

MEANING AND RULE-FOLLOWING

Richard Holton

The rule following considerations consist of a cluster of arguments which purport to show that the ordinary notion of following a rule is illusory; this in turn has profound consequences for the concept of meaning. In response, some have tried to show that it is necessary to revise the ordinary concept of meaning; others have tried to find flaws in the arguments.

I. INTRODUCTION

It is a common thought that an individual's words mean what they do in virtue of the fact that the individual is following a rule for the use of that word: a rule which determines what is a correct, and what an incorrect, use of the term. This idea has been challenged in a cluster of arguments given by Wittgenstein, which in turn have received great impetus in the interpretation given by Saul Kripke. According to Kripke, the central problem is this: there does not seem to be any set of facts which determines that a speaker is following one rule rather than another. In a little more detail: suppose that, in using a certain word, an individual seems to be following a certain rule, R. Then we can always come up with a different rule, R', which fits all the facts about the individual's use of the word just as well as R. In virtue of what is the individual following R rather than R'? According to Kripke there is no straightforward answer to this question. The ordinary conception of what it is to follow a rule is simply mistaken; so in turn is the ordinary conception of what it is for a word to be meaningful. The best that is available a surrogate for the ordinary conception; what Kripke calls a 'skeptical solution'. This takes the form of a social criterion of rule following: an individual follows a rule in so far as their behaviour conforms to that of the other members of their community. A mistake simply consists in deviation from what the others do. Thus an individual's words mean whatever the other members of their community use them to mean. This in turn can be seen as providing the basis for Wittgenstein's famous denial of the possibility of a private language. A private language is impossible since there would be no community-wide use to provide the correct use of its words, and hence they would have no meaning.

Kripke's presentation of the rule following worries has engendered a large literature. His skeptical solution has not been widely accepted. Responses have typically fallen into one of two classes. Either it has been argued that there are further facts about an individual, overlooked by Kripke, which determine which rule they are following. Or it has been argued that Kripke's whole presentation is vitiated by an over reductionism approach. These responses will be examined in what follows; first it is necessary to get clearer on the nature of the rule following problem itself.

2. THE RULE-FOLLOWING PROBLEM

The passages of Wittgenstein that provide the inspiration for the argument appear, in typically compressed and gnomic form, in Wittgenstein 1953, ¶138–242, and Wittgenstein 1978 §VI. Kripke's account was published in its final form in Kripke 1980, having been widely presented and discussed before; a similar interpretation was given independently in Fogelin 1987 (first edition 1976). There has been some controversy as to how accurate Kripke's presentation of Wittgenstein's views are; a number of authors have embraced the term 'Kripkenstein' to refer to the imaginary proponent of the views discussed by Kripke. The exegetical issue will not be of concern here. The concern will be only with the argument presented by Kripke. (For discussion of the exegetical issues see Boghossian, 1991)

The central idea is easily put. Imagine an individual who makes statements using the sign '+'. For instance, they say ' $14 + 7 = 21$ ', ' $3 + 23 = 26$ '. It might be thought that they are following the rule that '+' denotes the plus function. But consider the sums using '+' that they have never performed before (there must be infinitely many of these, since they can only have performed finitely many sums). Suppose that ' $68 + 57$ ' is one such. Now consider the quus function, which is stipulated to be just like the plus function, except that 68 quus 57 is 5. What is it about the individual that makes it true that they have been using '=' to denote the plus function rather than the quus function? By hypothesis it cannot be that they have returned the answer '125' to the question 'What is $68 + 57$?', since they have never performed that sum before.

The immediate response is that the individual meant plus in virtue of having mastered some further rule: for instance, the rule that, to obtain the answer to the question 'What is $68 + 57$?' one counts out a heap of 68 marbles, counts out another of 57, combines the two heaps, and then counts the result. But now reapply the worry. How can it be known that by 'count' the individual did not mean 'quount', where, of course, this means the same as 'count' except when applied to a heap constructed from two piles, one containing 68 objects, the other containing 57, in which case one correctly quounts the pile if one simply returns the answer 5? One might try to fix the meaning of 'count' by some further rule; but this will just invite further worries about what is meant by the words occurring in that rule. Clearly there is a regress. Any rule that is offered to fix the interpretation of a rule will always be open to further interpretations itself.

