THE EPISTEMOLOGY OF SELF-KNOWLEDGE AND THE PRESUPPOSITIONS OF RULE-FOLLOWING*

Phenomena such as our "understanding in a flash" and our immediate knowledge of the meaning of our own utterances seem to point to problems that call for philosophical explanation. Even though the meaning of an utterance appears to depend on where and when we use it, on what we use it for and on what we expect in response, we do not examine such circumstances when asked what we mean. Instead we simply say what we mean. Similarly, our having grasped a rule is something shown by how we perform certain tasks and respond to certain requests. But we frequently declare that we have indeed grasped a rule without paying any attention to those overt performances and, despite this, we are normally correct. These facts seem puzzling and impel us towards a certain philosophical picture of meaning and understanding. This picture identifies the meaning of a subject's utterances, and his understanding of the rules that he follows, with some kind of structure of which he has immediate knowledge. By virtue of their connection with these private, meaning-constituting phenomena, the public manifestations of meaning and understanding are invaluable as clues to the meaning of an utterance or a subject's understanding of a rule. But the latter are, nevertheless, ultimately fixed by the inner structures to which only the subject in question has immediate access and upon which he or she therefore has authority. This picture prompts a host of perplexing questions. "What are these immediately-knowable structures?" "What does it mean to say that we have immediate acquaintance with them?" "How are these private structures connected with those public performances?" These questions are notoriously difficult to answer.

Wittgenstein's approach to philosophical problems does not directly address the questions that they raise. Characteristically he examines earlier stages in the genesis of these difficulties, stages at which a mundane

"The Epistemology of Self-knowledge and the Presuppositions of Rule-following" by Denis McManus, *The Monist*, vol. 78, no. 4, pp. 496–514. Copyright © 1995, THE MONIST, La Salle, Illinois 61301.

fact may have been misinterpreted or a connection with a broader context overlooked. Part I of this paper attempts to show that this approach can free us from the awkward philosophical questions that are generated by the self-ascription of meaning and understanding. It performs this task by arguing that our general accuracy in such judgments is not a fact to be explained by the kind of philosophical account described above but is instead a condition of our treating a subject as capable of speaking meaningfully or acting with understanding. Although it is tempting to depict our proficiency in such judgments as some sort of cognitive achievement, such a depiction loses its appeal once we recognize that a collapse in that proficiency cannot be regarded as ignorance. Rather a person's being repeatedly wrong in such judgments puts in doubt his or her status as an intelligible agent. Accuracy in our self-ascription of the relevant states is interwoven with the possession of those states in such a manner that, when we try to construct a picture of what "ignorance" would be like, we find ourselves excluding not only knowledge of the states but also the states themselves.

Part II explores the impact that a failure to recognize these truths has had on our understanding of rule-following. Much as ignorance of one's own intentions undermines one's possession of intentions, a certain minimal self-transparency is a presupposition of a creature's following rules and I demonstrate that one of the most popular arguments offered in the recent revival of "community views" of rule-following rests squarely on a failure to acknowledge this fact. Although presented as proving that rule-following only makes sense in the context of a community, the argument in question merely shows that anyone we see as a rule-follower must manifest a degree of self-understanding. This realisation dissolves the apparent tension between the authority of individuals over the meaning of their own utterances and the potentially divergent evidence that their use of these utterances provides. Although they may appear to be rivals in the interpretation of a person's actions, these two "sources" must be in step with each other if we are to see the person as acting meaningfully. Rather than the expression of two incompatible meanings, their divergence would signal the breakdown of meaningful action.

I

Several aspects of our understanding of meaning encourage the notion that meaning is a kind of inner experience. For instance, if I wish to know what a person means by one of his utterances, I ask *him* to explain, rather than anybody else, and it makes no sense to say in reply to him "Are you sure?" or "How do you know?" In this respect, a person's knowledge of the meaning of his own utterances is similar to his knowledge that, for example, he is in pain. Also when learning the use of a new term or the application of an unfamiliar rule, we often say things such as "Oh I see!," "Now I get it!" or "Yes, now I can go on!," proceeding then to demonstrate the sought-after ability without having displayed any mastery up until that point. Although understanding manifests itself in an ability extended over time, this phenomenon of "understanding in a flash" also encourages the identification of meaning with some sort of experience, something which can appear in an instant and to which the learning subject has special access. Wittgenstein depicts this interpretation of our knowledge of our own understanding as a "mythological description" (Wittgenstein 1967 sec. 221). Not only do we seem to have no real acquaintance with the peculiar structures that this view postulates, but Wittgenstein argues convincingly that these structures cannot, in any case, explain what they were invoked to explain.

