces. But this caricatures the proposal. There would be no need for all this portable semantic hardware. This is clear if we pursue the analogy with the use of rulers: it is true that, without a ruler, we can only guess at length; but after the introduction of an ad hoc paradigm for colours, the use of colour predicates will presumably be analogous not to that of expressions like 'two feet long', but rather to that of expressions like 'less than two feet long', i.e. expressions of a *range* of lengths. Of such expressions the criterion of application is still measurement; but unless the example is a peripheral one, we can tell *without* measuring what the outcome of measurement would be. Training in the use of paradigms might be essential if one is to grasp the sense of such expressions; but, once grasped, most cases of practical application could be decided without the use of paradigms—for most practical purposes, the wheelbarrow could be left behind.

It would appear, then, that if we adopted such stipulations as a general strategy, it would not have to affect our use of observational language very much at all. At present we can tell of anything red that

it is so just by looking at it. This would still usually be true after the proposed stipulation; and if the new distinction was suitably located, cases where it was not true would in general coincide with borderline-cases of the old red/orange distinction. The use of predicates so refined could thus greatly resemble their present use; the distinctions which they expressed would be empirically decidable; and there would be one crucial disanalogy—they would be tolerance-free.

It is apparent that exactly parallel considerations may be brought to bear upon the earlier treatment of the first and third examples. Even after a precise re-definition of 'heap', we would be able to learn to tell in most cases just by casual observation what verdict the new criterion would give if applied; it would seldom be necessary actually to count the grains. And the distinction between red and orange, supposing an exact distinction were drawn by means of a chart, would be unmemorable only within that small range of shades which could not by unaided memory be distinguished from the last red sample. It would thus appear that the cost of eliminating tolerance in cases of these two types need not after all be high, since we could expect to be able to tell in general just by looking at, etc., an item on which side of the dividing line it would fall.

If there need, after all, be no substantial sacrifice in endowing formerly observational predicates with exact boundaries, what has become of the alleged profound tension between phenomenal continuity and language designed to express how things seem to us? The answer, of course, is that it has simply been swept under the carpet. The possibility of our dispensing with paradigms for most practical purposes depends upon our capacity e.g. to distinguish between cases where we could tell whether or not 'red' applied just by looking and cases where we could not, where we should have recourse to a chart. But if we are able to make such a distinction, there can be no objection to introducing a predicate to express it. And then, it seems, the semantics of *this* predicate will have to be observational. On what other basis should we decide whether something looks as though comparison with a chart would determine it to be red than how it looks? Of any pair of colour patches which look exactly alike, if either looks as though the chart would deliver the verdict 'red', both must. So the new predicate, introduced to reflect our capacity to make this distinction, will be applicable to both members of any pair of matching colour-samples, if to either.

It is not that there is any compelling reason to have such a predicate;

only that there is no reason not to. If we were sometimes able to tell without using a chart whether something was red, it would surely be possible to make intelligible to us a predicate designed to apply in just such circumstances. So it transpires that a language all of whose observational concepts were based on paradigms would avoid containing tolerant predicates only by not containing means of expression of all the observational distinctions which we are in fact able to make. The dispensability of the wheelbarrow requires the exercise of observational concepts.

It would of course be absurd to propose that the tolerance of such new predicates—'looks as though it would lie to the left of the last red shade', 'looks as though it contains fewer than ten thousand grains', etc.—might in turn be stipulated away. Their meaning will not permit it; it cannot be allowed of things which look exactly alike that one may look as though it satisfies some condition which the other looks as though it does not, unless how a thing looks may not be determined by looking. The earlier discussion of the first and third examples involved an over-estimation of our interest in preserving the tolerance of the predicates involved only if we possess a coherent understanding of these new predicates; if the first and third examples do not, after all, pose a substantial problem for the governing view, it is because of our capacity to handle expressions falling within the scope of the fourth example.

VI

Let us, then, finally review the character of the difficulty for the governing view which originates in the fourth example.

It is a fundamental fact about us that we can learn to classify items according to their appearance, that we are able, consistently as it seems to us, to apply or to withhold descriptions just on the basis of how things strike the senses. In a discrete phenomenal world, there would be no special difficulty—no difficulty not inherent in the idea of a semantic rule as such—in viewing our use of such expressions as essentially nothing but the following of rules of which it was a consequence that indiscriminable phenomena should receive the same description. But if mutually exclusive use is made of a pair of such predicates, and if cases to which one applies permit of continuous transformation into cases where the other applies, it cannot be correct to represent the use made of either predicate just as the doing of what

is required by a set of rules with such a consequence. Yet we are forced—if the relevance is allowed of considerations to do with what we should regard as adequate explanation of such expressions, or with certain criteria which we should accept of misunderstanding such an expression—to attribute to the rules governing these predicates precisely such an implication; and *all* the phenomena which we confront in our world are in principle capable of seemingly continuous variation.