At this point it might be suggested that it is a fact about the individual's *dispositions* that determines that they meant plus: the relevant fact is that they would have answered '125' if one had asked them 'What is $68 + 57$?'. But that response falls foul of the normativity constraint on rules. Rules are things that tell us what we *ought* to do; dispositions are simply facts about what we *would* do. To think that facts about the rules we follow can be derived from facts about our dispositions is to try, illegitimately, to derive an 'ought' from an 'is'. Thus even if, due to some cognitive blip, the individual would in fact have answered '5' to the question 'What is $68 + 57$?', one would still want to say that by '+' they

meant plus, but that they simply made a mistake. If one identifies what their rule requires them to do with what they would in fact do, no possibility is left that they might go wrong. Yet it is part of the very idea of a rule that one can make sense of the idea of breaking it. So one cannot identify the rule that an individual would follow with what they are disposed to do. No answer has been found to the question of what the individual means by their words.

It is important to see just how wide the repercussions of this are. First note that although Kripke presents the problem as an epistemic one (how can one *know* that the individual meant plus rather than quus), the deeper problem is constitutive: there seems to be nothing for meaning to consist in. Second note that the problems do not arise just because Kripke restricts himself to behavioural evidence. The argument is that there is nothing in the individual's mind, not just their behaviour, that can determine what they mean. (Here there is an important difference between Kripke's arguments and the influential arguments for the indeterminacy of meaning given by Quine (1960); Quine does restrict himself to behavioural evidence.) Third, note that it is not just a matter of the meaning of words in a public language. The same considerations apply to the very contents of an individual's thoughts themselves. How can an individual know that their thought that $2 + 2 = 4$ is a thought about plus, and not a thought about quus? This makes the worry go much deeper; and it also makes Kripke's skeptical solution much more shocking. For it is one thing to say that the meaning of words in a public language are determined by the behaviour of the community; there is a sense in which, given the conventional nature of language, something like that is bound to be true. It is quite another to say that the very contents of an individual's thoughts are somehow determined by the community around them. It is now time to examine the skeptical solution a little more deeply

3. THE SKEPTICAL SOLUTION

Recall that the basic problem that arose in identifying following a rule with having a disposition: it gives us no purchase on the idea of making a mistake. But once one identifies following a rule with conforming to the dispositions of the *community*, the idea of a mistake can come back: an individual makes a mistake in so far as they do not conform. What this approach can make no sense of is the idea of the whole community going wrong. In proposing this solution, Kripke is not suggesting that he has recaptured everything in the ordinary notion of meaning. After all, it is normally thought that the reason that the whole community gives the same answer to 'What is $68 + 57$?' is because they have grasped the same concept. But as Kripke points out (p. 97) the skeptical solution does not allow us to say that; grasping the same concept isn't some independent state which explains convergence in judgement. Rather, the skeptical solution is meant as a *surrogate* for our ordinary conception of meaning; it is the best that there is.

However, it is questionable whether it is good enough. It might be accepted that the ordinary philosophical picture of the nature of meaning is a false one, one which the

skeptical solution corrects. It would be much worse if the skeptical solution were at odds with the ordinary practice of making judgements. Yet it seems that it is. As Paul Boghossian has noted, there are plenty of cases in which, for ordinary predictably reasons, the whole community tends to go wrong (Boghossian 1990 pp. 535–6). Thus consider a good magician. With sleight of hand she can get us all to believe that she has put the egg into the hat; only she knows that it is under the table. So the community agrees that the predicate ‘is in the hat’ applies to the egg; the magician is the odd one out. Yet it would be ridiculous to say, as it seems the skeptical solution must, that the community is right and the magician wrong. She is the one who knows the truth. The rest of the community have simply been tricked.

