# VAGUENESS, RESPONSE DEPENDENCE
# AND RULE-FOLLOWING: SOME REFLECTIONS

## *Crispin Wright*

1. Vagueness gives rise to some of the most perplexing philosophical problems and paradoxes. Frege, as is well known, came to be thoroughly mistrustful of the phenomenon, supposing that a language fit for the purpose of articulating scientific and mathematical knowledge would have to be purified of it. There is little explicit argument for this view in his writings, but what there is gives expression to the idea that vagueness somehow threatens the stability of basic logic. One might expect that he would have cited the Sorites paradox in that connection, but the little he actually says is indicative of a rather less obvious thought. Frege takes it that if a term is not sharply defined, that will be tantamount to its having a limited range of significance: if a predicate is only partially defined, then it will only be of things of a certain kind that it will make sense to say that it either applies to them or that it does not. But that will give rise to failures not just of excluded middle, but of other laws too – contraposition, for example. Everything to which the predicate applies will be a thing of the presupposed kind; but it will not be correct to affirm, conversely, that everything not of that kind is something to which the predicate doesn't apply – since the range of cases of which the predicate may significantly be denied is likewise restricted to the kind of thing in question[1].

[1]  I take this to be the argument of *Grundgesetze*, vol. I, §65. Frege does mention the Heap itself at *Begriffschrift* § 27, in the context of his definition of what it is for a property to be hereditary in a series, but remarks on no threat to logic, suggesting merely that, since 'heap' is not everywhere sharply defined, we may regard the major premise as

This train of thought is interesting in two respects. The first is Frege's equation of an expression's being merely *partially defined* with its having a *limited range of significance* – a limited range of cases where it so much as makes sense to enquire whether it applies or not. This equation should be contested. Consider Tim Williamson's example: the predicate, 'dommal', whose satisfaction conditions are stipulated as met by a creature which is a dog, and failed by a creature which is not a mammal[2]. Nothing else is said by way of determination of its meaning. So the effect is that it is undefined whether 'dommal' applies to mammals other than canines. But it seems wrong to represent that as a situation in which the term 'dommal' has been given a limited range of significance in such a way that the validity of contraposition, and other related laws, has thereby been put in jeopardy. That would require that there now be some kind of thing, G, such that one of (a) 'All dommals are Gs' or (b) 'All G's are dommals' holds good in virtue of the partial definition while its contraposition is unassertible precisely because the definition is partial. But what selection for G makes the point? Not *dog*, since both (a) and its contraposition are then unassertible, while both (b) and its contraposition are true. Not *either a dog or a non-mammal*, since then again both (a) and its contraposition are unassertible, while both (b) and its contraposition are false. There is no other obvious candidate. Limitation on the range of cases to which an expression is determined as either truly or falsely applicable and limitation on the range of cases to which it may be significantly applied or withheld are simply distinct forms of limitation, which Frege would appear to have confused.

The second interesting aspect of Frege's line of thought may also be confused. But if – as I believe – it is, it is a much more seductive

indeterminate. (That is no option of course if we take it that the minor premise, that 0 grains of sand cannot constitute a heap, and the conclusion, that – say – 200,000 grains of sand cannot constitute a heap, are, respectively, true and false. An indeterminate statement cannot be inconsistent with the facts! Even Frege, it seems, underestimated the Sorites.)

[2]    Timothy Williamson *Identity and Discrimination*, (Oxford: Blackwell 1990) at p. 107.

and important confusion – one which has, indeed, conditioned virtually all mainstream work on vagueness until quite recently. It is the idea that vagueness – I mean vagueness of the kind which gives rise to the Sorites paradoxes – is a *semantic* feature of expressions which possess it, usefully comparable to the kind of semantic indeterminacy artificially generated in the example of 'dommal'. In that example, we are given rules for the application of 'dommal' and rules for denying it, which between them leave space for a *gap* – a range of cases where the rules simply don't give us any instruction what to do. The thought about vagueness, of which Frege's response is a prototype, is that vagueness in natural language is in general the naturally occurring counterpart of the artificially generated vagueness of 'dommal'. Our training gives us rules for the application of 'bald' – roughly, looking sufficiently like certain paradigms – rules for applying 'not bald' – again, looking sufficiently like certain paradigms – and these rules once again fail between them to cater for all cases. The exceptions – the borderline cases – are those for which we lack any sufficient instruction. They stand to 'bald' essentially as non-canine mammals stand to 'dommal'.

This way of looking at the matter may actually have implicitly originated with Frege, or it may be much older. In any case it is, undeniably, extremely natural. To be sure, not all writers about vagueness nowadays think of it as a semantic phenomenon at all, but among the majority who do, the Fregean conception of vagueness as a kind of semantic incompleteness is entrenched. It is, for example, a presupposition of the idea that vague expressions allow of a variety of alternative but admissible sharpenings, whose effect will be to include or exclude certain items from their range of application whose status was previously indeterminate. This idea in turn, of course, is presupposed by the widely accepted supervaluational approaches to the semantics and logic of vague discourse.