Wittgenstein's demolition of this familiar account of first-person knowledge of meaning may seem to threaten us with some sort of free-fall, in which our claim to know the meaning of our own words appears unjustified. Having undermined our favoured theory, we are left to seek, with an added urgency, a new, improved solution to our problem. Such an outcome falls short of the ideal of philosophical analysis that Wittgenstein proposed. The novelty of his approach is not captured by the insight that philosophical solutions are generally inadequate, since that is something of which non-Wittgensteinian philosophers are well aware. Rather what is presented as a puzzling phenomenon, necessitating some kind of philosophical explanation, must be shown to be nothing of the sort, must be shown not to merit such a response. By offering what I would suggest is a "dissolution" of the "problem of first-person authority," this paper attempts to show that this Wittgensteinian conception of philosophy can be given some substance.

What I wish to suggest is that the authority of self-ascriptions lies in their implication within a network of familiar facts whose significance has gone unappreciated. Viewed from this perspective, we can understand our need (under all but extreme circumstances) to take a speaker's word as the last word on her own meanings and we can do so without needing to commit ourselves to the reality of self-interpreting rules ("last interpretations" [Wittgenstein 1969 p. 34]) or mental entities which "are their own descriptions" [Wittgenstein 1971 p. 235]). Such postulates are wheels turning idly (Wittgenstein 1967 sec. 271) in that a general accuracy in claims to understand "in a flash" is presupposed when we treat someone as a subject. It is a precondition of regarding someone as a rule-follower that we take seriously her claims to understand "in a flash" and it is this truth that is "mythologically described" by our philosophical fantasies of, for example, immediate acquaintance. Once we gain an *Übersicht*, an overview of how we handle the relevant intentional concepts, we can see the "real foundation" of the authority of self-ascription which, though "always before [our] eyes," has been hidden by its "simplicity and familiarity" (sec. 129).

The Wittgensteinian question that I suggest we begin with is: What would life be like without these "peripherals"? For example, what if people generally could *not* say when they had understood what they were being taught? Certainly they can be wrong in making such judgments. But what if they were *usually* wrong? One difference between individuals who are incompetent in this way and "the rest of us" is the absence of a strong, inductively-establishable relation between the tendency to self-ascribe understanding and the manifestation of understanding.[1] The absence of that relation makes our imagined situation appear to be an instance of ignorance, of reprehensible extrapolation. In such cases, we would not be inclined to take the person at his word when he says he understands what he is being taught. But the problem is deeper than his merely being rash. Consider how we understand the utterance "Now I can go on!" in the mouth of such a person. What inclines us to give the "usual translation" to this utterance? Looking at the inductive evidence, his utterance of these words constitutes nothing like a serious signal of understanding, let alone an "expression of understanding." It is almost as if his parents had mistakenly taught him to say "Now I can go on!" instead of "I still haven't quite got it yet." One could perhaps argue that this person really does mean what he says but just happens to be a very bad judge of these matters. The problem with this response is that certain basic capacities need to be in evidence before we extend to someone the grace of taking him at his word in situations where circumstances mitigate against it. A certain level of proficiency is necessary before we will even say someone *is* performing a particular activity *badly*. Consider, for example, how we would distinguish someone who simply writes down patterns of numbers from someone who calculates incompetently. The difference seems to have something to do with the circumstances in which she makes her "errors" and how she reacts to them when they occur. If what we deem to be an error occurs repeatedly under conditions that we would see as ideal (such as others carefully and clearly drawing the subject's attention to her "error") we would seem to have incorrectly interpreted the subject's "error" as an error. Much as "[t]here is a continuum between an error in calculation and a different mode of calculating" (Wittgenstein 1977 Part III sec. 293), deeply incompetent calculation merges into the random manipulation of mathematical symbols.