It will not quite do, though, to present the difficulty as that of the inadequacy of any inconsistent set of rules to explain a consistent pattern of behaviour. To begin with, it is unclear how far our use of e.g. the vocabulary of colours *is* consistent. The descriptions given of awkward cases may vary from occasion to occasion. Besides that, the notion of using a predicate consistently would appear to require some objective criteria for variation in relevant respects among items to be described in terms of it; but what is distinctive about observational predicates is exactly the lack of such criteria. So we may not lean too heavily, as though it were a matter of hard fact, upon the consistency of our employment of colour predicates. The point rather has to do with the fact that our use of these predicates is largely *successful*; the expectations which we form on the basis of others' ascriptions of colour are not usually disappointed. Agreement is generally possible about how colours are to be described; which is equivalent to saying that others *seem* to use colour predicates in a largely consistent way.

It is this fact of which the governing view can provide no account. A semantic rule is supposed to contribute towards determining what is an admissible use of its associated expression. The picture evoked by the first thesis of the governing view is that there is, for any particular expression in the language, a set of such rules *completely* determinant of when the expression is used correctly; such a set thus provides a model of the information of which a master of the use of the expression may be deemed to be in possession. Clearly, however, the feasibility of such a picture requires that the rules associated with an expression, about whose use we generally agree, be consistent. For if they issue conflicting verdicts upon the correctness of a particular application of the expression, it cannot be explained just by appeal to the rules why we agree that the application is e.g. correct.

The problem presented for the first thesis by the occurrence of tolerant predicates, or of any kind of semantically incoherent expression, is not that, in a clear-cut way, nothing can be done to implement

an inconsistent set of instructions. Strictly, of course, anything that is done will conflict with a part of them. But we can imagine a game whose rules conflict, but which is nevertheless regularly and enjoyably played to a conclusion by members of some community because, for perhaps quite fortuitous reasons, whenever an occasion arises to appeal to the rules, the players concur about which element in the rules is to be appealed to, so that an impasse never comes about. We need not enquire whether they have noticed the inconsistency in the rules. The point of the analogy is that in practice they always agree whether a move is admissible, as we generally agree whether something is red. The analogue of the first thesis in relation to this example is the notion that the rules completely determine when a particular move is admissible. But while it may be true that the authority of the rules can be cited for any of the moves the community actually make, it is plain that the rules alone do not provide a satisfactory account of the practice of the game. For someone could master the rules yet still not be able to join in the game, because he was unable to guess what sort of eclectic application of them an opponent was likely to make in relation to any given move.

An outsider attempting to grasp our use of a tolerant predicate would presumably not encounter exactly this difficulty; it would be clear that we were not prepared to allow remote consequences of its tolerance, inferred by means of reasoning of the Sorites type. The difficulty of principle for the first thesis, however, is the same. The rules of the game cannot provide an account of how the game is played, for it is possible that someone might grasp them yet be unable to participate. The semantic rules for an expression are supposed to provide an account of its correct use; they cannot do so if someone whose use of it differed radically from ours could still be thought of as in possession of exactly the same brief—as he can be, if it consists in an inconsistent set of instructions.

What, then, have we learned? The comparison of language with a game is an extremely natural one. What better explanation could there be of our ability to agree in our use of language than if, as in a game, we are playing by the same rules? We are thus attracted towards the assimilation of our situation to that of people to whom the practice of a highly ramified, complex game has been handed down via many generations, but of which the theory has been lost. Our central task, as philosophers of language, is to work towards the recovery of such a theory: a theory which will explain the mechanism

of our recognition of the senses of new complex expressions by displaying them as functions of the senses of their constituents and their mode of combination, which will explicate our apprehension of valid inferences—which, in short, will explicate the overall character of our mastery of the language game. What we have learned is that we probably cannot combine this conception of what a theory of meaning should accomplish with the notion that the investigation is something which, as masters of the language in question, we are better placed to carry out than an observer of our practice. We have to avoid appeal, at any rate, to a range of considerations which it is our antecedent prejudice to consider must be relevant: considerations to do with what we should deem a proper explanation of the sense of an expression, the criteria which we should employ for determining that someone misunderstands it, what we use the expression for, i.e., what issues turn on its application, the limitations imposed by our senses and memories on the information which we can absorb from our linguistic training, and our general conception of what justifies the application of the expression. And what privilege do we enjoy in the quest for a theory of meaning which an observer of our usage does not, if all these traditionally accepted guidelines for sense are dismissed? But we have seen that we must dismiss them if we want a coherent account of the senses of vague expressions. The methodological approach to *these* expressions, at any rate, must be more purely behaviouristic and anti-reflective, if a general theory of meaning is to be possible at all.