Natural as it may be, however, the Fregean conception seems to me to be open to serious objection on at least two major counts. First, it is badly mispredictive. It gives the wrong prediction about our responses to – and responses to responses to – borderline cases of standard Sorites-prone predicates. Someone who has mastered

'dommal' will know better than to apply it, or its contrary, to non-canine mammals. Asked if a cat is a 'dommal', he will say that he is not empowered to judge – for all we have to go on is a sufficient condition for being a 'dommal' and a necessary condition for being a 'dommal', and cats neither pass the first nor fail the second. By contrast, what we find in borderline cases of the distinction between, say, men who are bald and men who aren't it is not a recognition that there is no competent verdict to return but rather a phenomenon, spreading both among the opinions of normally competent judges and, across time, among the opinions of single competent judge, of weak but conflicting opinions, unstable opinions and – between different judges – agreement to differ. It is true that sometimes a competent judge may simply be unable to come to a view but it is not a characteristic of the borderline case region that it comprises just – or even any – cases where competent judges *agree* in failing to come to a view; and any case about which a competent judge fails to come to a view may, without compromising his competence, provoke a (weak) positive or negative response from him on another occasion. Moreover, failure to come to a view is not the same as the judgement that there is no competent view to take – and it is the latter that is appropriately made of a cat by someone competent with the use of 'dommal'.

These considerations are radically at odds with what the Fregean conception would lead one to expect. If someone takes the view that some particular cat is a dommal, then, ceteris paribus, they show that they have misunderstood the explanation of the word. If someone understands the explanation properly, they won't return a verdict about a cat. In contrast, our responses to those who do return verdicts in the borderline area of 'bald' is that, provided those verdicts are suitably sensitive and qualified, it is permissible so to do. We are *liberal* about judgements in borderline cases. Our thought is not that they are cases about which one ought to have no view but rather merely that they are cases about which it is probably pointless to work through differences of opinion. The psychologist testing the responses of a variety of, by normal criteria, competent subjects down a Sorites series would *expect* divergences in the borderline area – indeed, that is what the borderline area *is*: an area of expectable

and admissible divergence. If the 'dommal' model were correct, the expectation would be of consensual silence.

The second major area of difficulty for the Fregean conception concerns one of the most disconcerting phenomena associated with the kind of vagueness that interests us, that of so called *higher order* vagueness. There are various ways of eliciting the impression that the phenomenon is real and needs to be reckoned with. One line of thought is Dummett's[3]: if I was to introduce a new word – say 'sparsey' – to apply to just to borderline bald men, we would find that the boundary between the bald men and the sparsey men was itself vague. Another line of thought is to observe that, in a typical Sorites series, there will be no determinate first case of which we are content to judge that something other than the initial positive verdict is appropriate. How is this to be explained under the aegis of Fregean conception? According to the Fregean conception, borderline cases are cases where we have not provided for a verdict – cases which we failed to cover by the relevant semantic rules; so the remaining cases are ones where we *have* so provided and a negative or positive version is appropriate. How can *this* distinction in turn be one for the drawing of which we could have made insufficient provision? No doubt it is up to us what provision we've made – we might for example have provided even less for 'dommal', restricting the sufficient condition to Corgis. But whatever provision we've made, it shouldn't then need a *further* provision to settle which cases it does and doesn't cover respectively. Giving just the provision that he did, Williamson thereby settled that the borderline cases of 'dommal' comprise all non-canine mammals. He didn't merely settle that *some* kinds of case are to be borderline for 'dommal', leaving it open whether others are borderline or are cases where a determinate verdict is mandated. That matter was determined for him by the nature of his omission. It does not need further determination by him – indeed, it's not for him to stipulate at all – how far the omission extends.

[3] In "Wittgenstein's Philosophy of Mathematics" *Philosophical Review* LXVIII (1959), 324-48, (reprinted in M. Dummett *Truth and Other Enigmas* (London: Duckworth 1978); the cited line of thought occurs at p. 182).

A defender of the Fregean conception has possible responses, of course. It is intelligible, for instance, how an expression with a semantic architecture like that of 'dommal' might nevertheless allow vagueness on the boundary between the things which satisfied its sufficient condition and the things that neither satisfied its sufficient condition nor failed its necessary one. These, second order borderline cases, will precisely be things for which it was indeterminate whether or not they satisfied the relevant sufficient condition. Such indeterminacy might arise because the concept giving that sufficient condition – *dog* in the Williamson example – might itself have the same kind of semantic architecture. Similar possibilities would apply to the concept – *mammal* – which gave the original necessary condition. Since there seems to be no limit to the extent to which the pattern might be iterated, it looks as though the Fregean conception might after all be able to recover some of our preconceptions about higher order vagueness, to whatever extent they run, provided the relevant concepts that would be successively invoked were all of the illustrated semantic-architectural kind.

That is a proposal. But it seems to have very little mileage in it. When we reflect on the prototypical Sorites-prone predicates – predicates like 'red', 'bald', 'heap,' and so on – the most salient feature about them is their immediacy: it simply isn't credible that our conception of their conditions of proper application is informed by an indefinitely extending structure of partial definitions, each one deploying novel concepts distinct from those employed in the sufficient, or necessary, conditions articulated by its predecessor.