Distinguishing the mathematically incompetent from the mathematically innocent required that we examine the circumstances under which "errors" were made and the subject's reaction to those "errors." Turning to our case of the person who is characteristically wrong when he says "Now I can go on," we should be able to tell how plausible it is to see this as a matter of *incompetence* by imagining what the rest of this person's behaviour is like. If it reflects the usual understanding of "Now I can go on!," the rest of his behaviour must surely be a smiling and resolute confidence in a new-found ability. If, on the other hand, it mirrors the inductive evidence, it would be hesitant, confused and unassertive. What kind of people do we envisage here? The latter would appear, at best, to be joking when he says the crucial phrase. The former would appear to be mentally disturbed. He happily ascribes understanding to himself and then demonstrates total inability. And not once, but time and time again. He does not check his claims and fails to manifest any recognition of his errors. If this were insincerity, it would be of the order of a Walter Mitty. If it were incompetence, it would be the kind that gets a person certified as insane. His self-ascriptions are like the pantomimic pronouncements of a child. Described in the abstract, this person's "disability" sounds reasonably delimited. But when we fill out the picture of what such a person would be like, we see that his "disability" has ramifications which penetrate the entire human personality.

The notion that one might be unreliable in the assessment of one's own intentions, while remaining an otherwise "intact" agent, seems similarly unintelligible. We can, of course, be mistaken about what we think we "really want" and we can worry whether the way we feel about a matter right now reflects how we "really feel" about it. But should these conditions become anything other than extremely rare, we would have difficulty in ascribing intentions to the person in question. Once again, the circumstances of this unreliability are crucial in determining how it is to be interpreted. If someone says she will perform an act and subsequently does not, we assume that she has forgotten about performing the act, has changed her mind or lied initially. To determine which interpretation is appropriate, we examine the rest of her verbal and non-verbal behaviour, looking for the specific patterns of action and reaction that characteristically accompany these different possibilities. If, however, someone says she will perform an act, subsequently does not and is then *surprised*, we would not know what to say. At best, we might start groping towards terms like "dissociation of the self" or "self-alienation." If this were a one-off incident, we could warily re-establish normal relations with the person involved. If, however, this kind of incident occurred frequently, we would begin to suspect that the person had lost her mind. It is tragic understatement to call the difference between someone who knows her own intentions and someone who does not (in the radical way described) a difference in *knowledge*. The latter "individual" is so "dissociated," so "decentred" in a real sense, that she lacks the kind of structure within which one might start ascribing privative forms of awareness. Calling her "ignorant" is almost a pathetic fallacy.[2]

This example suggests that if we are to see an utterance as a statement of intention we must be able to see it as correct in the vast majority of cases in which circumstances prevent its interpretation as a joke, a lie or a slip of the tongue. We have uncovered here a sense in which we must indeed "know" our own intentions in that our very *possession* of intentions appears to depend on the actions we say we will perform generally proving to be the actions that we do go on to perform. Our example suggests that although one can build on to the possession of intentions a first-person knowledge of those intentions (meaning by that no more than a subject's tending to make utterances that can be interpreted as correct statements of his own intentions) one cannot build on *anything less*. That is to say, an ability to judge one's own intentions which is only more or less reliable (and results in a significant number of mistaken claims) undermines the very possession of intentions on to which this ability would be built. The building-block metaphor is, therefore, misleading. What I am *not* suggesting is that only creatures that have certain linguistic capacities can possess intentions (although there may be some truth in that claim). The claim I want to make is merely that *if* a creature makes assertions about its own intentions, these must generally be correct. The form of ignorance upon which I wish to cast doubt involves not an *absence* of knowledge claims but the *failure* of knowledge claims.

Some will see this approach to self-knowledge as superficial. Consider the following argument made by Hugh Mellor:

> I perceive my own beliefs without using my outer senses. But some perceptual mechanism there must be. Assent does not occur by magic, nor is it an accident that it generally reveals what I believe. (Mellor 1978 p. 98)

According to the perspective I have offered, the reason why it is indeed not by magic that self-knowledge comes about is that only someone who manifests a certain minimal self-transparency would we treat as an intelligible subject. Again that is not to say that only creatures which can perform acts identifiable as self-ascriptions of intention can have intentions. But if we interpret some of a creature's acts as self-ascriptions, we must also generally interpret these judgments as accurate. Hence it seems correct to say that:

> (1) If we are to see someone as possessing intentions, her knowledge of her own intentions must generally be unchallengable.