A natural response to this is to say that not every disposition of the community should be respected. It is important to pick and choose. In this case, the disposition of the magician to say that the egg isn’t in the hat should count for far more than that of the rest of the community to say that it is. However, once that move is made, the primary motivation for the skeptical solution has been lost. The whole point was that using the actual dispositions of an individual could not give rise to the possibility of error; so it was necessary to look to the actual dispositions of the community instead. But once one starts to pick and choose amongst the dispositions of the community, saying that some must be given greater weight than others, the possibility of error is reintroduced into the individual case. Here too, if some dispositions are favoured over others, then the individual can be thought of as falling into error: they will be in error whenever they go with a less favoured disposition over a more favoured. To say this is not to deny the role of the community altogether; for instance, it could be that in sorting out the preferred dispositions it will be necessary to have recourse to the dispositions of the experts, and this is a position conferred by the community. But the *essential* role of the community will have been lost, and with it what is distinctive about the skeptical solution. Let us now turn to this alternative proposal, and examine in more detail the grounds on which some dispositions might be preferred to others.

4. LAUNDERING THE DISPOSITIONS

The central thought is that the correct application of a term cannot be identified with that which an individual is disposed to apply it to, nor with what the community is disposed to apply it to; but it might be possible to identify it with what the individual or the community is disposed to apply it to *under certain circumstances*. This is the device that enables some dispositions to be preferred to others. The proposal here parallels certain versions of the dispositional theory of colour, and useful lessons can be gained from considering that.

According to a crude dispositional theory of colour, an object is, say, red, if and only if observers have the disposition to judge it as red. But that crude form of the theory is hopeless. Some observers are colour-blind; some circumstances, for instance coloured lighting, will distort the colour judgements even of someone who is not. So a more

plausible dispositional account of red will say that an object is red if and only if competent observers in favourable circumstances have the disposition to judge it as red. This makes the account more plausible; and it re-introduces the possibility of error. Yet it raises a difficulty, that of specifying what competent observers and favourable circumstances are. For if all that can be said about them is that they are, respectively, the people who are good at identifying red things, and the circumstances that are good for such identifications, then the account will be circular.

Parallel issues arise for the rule following concerns. Suppose it were said that an individual means plus rather than quus by '+' just in case, *under favourable circumstances*, they have the disposition to use it to denote the addition function, i.e. to answer '125' to 'What is $68 + 57$?'; '2' to 'What is $1 + 1$?' and so on. Now the possibility of error has been reintroduced, since sometimes the circumstances will not be favourable. But how should the favourable circumstances be specified? They cannot be specified as those circumstances in which the individual uses '+' to denote the addition function; for that makes the proposal trivial. Worse, the proposal can provide no support for thinking that the individual does in fact use '+' to mean plus, since it is equally true that they have a disposition to use '+' to denote quus in those circumstances in which they use it to denote the quaddition function.

Moreover, it seems that it won't do to say that favourable circumstances are those in which the individual is thinking hard, and has no distractions, for even here people make mistakes. Everyone does. Of course, in such cases the individual will not be thinking *hard enough*; but if that is shorthand for 'not thinking hard enough to get it right', then the proposal has lapsed back to triviality.

What other ways might there be to launder dispositions without triviality? One response, stemming from those who seek to give a biological foundation to meaning, is to seek to spell out competent speakers and favourable circumstances in terms of *proper functioning*, where this in turn is spelled out in terms of the function that has been selected for. This is typically presented as part of a broader project of understanding semantic notions in terms of evolutionary ones. Whatever the merits of the broader project, it faces great difficulties in providing a response to the rule following worries. One kind of problem arises when we try to give an account of highly abstract terms, terms which do not bring obvious evolutionary advantage to those who grasp them: the case of addition, discussed so far, provides a case in point, let alone the concepts of, say, algebraic topology. But suppose one were to consider more down to Earth concepts, like that of *cow*. Could it be argued that the reason a given individual means *cow* by 'cow' rather than, say, *cow unless today is 25th January 2150, in which case horse* is that they have been selected to do so? It appears not; for, since we have not yet reached 25th January 2150, it will be true of every cow that the individual has in fact met that they are: a cow unless today is 25th January in which case they are a horse. (This example makes clear the similarities that the rule following problem bears to Goodman's 'New Riddle of Induction' and the famous example of grue; see (Goodman 1973), and Kripke's discussion pp. 20, 58-9. Similar examples could be given using disjunctive concepts, for instance that of being *a cow or a marshmallow*: it is true of every cow that it is either a cow or a marshmallow. For discussion, and a suggested way of