A yet more serious problem for the Fregean conception posed by higher-order vagueness is, as I have stressed in other work, that there is an overarching formal incoherence to reckon with[4]. The borderline cases of 'dommal' are cases where there is mandate neither to apply 'dommal' nor to apply is contrary. Borderline cases of borderline cases, on this model, will thus be cases where there is neither mandate to apply either a predicate or its contrary, nor mandate to

[4] See C. Wright, "Vagueness: a Fifth Column Approach" in J.C. Beall and Michael Glanzberg (eds.), *Liars & Heaps: New Essays on Paradox* (Oxford: Oxford University Press 2003) at p. 89.

regard them as borderline – to regard them as cases where neither the predicate nor its contrary is mandated to apply. So they are cases where there is no mandate to apply the original predicate (otherwise they wouldn't be borderline cases at all), no mandate to apply its contrary (for the same reason), but also no mandate to characterise as cases where there is neither mandate to apply the predicate nor to apply its contrary. But that's incoherent: whenever – as the first two conjuncts say – a case is such that there is no mandate to apply the predicate to it and no mandate to apply its contrary, then it will be *true* – and hence mandated – to say just that; but that is what the third conjunct denies.

The Fregean conception, then, is in tension with our actual sometimes positive or negative reactions to borderline cases, in tension with the liberality of our reactions to others' reactions to them, and in tension with the – at least apparent – phenomenon of higher order vagueness. It may be rejoined in mitigation that we should by now have learned to expect that *any* possible broad conception of vagueness will seem to be in tension with one or other of our preconceptions about the matter, or with other aspects of the phenomenology or linguistic practices surrounding the use of vague expressions. That, indeed, is what makes the whole issue so hard. Even if resolute prosecution of the Fregean conception is going to require a theory which is indeed in tension with certain aspects of the apparent data, that – it may be said – should be a decisive consideration against it only if some other theory can match the available advantages of such a theory, including those made possible by the apparatus of supervaluation, whilst causing fewer, or less serious casualties. But I do not buy that. I shall persist in the assumption that Fregean conception is mistaken: borderline cases of vague expressions of the type typified by the 'usual suspects' are not to be conceived in terms of the idea of, as it were, one's semantic instructions giving out, of the rules of language failing to provide guidance. But if they are not to be so conceived, what are the alternatives?

II. One thought is that the rules do indeed provide guidance right through a Sorites series, rather as a powerful river current guides items of flotsam over the waterfall – and that the deterioration of our

linguistic practice from broad consensus into paralysis and conflict is a function of our realisation that it is to disaster that the rules are taking us. Michael Dummett comes close to this – sometimes termed *nihilist* – proposal, in 'Wang's paradox'³. But actually it is, again, a poor explanation of our characteristic patterns of reaction in the borderline area. It's a poor explanation of our patterns of response to borderline cases – weak, defeasible opinions, conflicting opinions, unstable opinions, and failures to come to an opinion – for the simple reason that it doesn't need the context of a Sorites paradox to elicit these characteristic responses: they are elicited anyway by confrontation with borderline cases, from judges who have no inkling of the Sorites or even – in the case of, say young children – any capacity to follow the reasoning of the paradox and see its point.

More fundamentally, the suggestion that in our linguistic practice with vague expressions we follow incoherent rules – rules that do indeed actually mandate the application both of an affected predicate and of its contrary to the very same object – is powerless to explain the basic point that, when we do confront a Sorites paradox, we have *not the slightest inclination* to weigh the two limbs of the contradiction equally. There is not the slightest inclination to regard the verdict reached by the Sorites chain as correct. It is utterly dominated by the verdict with which it conflicts, and the paradox initially impresses as the merest trick. By contrast, where we really do have conflict in the rules governing a concept – for example, the concept of *course of values* introduced by Basic Law V of Frege's *Grundgezetse* – the two components of the paradox are balanced in our esteem, and we have not the slightest sense that one is to be preferred to the other. This is a basic datum which any satisfactory account of vagueness should accommodate and explain, but which – at least on the face of it – the nihilist view is powerless to explain.

I do not, by these remarks, mean to deny that we can make any kind of sense of an in some respects successful practice being informed by inconsistent rules. It might happen, for example, that a

³ Michael Dummett, "Wang's Paradox", *Synthese*, 30 (1975): 301-324.

society lays down rules of division which are correct except for the detail that they fail explicitly to exclude division by zero, but that they get away with it because it never occurs to anyone to attempt a calculation involving a ratio with zero as its denominator. Or it may be that practitioners are aware of the contradiction, but don't exploit it and manage to avoid situations in which it matters. In such cases, though, the presence of the contradiction in the rules is a matter of the way they are *explicitly* codified, and it wouldn't otherwise be motivated, just on the basis of the character of the informed practice, to propose a theory which represented its rules as inconsistent. The point is surely equally good for our pervasive and remarkably successful linguistic commerce involving vague expressions. To represent it as the product of inconsistent rules purely on the basis of a paradox which no-body actually accepts is merely bad empirical linguistics.

III. If the borderline cases of an expression are conceived neither as cases for which its governing rules fail to prescribe any verdict, nor as cases where they prescribe conflicting verdicts, then there seems to be no alternative but to accept both that verdicts are indeed prescribed in every case, and that they are consistent. In a Sorites series – at least any Sorites series of the normal 'monotonic' kind – that is tantamount to an acceptance that the rules governing the affected predicate mandate a sharp cut-off. Bivalence is therefore assured, and with it classical logic. What is missing, though, is any account of what vagueness is or why it arises.