One way of explaining this unchallengability is by seeing it as based on some kind of unusually accurate perceptual mechanism or on a form of immediate knowledge such as that which we seem to have of our own experiences.[3] These responses take at face-value our apparent inability to make sense of the negation of (1)'s consequent and attempts to explain this by postulating some inner mechanism. I suggest, however, that we *can* make sense of this negation, but that we can only do so by also denying the antecedent. Someone's self-ascriptions can be challenged but only by our ceasing to treat those utterances *as* self-ascriptions (and the subject as a possessor of intentions).[4] The mysterious necessity of (1)'s consequent vanishes when we focus on its antecedent, on what a big "if" (1) involves. When we think through examples of what a general failure in self-understanding is like, it becomes apparent that what looked like some strange metaphysical necessity was the inter-dependency between our interpreting someone as possessing intentions and our interpreting certain of her acts as statements of intention. If we frequently conclude that a person intends to do x while at the same time concluding that she believes she will perform not-x, we uncover not a general inaccuracy of judgment but either a failure to understand the person on our part or a disintegration of intelligible agency on hers.

This inter-dependency between the possession of certain intentional states and our first-person "knowledge" of those states is part of a larger pattern which characterises our understanding of subjecthood. As a result, inaccuracy in one's self-ascriptions does not merely jeopardise the possession of the particular states of which one is claiming knowledge. The fact that someone is usually correct when, for example, he ascribes intentions to himself is not a characteristic that we can line up alongside his other competencies and qualities. The loss of this "property" runs to the very heart of its "bearer" in that this apparently simple inability ramifies wildly throughout what little life such a person can be said to have.[5] This changes the way we approach the issue of whether one needs to demonstrate that one knows one's own intentions. In certain particular cases, one may need to, but the call for a *general* justification of one's claims to know one's own intentions appears to be a request for a justification of something like one's own sanity. Since "the unjustified" (those who, for example, claim to be able to "go on" when they rarely, if ever, can) are not so much unjustified in their claims as out of their minds, it makes little sense to say that, therefore, I *am* generally justified in making these claims. Although it takes something like the discussion above to make this apparent, the very fact that I might be thought capable of formulating a justification shows that my overall ability is not in doubt.

II

Philosophical problems often support one another, in that we may be unable to see that a problem rests on a confusion as long as certain other philosophical problems remain unchallenged. Conversely, once certain problems have been "dissolved," certain others may become tractable. In this second part of the paper, I will argue that our dissolution of the problem of the authority of self-ascriptions opens the way for a dissolution of a problem central to the current debate on rule-following. The recent revival of "community views" of rule-following has been driven in large part by the claim that a serious philosophical difficulty afflicts the notion that a solitary individual can be seen as following a rule. Its capacity to solve this problem has been offered as a reason for embracing a community view.[6] Rather than endorse that conclusion, I wish to argue that, by noting parallels with the conclusions of Part I, the motivating "problem" can be dissolved.

Claiming its roots in the rule-following considerations, the "communitarian" argument I wish to examine takes as its point of departure the thesis that the identity of a rule is fixed by its application. This thesis implies that if we come upon someone we believe to be following a rule, the rule that he is following must be read off his behaviour. When we come to consider the possibility of error, however, we encounter a problem. Before we can examine the person's application of the rule for errors, we must establish what counts as correctly following that rule. That is determined by the rule but the rule is, in turn, determined by application, that is, by the person's actual applications of the rule. If our initial thesis is correct, the very idea of "successful application" has vanished because what we had thought of as the "assessment" of the person's rule-following has turned out to be a comparison of his actual performance with his actual performance. If the identity of the rule is fixed by what the person does, his *actual* performance dictates what will count as *successful* performance. "[W]hatever is going to seem right to [him] is right" and that only means that here we can't talk about 'right'" (Wittgenstein 1967 sec. 258). Without an *independent* standard by which to assess someone's behaviour, the room does not exist for the application of normative notions. "Rule" is here "just a misleading substitute for 'regularity'" (Robinson 1992 p. 339) and what we are labeling "concepts" are merely "preconceptual dispositions to respond to stimuli in various ways" (O'Hear 1991 p. 57). What looks to be a mistake may be adaptively disadvantageous, may result in the individual suffering some misfortune, but no more than bird-call can it be seen as *incorrect* if we have no criterion of correctness.[7] Faced with these considerations, the communitarian concludes that if we are to make sense of rule-following, we have to postulate the existence of a community of other rule-followers:

> There must be a use of a sign that is *independent* of what an individual speaker does with it, in order for the latter's use of the sign to be correct or incorrect. . . . This independence-condition can be satisfied only if there is a community of speakers who use the sign in a customary way. (Malcolm 1989 p. 28)

"[O]ur being able to contrast the actions of the individual with the actions of the community" (Williams 1991 p. 113) makes available a criterion of correctness. Hence "the basic normative distinction between correct and incorrect action" (p. 93) can once again be applied to the "solitary" rule-follower.

In response, Baker and Hacker (1984) argue that certain characteristic kinds of behaviour make available the concept of error. Errors are identifiable not only by reference to the (elusive) rules which govern the acts in question, but also by the "corrective behaviour" that accompanies their recognition by the rule-follower, expressions of confusion and annoyance, a shaking of the head or a stamping of the foot. Hence one can argue that, rather than there needing to be a community whose behaviour can serve as a standard by which to assess that of the individual, all we require is that the individual exhibit "*regularities* of action of sufficient *complexity* to yield normativity" (Baker and Hacker 1984 p. 42), a complexity which makes room for the range of expression, excitement, frustration and uncertainty that we associate with a rule-following creature.[8] But how much, communitarians reply, does "corrective behaviour" show? We still have no distinction between what it is to follow the rule correctly and the particular subject's understanding of that rule. A subject's "corrective behaviour" would show us no more than what he takes erroneous following of the rule to be and where there is no distinction between seeming incorrect and being incorrect, why talk about "incorrect"? As Mounce puts it, "[w]e are permitted . . . to say that he is incorrect but only if he treats it as correct to do so" (Mounce 1986 p. 193).[9]

Rather than revealing a fatal incoherence in the notion of solitary rule-following, I wish to argue that this argument rests on a doubly-confused picture of how we interpret rule-following behaviour.[10] Firstly, the communitarian argument over-estimates the bond between what the rule-follower treats on any one occasion as an error and our assessment of what his rule must be, given how he applies it. The communitarian's mistake here stems from a failure to appreciate the holistic character of interpretation. In presenting what the communitarian sees as a loop-hole in this first criticism of mine, he immerses himself in a second confusion. The gap that can be substantiated between the solitary rule-follower's actual performance and what counts as correct performance vanishes once again when we consider judgments made in ideal conditions and this, the communitarian argues, re-establishes the need for a community view. I wish to suggest, however, that this "loop-hole" merely serves to confirm a claim regarding rule-following which parallels the conclusions of Part I, namely that anyone we see as a rule-follower must manifest a degree of self-understanding.

The communitarian misinterprets the fact that only where it makes sense to talk of "errors" does it make sense to talk of "rules." What he forgets is that as soon as we set about interpreting someone as a rule-follower, we set aside a certain, as yet undetermined, portion of her behaviour as "erroneous." In interpreting her actions, we construct rough hypotheses about which rules she is following. We assess the accuracy of these hypotheses by assessing how accurately they predict her behaviour. However, a certain number of those acts of hers that clash with what our postulated rules predict can, under appropriate conditions, be discounted. We discount them as *errors*. The presence of "corrective behaviour" is one circumstance that we may cite in justifying this interpretation. But there are others and these allow us to postulate *undectected* errors, the phenomena of which the communitarian declares we cannot make sense in cases of solitary rule-following.