avoiding this latter problem, see (Fodor 1990), Chs 3 & 4.) It might be argued that the property of being a cow unless today is 25th January 2150, in which case a horse, is not a very natural one. That is surely correct. But once that property is ruled out as an eligible referent for the term 'cow' on the grounds of its unnaturalness it is unclear that the talk of proper function is actually doing any work at all. (For an example of a teleological response to the rule following problem, see (Millikan 1990); for discussion of the problems with the approach in general see (Boghossian 1989) pp. 537ff. For a discussion of the advantages to be gained by treating naturalness as a primitive feature of properties, including an application to rule-following, see (Lewis 1983), esp. pp. 375–7)

A more promising approach to laundering dispositions is suggested by Philip Pettit (1999). Why not think of the ideal conditions as involving an idealization of our actual practices of correcting ourselves and others in the light of discrepancy and divergence? This reintroduces a social element: individuals mean plus rather than quus by '+' just in case the process of correcting themselves and others in which they actually engage will, if taken as far as it can be taken, result in them using it to denote the addition function. However, there is in this proposal no unwanted guarantee that the community, as it now stands, is right in its use of terms; for the process of correction might not yet have been carried out. And provided that the idealization is not understood in a trivializing way (i.e. as the process that will bring individuals to apply their terms correctly according to what they mean by them) the proposal will not be trivial. Nonetheless a worry remains. It seems that there is no *a priori* guarantee that a community's actual practices, even if idealized, will lead them to use their terms correctly. For couldn't it be that their practices are prey to a deep and systematic error, such that however much they are idealized, the error will remain? Perhaps, to extend the example given before, nature herself is the magician, and the community will never see through her tricks. In such a case, following Pettit's method of resolving differences and contradictions would perhaps lead them to a consistent theory, and one that they would be quite happy with; but it would be a theory involving misapplications of their own concepts—misapplications which, given their cast of mind, they would never be in a position to shake off. All this seems possible. Yet if what is meant by a word is what its users would be disposed to apply it to after they had resolved all discrepancies and divergences, no such errors would be possible. It might seem very implausible that communities are in fact prey to errors of this sort; but it is not clear that there should be an *a priori* guarantee that they are not.

5. ANTI-REDUCTIONIST APPROACHES

The rule following worry consists in the thought that we cannot give a characterization of what following one linguistic rule rather than another amounts to. But what kind of characterization is required? It might appear that what Kripke has shown to be unobtainable is a characterization of facts about rule-following in terms of other sorts of facts. What it shows then is that a certain kind of reductionism is unavailable. So another response to the rule following worries is just to give up on reductionism: conclude that

meaning facts are simply not reducible to other sorts of facts, in particular, conclude that they are not reducible to facts about actual or possible behaviour. This is the approach taken by McDowell (1984, 1993); see also (Wright 1989).

The flight from reductionism has been a feature of much recent analytic philosophy; and social scientists schooled in the *verstehen* approach will find the response natural. But what exactly does a rejection of reductionism with respect to meaning consist in? One response would be to say that each individual meaning fact (that 'plus' means plus, for example) is *sui generis* and primitive; one simply grasps it or one doesn't. Kripke briefly discusses and rejects such an approach (p. 51), and he is surely right that it is most implausible. No matter how skeptical one might be about the existence of analytic equivalences, it cannot be denied that there are systematic connections between different meaning facts; how else could we explain the meanings of new words in terms of old? A more plausible claim is that, whilst there are connections between meaning facts, no meaning fact can be reduced to a non-meaning fact. This is how proponents of anti-reductionism typically conceive of it. The contention is that the rule following worries only arise if one tries to give an account of meaning facts in terms of facts that make no reference to meaning. (Note: if this really is to be anti-reductionism, the position had better be that there is no way whatsoever of deriving meaning facts from non-meaning facts, not just that there are ways of formulating meaning facts that do not admit of reduction. Thus, for instance, Paul Horwich denies that there is any reductive substitution for 'F-ness' in the sentence "'Plus' means F-ness'; but he still claims that meaning facts are reducible to facts about use. As he correctly says, this is a rejection of the non-reductionist position. See Horwich 1998.)