As we know from the work of Williamson and others[6], this can be made to be a much more resilient proposal than it first appears. Its proponent – the *epistemicist* – is thinking of our use of vague expressions as broadly comparable to the sorts of judgements we should make about the application of precise predicates like 'is more than six feet tall' if excluded from reliance on canonical means of

[6] The leading systematic account of the epistemicist view is Timothy Williamson's *Vagueness*, London, Routledge, 1994. See also ch. 6 of Roy Sorensen's *Blindspots*, Oxford, Oxford University Press, 1988.

measurement and forced to estimate by eye. There would be cases where it would be clear that a subject was more than six feet tall, cases where it was clear that he was not, and cases where observation left one unable to form a view, or where our views were weak and unstable. This model thus predicts something very like our actual patterns of application of vague expressions in borderline cases. Presumably these patterns of reaction would be picked up by someone who was trained in the use of 'is more than six feet tall' purely by immersion in the practice of those restricted to visual estimation and to whom no communication was made of the real meaning of the predicate, in virtue of which its extension is actually sharp. We can construct a fantasy around that idea. Call those privy to the real meaning the *Priests* and those kept in the dark the *Accolytes*. We might imagine that, as time goes by, the Priests die out but the practice of the Accolytes survives in more or less stable form, though uninformed by any adequate conception of what determines the extension of 'lessonsicksfittle' as the phrase begins to be written in their dialect.

I do not know if Williamson, or any other defender of the epistemic conception, would take any comfort in this parable. Certainly, it is no part of their view that we should think of ourselves as having, over the generations, lost touch with earlier fully explicit conceptions of the meanings of vague expressions. What cannot be avoided, however, is the admission that, if the epistemicist view is right, we do not as a matter of fact *have* satisfactory such conceptions and have no idea how to go about recovering them. This is something that epistemicism needs to explain, and – so far as I am aware – no explanation has ever been offered. It's true that Williamson, for one, has worked hard to explain why we cannot know where the (alleged) sharp cut-offs come in Sorites series[7]. But it is not an explanation of that that's being asked for. Estimations of the application of 'is less than six feet tall' based on unaided observation would indeed be subject to margins of error, and for my present purposes we can concede that Williamson explains why, accordingly, we cannot know by such means where, in a

See especially ch. 8 of his *Vagueness*.

Sorites series of fifty men beginning at five foot six inches tall and increasing by small variable margins less than one quarter of an inch each time, the last man less than six feet tall is to be found. But that's not to explain why I cannot know the alleged principle or principles that stand to my use of 'bald' as the condition expressed by the predicate, 'is less than six foot tall' stands to the practice – at least according to the suggestion of the parable – of the Accolytes.

The real difficulty, though, is to make anything of the suggestion that the Accolytes' practice is indeed actually so informed. It is only because the Priests understood what it is to be less than six feet tall that they had any inkling about when it is safe to apply the predicate, or to deny its application, under circumstances of unaided observation. The practice so informed may be imitable in a relatively stable and transmissible way, but insofar as it is, there is no longer any motivation for saying that it is driven by the original precise principle. What drives it, rather, is a sense of the patterns of use – in which, for the Priests, the precise principle was once instrumental – purely as patterns *in their own right*. There is no sense in the idea of a continuing additional undertow exerted by an originating principle which nobody any longer grasps. The most fundamental difficulty with the epistemicist proposal is not merely the element of superstition involved in the suggestion that there are indeed, in the case of any vague expression which contributes to a successful linguistic practice, underlying principles which determine sharp extensions. The charge of superstition might be thought to be deflectable, at least to a degree, by the reflection that what at this point may appear to be the only possible alternatives – semantic incompleteness, *à la* Frege, and nihilism – appear radically unsatisfactory in their own right. But the more acute problem is that, so long as the epistemicist has to concede that we have no inkling of what the relevant principles are or how they might be ascertained, it is simply bad philosophy of mind to suppose that our linguistic practice consists, in some sense, in their implementation – that it is, in any meaningful sense, *regulated* by them at all.

IV. We appear then to have come to an impasse. How are we to think of the rules which govern the use of vague expressions as comporting themselves as we advance down a Sorites series of cases into

the borderline area? It seems that there are just three possibilities: either there's no prescription in the borderline area, or the prescription remains the same as it was, or it changes abruptly – semantic incompleteness, nihilism, and epistemicism. Yet we have reviewed serious causes for discontent with each of the three proposals. Each, indeed, comes short in the most basic way, by failing to offer a satisfactory explanation of one or another aspect of competent speakers' linguistic practice with vague expressions. Semantic incompleteness fails to explain our tolerance of conflicting verdicts in the borderline region, nihilism fails to explain the disparity in our reaction to the two components of the contradiction, and epistemicism, in so far as it is content to appeal to underlying semantic features transcending any sense that speakers have of competent linguistic performance, seems to disclaim any ambition of giving an explanation of our linguistic practice with vague expressions at all.

The puzzle is intense. But we get a pointer to what I've long believed to be the direction in which to find the correct response to it if we consider the well-known 'Forced March' variation on the Sorites paradox. The Forced March involves no reasoning to a contradiction from premises, plausible or otherwise. Rather we simply take a hapless subject case-by-case down a Sorites series, ranging from things which are clearly F at one end to things which are clearly not F at the other, and demand a verdict at every point. The subject starts out, naturally, with the verdict, F. And there are just two possibilities for what happens afterwards. Either he goes on returning that verdict all the way down, or at some point he does something different – if only refusing to issue a verdict at all. If he does the former, then eventually he'll say something false, and hence will betray incompetence. And if he does the latter, then he will draw a distinction by his responses which (i) will have no force of precedent for subsequent trials, and (ii) will correspond to no relevant distinction that he can call attention to between the last case where he gave the original verdict and the first case where he changes. So his verdicts will be unprincipled. Conclusion: anyone who uses a Sorites-prone expression can be forced to use it in ways which are either incompetent or unprincipled. If we add the plausible-seeming supposition that com-

petent linguistic practice is always essentially principled – always consists in the proper observance of semantic and grammatical rules – then the conclusion, on either horn, is that the use of vague language is bound to be incompetent.