Let us imagine a case. On the basis of our observations of his behaviour we have arrived at a formulation of a rule which we believe an individual is following. In so doing, we have relied upon his manifestation of "corrective behaviour" and this has allowed us to discount certain acts as "mistakes." Now we encounter an act which appears to be an instance of the attempt to follow the rule in that it takes place in similar circumstances, in response to similar events etc.. No "corrective behaviour" is manifest but in this case the act is *not* that which is called for by the rule we have postulated. At this point, the communitarian insists we must postulate a different rule. But this is to forget the holistic character of interpretation. Faced with this unexpected outcome, we broaden the context, asking, for example, whether the act takes place in circumstances similar to those in which the subject has in the past committed what we have interpreted as "errors." A host of other questions arise. "How well founded are our generalizations based on previous instances of 'corrective behaviour'?" "Could some of these instances have been *mistaken* 'corrections', for example?" "And by how much must we complicate the rule in order to understand this act as a correct response?" On the basis of the answers to these questions and others like them, we can choose either to abandon our postulated rule and substitute a more complicated one, *or* to ascribe to the individual the committal of an unnoticed error. The grounds are, of course, defeasible but to deny them the very status of being grounds is to treat the evidential *slack* between overt behaviour and its intentional description as an evidential *gulf*. It is true that we cannot conclusively identify what counts as an error without knowing what the relevant rule is, and that we cannot conclusively identify the rule without knowing which acts are errors. But the communitarian is mistaken in suggesting that this closed circle prevents "rule" from gaining a foothold. Rather, this circle is a reflection of the fact that our ascription of errors to a subject and our postulation of rules that we believe she is following take place within a broader, holistically-structured interpretation of her behaviour.

The communitarian mistakenly assumes that one must determine which acts are "errors" and which "correct" *in advance* of determining "the rule followed." Without this independent specification, we may seem to be involving the very conclusion we wish to establish in the interpretation of our "evidential base." We seem to be tailoring our "findings" to the conclusion we wish to reach. However, we engage in this "malpractice" merely because the fixing of what counts as a "rule," a "correct application," an "error," a "correction" and an "unnoticed error" happens, in a sense, *simultaneously*. What counts as the correct application of the subject's rule is indeed read off her behaviour but not in the straightforward way that the communitarian imagines. Any one particular application of her rule is mediated by, among other things, her beliefs about the situation in question and sometimes conditions will be such that it makes more sense, given how we have so far interpreted her acts, to ascribe to her an error even though she manifests no "corrective behaviour." No finite set of characteristics exhaustively and exclusively specifies those acts which are instances of errors, but we may still have (defeasible) grounds for applying the concept of "error" (and hence "rule") to a solitary rule-follower.

When we examine how we actually go about interpreting the behaviour of a rule-follower, we see that our communitarian argument uncovers nothing more than the holistic character of intentional description. To insist that the argument supports a community view we would now have to deny the truth of psychological holism, a widely-accepted view that enjoys much independent support.[11] There may appear, however, to be a loop-hole left open by this response to the communitarian. It is this loop-hole which leads on to my second criticism of the communitarian argument and back to the conclusions of Part I. The objection is that, even accepting the holistically-mediated character of error-ascription, in certain cases we *are* obliged to take what the subject actually does as fixing what constitutes the correct following of the rule. We may be able to argue that certain acts constitute unnoticed errors but this seems to assume that in the cases in question, were the subject's attention drawn to the error, he would revise his judgment. Doesn't this suggest that we are permitted to say that he is incorrect only if he ultimately treats it as correct to do so? Even if we accept that our application of "rule" and "error" are holistically-mediated, in the crucial case in which conditions are ideal and time for reflection has been more than adequate, the arbitrator on what counts as correct following of the rule would still be the subject himself, and we seem unable to construct here a distinction between "is right" and "after careful reflection, ultimately seems right." Mounce argues that, without such a distinction, the "normative" can amount to no more than the merely regular. Regarding a child learning mathematics who repeatedly gives what we would call "wrong answers," Mounce argues:

> We may call him wrong even if, after the fullest reflection, he *does* feel compelled to proceed in that way. This is possible because he is being trained in a practice that exists independently of him, or to put it another way, because it is *not* only his own thoughts and movements which are relevant in determining the correctness of what he does. (Mounce 1986 p. 194)

If we can say that someone is trying to follow the same rule as the community at large, we will appear to have freed ourselves from the situation in which "there seems no difference between saying our individual proceeds correctly and saying he proceeds as, on reflection, he feels inclined or compelled to proceed" (Mounce p. 193).