An obvious difficulty with this anti-reductionist approach concerns how to fit it into a naturalistic world picture. Naturalism has been understood by many to require that every fact is ultimately reducible to a fact of natural science. This general issue of naturalistic reducibility is an enormous one, which cannot be pursued here. Clearly the success of an anti-reductive response to the rule-following worries will depend on the success of anti-reductive approaches more generally. However, the worry can be somewhat lessened by pointing out that whilst anti-reductionists reject the idea that meaning facts are *reducible* to non-meaning facts, it is not obvious that they need reject the idea that they *supervene* on natural facts. Reduction requires that, for every meaning fact, a corresponding non-meaning fact can be given. Supervenience requires only the weaker thesis that the class of meaning facts are somehow made true by the class of non-meaning facts. That is, someone who embraces supervenience without reduction will accept that non-meaning facts (in particular, facts about usage, belief and the like) make true the meaning facts, in the sense that there could not be a change in the meaning facts without a change in the non-meaning facts. Whether such a picture can really be made good, however, remains to be seen. (There is a gesture towards such a picture in McDowell 1984, p. 348; for a fuller discussion of supervenience of meaning on use, without reduction, see Williamson 1994, pp. 205–9.)

BIBLIOGRAPHY

- Blackburn S., 1984: The Individual Strikes Back. *Synthese* 58: 281–301
- Boghossian P., 1989: The Rule-Following Considerations. *Mind* 98: 507–549
- Boghossian P., 1991: The Problem of Meaning in Wittgenstein. In: K Puhl (ed.) *Meaning Skepticism*. De Gruyter, Berlin
- Chomsky N., 1986: *Knowledge of Language*. Praeger, New York. Chapter 4
- Fodor J., 1990: *A Theory of Content and Other Essays*. MIT Press, Cambridge MA
- Fogelin R., 1987: *Wittgenstein*. Routledge, London
- Goodman N., 1973: *Fact, Fiction, and Forecast* 3rd ed. Bobbs-Merrill, Indianapolis.
- Horwich, P., 1998: *Meaning*. Clarendon Press, Oxford
- Kripke S., 1982: *Wittgenstein on Rules and Private Language*. Harvard University Press, Cambridge MA
- Lewis D., 1983: New Work for a Theory of Universals. *Australasian Journal of Philosophy* 61: 343–77
- McDowell J., 1984: Wittgenstein on Following a Rule. *Synthese* 58: 325–63
- McDowell J., 1993: Meaning and Intentionality in Wittgenstein's Later Philosophy. In P French, T Uehling, H Wettstein (eds.) *Midwest Studies in Philosophy* 17. University of Notre Dame Press, Notre Dame Indiana
- McGinn C., 1984: *Wittgenstein on Meaning*. Basil Blackwell, Oxford
- Millikan R., 1990: Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox. *Philosophical Review* 99: 323–53
- Pettit P., 1990: The Reality of Rule Following. *Mind* 99: 1–21
- Pettit P., 1999: A Theory of Normal and Ideal Conditions. *Philosophical Studies* 96: 21–44
- Quine W., 1960: *Word and Object*. MIT Press, Cambridge MA
- Williamson T., 1994: *Vagueness*. Routledge, London
- Wittgenstein L., 1958: *Philosophical Investigations*. Basil Blackwell, Oxford

- Wittgenstein L., 1978: *Remarks on the Foundations of Mathematics*. Basil Blackwell, Oxford
- Wright C., 1984: Kripke's Account of the Argument Against Private Language. *Journal of Philosophy* 81: 759–78
- Wright C., 1989: Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics. In A George (ed.) *Reflections on Chomsky*. Basil Blackwell, Oxford