The solution must be to break the tie between 'unprincipled' and 'incompetent'. But that's to say there has to be a sense in which competent classification utilising vague expressions does not consist in the implementation of the requirements of semantic rules. When the question is, what do the semantic rules subservience to which constitutes competence for a vague expression require when it comes to borderline cases, the answer we should give is not *any* of the three canvassed. The position is neither that the requirements give out, nor that they remain in force driving us on towards paradox, nor that they mandate a sharp cut-off. Rather, it is that competence with basic vague expressions is *not a matter of subservience to the requirements of rule at all*. This is, indeed, the conclusion to which I came in work on the Sorites paradox published thirty years ago[8]. The great difficulty now, as then, is to understand what it means.

The difficulties are the more acute when one reflects that, on at least one way of understanding the issue, to claim that basic classifications effected using Sorites-prone expressions are not rule-governed is absurd. It's absurd because such classifications are, of course, genuine classifications, apt to be *correct* or *incorrect*. And correctness, or incorrectness, will be a matter of fit, or failure of fit, between an actually delivered verdict and *what ought to be said*. There is therefore no alternative but to construe vague classification as, in some sense, subject to norm – and now, if the correctness of our verdicts in general is to be a matter of competence, rather than accident, there seems no option but to concede that we are in some way masters of these norms and follow them in our linguistic practice.

8  C. Wright "On the Coherence of Vague Predicates", *Synthese*, 30 (1975): 325-365 and "Language-Mastery and the Sorites Paradox", in G. Evans and J. McDowell (eds.), *Truth and Meaning*, Oxford, Oxford University Press, 1976, 223-247. Reprinted in R. Keefe and P. Smith (eds.), *Vagueness: a Reader*, Cambridge, Mass., Bradford/MIT, 1996.

I think there is indeed no option but to concede that. The point, however, is that such a concession may not amount to very much – in particular, it may not amount to anything which sets up the trilemma confronted above. In order to illustrate how this may be so, we can invoke a comparison with some thoughts about truth which I have canvassed in other work[9]. To speak the truth is to 'tell it like it is', represent things as they are, state what corresponds to the facts. Understood in one way, these phrases are platitudes and incorporate no substantial metaphysics of truth, at variance with, say, coherentist or pragmatist conceptions. The patter of correspondence, platitudinously understood, should motivate no questions about the nature of *facts* – what kind of entities they might be (as it were, sentence-shaped objects?) or how they might somehow be fitted to *correspond* in an appropriate way to beliefs and thoughts. I want to suggest that, in a similar way, the conception of basic classificatory linguistic practice as consisting in learning rules and following them is likewise open to a minimal, or platitudinous construal, but that the trilemma:

<semantic incompleteness, nihilism, epistemicism>,

arises only on a richer, non-deflated construal. The proper understanding of the idea that such classificatory competence is unprincipled is exactly that the relevant kind of richer construal is inappropriate.

So: that's the shape of the proposal. The question is how to fill it out. I shall conclude by outlining two possible ways.

V. The first proposal is to invoke the idea of *response-dependence*, in a sense in which to regard a concept as response-dependent is to regard the basic classifications which we make in exercising the concept not as answerable to facts settled independently of them but as somehow implicated in determining the extension of the concept

    [9]  See for instance "Truth: A Traditional Debate Reviewed" in supplementary volume 24 (1998) of the *Canadian Journal of Philosophy* (special issue on Pragmatism, guest edited by Cheryl Misak) 31-74; reprinted in *Truth*, ed. Simon Blackburn and Keith Simmons, Oxford, Oxford University Press, 1999, 203-38, and now in C. Wright, *Saving the Differences*, Cambridge, Mass., Harvard University Press, 2001, 241-287.

in question. So conceived, our basic classifications of colour, for instance, are not thought of as true or false in virtue of the independently constituted colours of things, but rather play a constitutive role in determining what colours things may correctly be said to have. A simple form of the idea is expressed in the Sellarsian proposal that to be red is to look red to standard observers under standard conditions, where this is intended as an expression not of the idea that colours cannot but shine through, as it were, in standard conditions but rather that how things strike standard observers under such conditions is *all there is* to their having the colours they do.

The suggestion, then, is that we look to the broad contrast between response- dependent and response-independent judgements to try to fill out the idea that, when it comes to basic classifications effected using vague concepts, there is no rule-following in any sense that subserves the awkward trilemma. Where there is response independence, truth is settled independently of our propensities of judgement, so our judgements had best keep track of the truths so settled. In such cases what it is correct to say or think is constitutively independent of what, if anything, we are inclined to say or think, even under the very best conditions for taking a view. If one thinks of classifications of particular shades of colour as 'red', or particular men as 'bald', in this spirit, then it should indeed make perfect sense to enquire what the rules really require of us in the borderline region, as our actual judgements give out or fall into instability and conflict, with just the same three possible forms of answer as before. But if one thinks instead of vague classifications as response-dependent, the motivation for the trilemma disappears. Now we no longer think of the propriety of a description as settled purely by the rules of description and the prevailing facts. Rather, it is our judgements that, in some way or other to be specified, contribute towards determining what the proper description is. Since the standard of correctness is one which our basic classificatory propensities contribute towards determining, there is no sense in asking what it demands of us – what the *real* requirement is – when those classificatory propensities go into disorder in the manner typically manifested in the borderline region. The question which invites the trilemma does not get off the ground.