Once again, however, the "problem" the community view claims to solve is nothing of the sort. The form of independence which Mounce believes only a community view can accommodate (independence from the subject's ideally reflective best judgment) in fact plays no significant role in our understanding of rule-following. To think otherwise is to forget that people can only be said to be failing to follow a rule if that rule has some sort of anchor in them. They may not act in accordance with a rule about which they do not know, but they cannot be said to fail to follow such a rule. If, in ideal conditions, an individual shows no inclination to revise his "non-standard judgment," we are then inclined to say that that is what *his rule* prescribes, our rule having become irrelevant. Under such circumstances, we would conclude that, despite its unusual character, what a subject seems to be presenting as his rule is indeed the rule he chooses to follow.

For error and correctness to be conceivable, a certain form of independence must hold between a rule and the behaviour it governs. My first criticism of the communitarian argument was that a non-communitarian can accommodate this species of independence. What is then offered as a remaining advantage of the community view is its ability to explain a further form of independence, allowing a subject to be still (totally ineptly) following a rule of which he shows no recognition whatsoever. My second criticism of the communitarian is that in such cases, the subject simply isn't following that rule, ineptly or otherwise. Mounce presents a case in which, without the aid of the community view, we seem to have no choice but to take what the subject says is right as being right. I am suggesting that, in such cases, that *is* the appropriate response. The kind of incompetence that Mounce feels the need to accommodate (the possibility of the subject's ideally reflective judgment about the rule he follows being mistaken) is not a kind that rule-followers display. What is here presented as a kind of error in the following of a rule is, at best, the following of some other rule and, at worst, a failure to follow any rule at all. A certain degree of competence is necessary for us to ascribe to someone the mere effort to follow a rule, and what Mounce claims the non-communitarian cannot explain is the rule-follower's continuing to follow his rule even when that minimal grasp has been *lost*.

That the ultimate authority on what counts as the correct following of a rule is the individual rule-follower himself is a reflection of the fact that a certain degree of understanding of his own rules is a requirement if we are to view people as attempting to follow those rules. This does not, however, imply that we must accept whatever a person says on that topic because one interpretative possibility is that we cease to regard them as rule-followers. If a subject appears to insist that her acts instantiate a particular rule even though every careful observer cannot see how that could possibly be the rule that she is following, doubts emerge as to whether what we understood to be self-ascriptions were indeed self-ascriptions. That is to say, there must be a certain minimal coherence between the application a subject makes of a rule and her own statements regarding its meaning if we are to continue to see these acts *as* "applications of a rule" and "statements of a rule's meaning." Consider a corresponding case. If someone performs an act out of line with a rule that we follow, we can assess whether he is making an error (that is to say, whether he is attempting to follow *our* rule) by examining his other actions and reactions. If the evidence that we gather suggests that he *is* attempting to follow our rule, is free of sensory and memory problems, but retains his "erroneous" judgment in ideal conditions, then we will be inclined to think that he is just plain crazy. Faced with such a result, we start to reassess what we thought was evidence that he was trying to follow our rule. One thing we do not conclude is that *we*, therefore, are the ultimate authorities on which rule he is following (basing our judgment on, perhaps, his application of the rule rather than on the account he offers). The crucial question here is: Authorities on *what*? There would appear not to be any rule-following *going on* in such cases. Our reassessment of what we had taken to be evidence that the subject was trying to follow our rule is that this was not the following of a rule at all.

There is then something misleading about the tension that might be thought to exist between one's first-person authority in explaining the rules one follows and the claim that the identity of a rule is dependent on its application. Gripped by the thought that these two forms of "evidence" might come into conflict with each other, we could attempt to construct a theory which would show that one is derived from the other. The need for such a theory may seem all the more desperate precisely because the rule-following considerations suggest that any attempt to derive use from some kind of meaning-constituting structure (with which we might have immediate acquaintance) is doomed to failure and phenomena like "understanding in a flash" seem to rule out self-ascriptions being derived from the application of rules (*ex hypothesi*, the subject has not made any correct applications of the relevant rule up until that point).