I think this is an attractive direction[10], but its successful pursuit will be hostage to the outcome of three very substantial theoretical challenges. The first is the need for a *sharp characterisation* of response-dependence, or – more likely – of the varieties of response-dependence. Second, the characterisation has to be delivered in such a way as to minimise scope for controversy – it has to proceed in such terms that we can go on to *settle* whether, for example, judgements of colour, or judgements of aesthetic value, are indeed response-dependent judgements, and whether judgements of pure mathematics, or theoretical science, say, are not. Third, the account must deliver the means to show that basic classifications involving vague concepts are indeed response-dependent *as a class*. For if there are exceptions, the problems which the response-dependence proposal is meant to avoid will still arise locally, and be no less pressing or awkward for that.

That these are, potentially, formidable difficulties becomes very clear if one reflects on the lie of the land when illuminated by a proposal about response-dependence which I offered in *Truth and Objectivity*[11]. According to that proposal, we may consider that a predication is response-dependent if first, a true 'provisional biconditional' can be constructed – that is, where C are conditions on the judging subject and the object judged, a true subjunctive conditional of the form

$$C(Sx) \rightarrow (Fx \text{ iff } S \text{ judges: } Fx),$$

– were condition C to obtain, then whether x was F would coincide with S's impression of whether it was – and if, second, or so I argued, the following four conditions are met:

— that the conditions C are specified in fully explicit substantial terms (so as to avoid the kind of triviality that would be engendered by formulations like 'under conditions in which judges get these matters right'),

[10] I first mooted it in "Further Reflections on the Sorites Paradox", *Philosophical Topics*, XV, 1, 1987, 227-290. Reprinted in R. Keefe and P. Smith (eds.), *Vagueness: a Reader*, Cambridge, Mass., Bradford/MIT, 1996.

[11] C. Wright, *Truth and Objectivity*, Cambridge, Mass., Harvard University Press, 1992.

— that the conditions C are specified in such a way as to carry no unilateral implications about which things actually are F,

— that the provisional biconditional as a whole holds good a priori,

and that, finally, there is no explanation of the satisfaction of the former three conditions in terms of further features of our concept of what it is for something to be F which involve no reference to human judgements or to the conditions under which such judgements are at their best.

I will not here enlarge on the details of the motivation for imposing just these constraints. Broadly, they had to do with the idea that F's response-dependence should be a matter of our *concept* of what it is for something to be F, and should consist in there being no conceptual space between what, under constructively specified best conditions, we take to be F and what really is F – where there being 'no conceptual space' for that distinction requires that the coincidence between its components is a matter not of the infallibility of our tracking F-ness under the relevant conditions but rather of our having no concept of what F-ness consists in which would allow it to be merely coincident with the outcome of C-conditioned judgement[12].

Some readers may feel, no doubt, that there should be more wieldy notions of response-dependence than this. Maybe so. My present point is only that anyone who thinks about it will realise rapidly that there is complexity in the notion and – as my own prototypical formulation illustrates – that the gap between arriving at a well-motivated template for response-dependence and delivering cogent argument that particular concepts fall under it can be a considerable one. I do think, for example, that one can make a reasonable fist of arguing that our basic colour concepts are response-dependent by the lights of the *Truth and Objectivity* template. But the C-conditions which one needs to outline in order to make that claim plausible are complicated (since the scope for possibilities of divergence between an object's colour and how it appears is quite wide), and there is plenty of room for resistance by one who is inclined to maintain, for example, that colours are natural kinds.

[12] See *Truth and Objectivity*, appendix to Chapter 3.

And if the argument in particular cases can be expected to be difficult, how might one hope to accomplish a *general* demonstration that vague concepts *per se* give rise to response-dependence?

VI. So let me canvass a different, though compatible, proposal – one which takes us to the heart of the *Philosophical Investigations*.

Take as a simple, uncontroversial example of a rule-governed practice the case of Castling in chess. The rule states (something like):

> If the squares between the King and one of its Rooks are unoccupied, and if neither the King nor the Rook have moved previously in the course of the game, and if the King is not in check, nor would move through or into check, then it may be moved two squares toward the Rook and the latter then placed on the adjacent square behind it.

Following this rule involves verifying that the antecedent of the conditional articulating the requirement of the rule is satisfied in the circumstances of the particular game, and then availing oneself of the permission incorporated in the consequent. Generalising, what is suggested is a model of rule-following which involves reasoning, of the form of a modus ponens, from a conditional statement whose antecedent formulates the initial conditions of the operation of the rule, and whose consequent then articulates the mandate, permission or prohibition that the rule involves. We can call this the *modus ponens model* of rule-following[13].

The appropriateness of the modus ponens model is not restricted to cases where rule-following is informed by self-conscious inference. Following a familiar rule may be, very often, phenomenologically immediate and unreflective. But the important consideration, as far as the appropriateness of the modus ponens model goes, concerns what the rule-follower would acknowledge as justifying his performance. It is enough, in order for the modus ponens model to be appro-

---

[13] This idea, and its limitations, are anticipated in section V of my "Wittgenstein's Rule Following Considerations and the Central Project of Theoretical Linguistics" in *Reflections on Chomsky*, ed. Alexander George, Oxford, Blackwell, 1989, 233-64; reprinted at 170-213 of C. Wright *Rails to Infinity*, Cambridge, Mass., Harvard University Press 2001.

priate, that the explicitly inferential structure of reasons it calls for should surface in that context. A practiced chess player may decide to castle without any conscious thought but that of protecting his King from an attack down the left flank. But if the legality of the move were questioned, he would be prepared to advert explicitly to the relevant pattern of reasons of the modus ponens type, whose conditional premise embodied a formulation of the rule for Castling.

Plausibly, the modus ponens model actually extends a long way into the class of phenomenologically immediate decisions and judgements. How far does it extend? In particular, can the competent application of an expression *always* be conceived as rule-following in accordance with this simple prototype? More specifically still, can we regard the classifications we effect using basic vague expressions, of the kind typified by the usual suspects, as an example of rule-governed activity in accordance with the modus ponens model? The answer, I believe, is no, and the reasons why not are instructive.

Suppose, to the contrary, that competent classification using "red", e.g., is a rule-governed practice in the sense articulated by the modus ponens model. Then there should be a conditional which specifies the conditions under which predication of "red" to an item should be assented to, and such that each competent classification can be seen as inferentially grounded in the recognition, of a presented item, that it fulfils the antecedent. The picture, in other words, is that each informed, competent classification of an item, x, as red is underwritten by reasons of the form:

If something is X, then it should be classified as "red"

and

x is X

So, what is X? What is the property whose realisation underwrites the proper application of "red" as fulfillment of the antecedent for the rule for Castling underwrites the judgement that the particular situation in the game is one in which one is permitted to Castle? It doesn't seem that we know of any plausible candidate answer, apt for all cases, except to identify X with *red*. The same goes for the

general run of predicates that make up the usual suspects: basic, vague predicates used to record the results of casual observation. In such cases, the correct answer to the question, what is the condition common to the minor premise and the antecedent of the conditional for the modus ponens?, is irreducibly homophonic.

This has a striking consequence. It is that the price of continuing adherence to the modus ponens model in these cases is that we are forced to think of grasp of the concept of what is to be, say, red as *underlying and informing* competent practice with the predicate 'red'. If the explanation of competent practice with "red" adverts to the intention to follow a rule grasp of which demands a prior understanding of what is something to be red, then the latter grasp has to stand independent of the linguistic competence. So in effect, we commit ourselves to the picture of language mastery displayed in the quotation from Augustine with which Wittgenstein begins the *Philosophical Investigations* – not, to be sure, in the aspect of that picture which involves thinking of the semantics of expressions generally on the model that of a name and its bearer – but rather the aspect which we find made explicit only as late as § 32,

> And now, I think, we can say: Augustine describes a learning of human language as if a child came to a strange country and did not understand the language of the country; that is, as if he already had a language, only not this one. Or again: it is as if the child could already *think*, only not yet speak. And "think" would here mean something like "talk to himself".

Wittgenstein, I need hardly say, presents the quotation from Augustine as the prototype of a huge mistake. If we agree with him and simply repudiate the Augustinian picture, then we will conclude that, at the level of the basic, casual-observational classifications expressed by the usual suspects, the modus ponens model is inappropriate. The classifications effected by means of vague expressions of this basic kind are not supported by reasons, for the only possible candidates to constitute their reasons would demand, when so conceived, that we think of grasp of the concepts expressed by the vague predicates in question as something prior to and independent of the mastery of these predicates.

In his later work Wittgenstein gives expression to an epistemology of understanding whereby language mastery is systematically con-

ceived not merely as a means for the expression of concepts but as the medium in which our possession of them has its very being. As a reaction against the diametrically opposed view, that thought – at whatever degree of sophistication – may always intelligibly be conceived as constitutively independent of the thinker's possessing means for its expression, Wittgenstein's stance is surely compelling. But the correct view of the matter overall is surely something more nuanced and intermediate. There is much possible intelligent activity which, though wordless, would naturally call for explanation by ascription to the agent involved of conceptual – content-bearing – states of varying degrees of sophistication. Imagine a chimpanzee who, off its own bat and after many days of manipulation in the corner of his cage, is suddenly able to solve Rubik's cube; or the wonderful behaviour, even in an unfamiliar city, of a well-trained guide dog. On the other hand there are plenty of concepts – take, for instance, the concept of a parameter in a step of existential elimination – of which it strains credibility beyond breaking point to suppose that they could be fully grasped by someone with no linguistic competence at all. So if its implication of the Augustinian picture when applied to "red", "bald", "heap", etc., is to be sufficient reason to reject the modus ponens model of competence in those cases, we need a special consideration why. Is there one?

Let me just gesture here at one line of thought. We can certainly imagine a creature without language – the chimpanzee again – exhibiting some sort of concepts of colour: he may, for example, manifest a preference, among variously coloured but identically smelling and tasting sweets, for green ones. More generally, there are any number of kinds of colour-sorting behaviours that could be – and in many cases, are – exhibited by pre-linguistic children. But grasping the colour concepts we actually have is not merely a matter of dispositions of appropriate response to paradigms. To grasp any classificatory concept, one needs not just to learn to respond appropriately to central cases but also to acquire a sense of its *limits*. With the usual suspects, however, it is the very pattern of our *linguistic* practice that sets the limits, imprecise though they may be. We learn by immersion in the language how far one can stretch from paradigm cases of red, or orange, before classification starts to become controversial or

difficult. Pre-linguistic children no doubt have some sort of grasp of colour saliencies. But the raw concepts they have, or could have, do not match the vagueness of colour concepts as linguistically captured. The reason is that it is precisely the onset of hesitancy, disorder and weak conflict in the *linguistic* verdicts of the competent that constitutes the gradual intrusion into the borderline area and with it the limits of the concepts in question. I do not deny that the chimpanzee might behave in ways which went some way towards giving sense to the idea that the concepts he was working with were vague – he might hesitate over a turquoise sweet, for example. What seems to make no sense, however, is that his concept of green so manifested might be identical, vagueness and all, to the concept which, miraculously given the power of speech and perfect basic English, he went on to display in his uses of "green". Competence with vague concepts is an essentially linguistic competence because the extent of their vagueness is an essentially linguistically manifest phenomenon.

If the broad direction of the foregoing is correct, we must abandon the modus ponens model if we are to understand the sense in which the exercise of vague classifications is a form of rule-governed activity. One reaction would be to deny that it is, properly speaking, rule-governed at all – that it is sensible to think of an activity as *governed* by a rule only if one could in principle articulate the content of what would be the appropriate rules in such a way that a suitable thinker could *generate* a competence by observance of them. If I'm right, that condition is indeed failed where competence with vague classifications is concerned. That competence is, precisely, *not* to be viewed as a product of any possible anterior body of information which, in principle, could be used explicitly to inform the moves that competence requires. But actually, I don't think it matters whether we say that there is, properly speaking, no rule-following in such cases or whether we say merely – as I originally announced – that we need a more minimalistic conception of what rule-following involves. Certainly there is still all of correctness and incorrectness, criticisability, proper responsibility, the intention to get things right, and a wide range of contexts in which it is important to succeed in that intention. But what there is not is a body of information which underlies

competence, in a way in which knowledge of the rule for Castling together with knowledge of the history of the particular game and the present configuration of pieces enables (what may well be an unreflective) awareness by an expert player that Castling is an option here if he wants[14].

If this is right, then we can see past the trilemma. The trilemma arises if but only if it is a good question concerning the rules which govern our competence with "red": what do they *really have to say* when our colour classifications fall into the complex patterns which show that we are in the borderline area? – what is their message, what would we do if we were to do exactly what they require and nothing else? That the three answers canvassed seem, between them, to exhaust all of the alternatives, yet each to be objectionable, might have made one suspect that there is something wrong with the question. Now we see how that suspicion might be substantiated. There *is* no requirement imposed by the rules – not if we understand such a requirement as something whose character may be belied by our practice in the borderline area and into which there is scope for independent inquiry. There is no such requirement because to suppose otherwise is implicitly to commit oneself to the modus ponens model of rule-following for the classifications in question, and hence, as we have seen, to open an – at least locally – unsustainable gap between conceptual competence and the linguistic capacities that manifest it.

There is a lot more that needs to be discussed. Two matters in particular are pressing. First, consideration needs to be given to the question what account of the nature of borderline cases *per se* is fostered by the recognition that classifications using vague expressions are rule-governed only in a minimal sense. Must the best such account provide houseroom, after all, for the idea of truth-value gaps? Will there still be provision for higher-order vagueness or can the appearance of such a phenomenon somehow be explained away in

---

[14]   Additional difficulties for the modus ponens model are raised in Paul Boghossian's "Meaning, Rules and Intention", forthcoming in *Philosophical Studies*.

the minimalist setting? Second, and relatedly, we need to understand better what, if any, accommodation can be provided by these ideas for our liberality in the borderline region – our intuitive inclination to be forgiving of weakly conflicting opinions in the borderline area and of failures to reach any opinion[15]. There are questions concerning 5the nature of borderline cases, the reasons they arise, and the attitudes and meta-attitudes towards them which competence involves, which much contemporary writing on the issues ignores and which any decent account of vagueness must in the end address. They may be a little easier after we drop the assumption that underwrites the trilemma. But they are still to answer.

[15] How for instance does it play in this setting with Verdict-Exclusion – the thesis that no verdict about a borderline case can be knowledgeable (and that this fact is itself known)?

# CLARIFYING THE LOGIC OF CLEAR CASES

*Roy T. Cook*

## 1. *What is Vagueness?*

In writing about vagueness, there are a number of issues that one might address. These include:

*Explanatory*
Providing an explanation of what makes a vague predicate vague.

*Semantic*
Providing a semantics for languages containing vague predicates.

*Logical*
Providing a solution to the *Sorites* paradox and related puzzles.

*Identificatory*
Providing a characterization of vague predicates that is neutral with respect to the various accounts which might be provided to complete one or more of the first three projects.

In order to distinguish between these four projects, we can draw an analogy between the present project and medical practice. A doctor, in investigating a new disease, needs to accomplish three tasks. First, he must identify the symptoms of the disease. Second, he needs to diagnose the deeper underlying cause(s) of the disease. Third, he needs to determine what remedy (if any) can be prescribed to the patient in order to cure (or control) the disease. The identificatory project is analogous to determination of the symptoms characteristic of the disease in question. Diagnosis corresponds to the explanatory