Fortunately, the above discussion renders the inadequacies of these derivations harmless by dissolving the problem they attempt to solve. The mistake is to see the individual's authority as some sort of privileged awareness which could come into cognitively-construed conflict with the "evidence" of use. Rather this authority is a *presupposition* of our seeing his utterances *as* used, as anything more than noises. Our interpretation of one set of acts as "statements of a rule's meaning" and another set as "applications of that rule" is only tenable as long as these acts, so interpreted, remain generally consistent with each other and, hence, it is only under certain very specific circumstances that it makes sense to over-rule a subject's own account of the meaning of his utterances by citing our observations of how he has been using those utterances. We can only do so if we thereby uncover another coherent pattern of thought and action which we can recognize as that of an intelligible agent. For example, we may come to see his apparently incongruous self-ascriptions as instances of lying or of self-deception. Or we may light upon a wholly different interpretation of what we had taken to be the self-ascription of a particular rule. But if, on examining the subject's broader pattern of behaviour, we cannot construct this kind of alternative reading (and still cannot accept the subject's utterances at face-value), we begin to wonder what on earth it is that we are observing. Here the "evidence" provided by use points not to a meaning other than that which the "subject" ascribes to himself. Rather it points to a breakdown in meaningful action. Moreover, this conclusion does not show that use, rather than self-ascription, is the "true determinant" of meaning. When use undermines self-ascription in this way, that very use also appears in a whole new light. Just as the over-ruled self-ascriptions are denied the status of actually being self-ascriptions, so too what looked like the meaningful use of an utterance is now revealed to be nothing but a pattern of inexpressive movement. Here we have reached the hazy region at the edge of the holistically-structured web of intentional understanding where rule-governed behaviour merges into mere regularity.

A large part of the recent literature on rule-following fails to acknowledge these lessons. It is undeniable that someone's following of a rule requires that what she happens to do should not dictate what the rule prescribes. On the basis of this fact, it is argued that any instance of rule-following involves a community of rule-followers by reference to which individuals could be said to be correctly or incorrectly following their rules. But once we acknowledge the holistic character of intentional description, this communitarian argument can only establish its conclusion by failing to see the nonsense of a community's dictating to an individual who shows no recognition of the existence, let alone authority, of the rule the community follows. To continue to insist on this "possibility" is to fail to see that individuals must be credited with a degree of self-understanding if we are to depict them as rule-followers. This does not imply that anything they say goes. Certain patterns of behaviour will still undermine their status as rule-followers. But if we wish to retain our conception of them as rule-followers, we must be prepared to extend to them a degree of authority over the rules they are said to follow. Such authority is an ineliminable part of the structure we label "the following of rules."[12]

Denis McManus

Emmanuel College, Cambridge
and Harvard University

NOTES

[1] The following argument is a detail of a larger picture dominated by the rule-following considerations. In order to focus on the particular issue of self-ascription, I make the simplifying assumption that the meaning of a person's utterances is generally evident in her actions, reactions and other utterances.

[2] Unlike the kind of radical dissociation discussed here, self-deception brackets an extremely small number of instances of a very particular kind of claim to self-knowledge and, in doing so, relies on the subject's self-knowledge remaining, for the very much greater part, intact. Significantly, Wittgenstein attacks those conceptions of the unconscious which suggest that the above case might be construed as self-deception, namely those which understand unconscious thoughts as self-contained entities which can become radically divorced from any "surface" expression (cf. Wittgenstein 1966 pp. 46–47, 49, 51–52 and 1969 pp. 22–23, 57–58).

[3] I ignore in this paper Wittgenstein's criticism of the notion that our relation to our own experiences is best thought of as a form of knowledge, in the light of which the modeling of "our knowledge of meaning" on "our knowledge of our experiences" appears doubly confused.

[4] These conclusions only cover, of course, the extreme case in which self-ascriptions are *generally* over-ruled.

5. Ethical standards and the philosophy of mind overlap here, in that while a certain level of consistency and reliability is expected of us as morally good agents, a certain level is also required of us in order to qualify simply as agents.

6. While sharing the belief that reference to a community is essential in any intelligible case of rule-following, the "revival" to which I am referring is characterised by different arguments and some different conclusions from those of earlier "community views," such as those of Wright (1980), Peacocke (1981) and Kripke (1982). Some of these "novelties" were, however, anticipated by Rhees (1954).

7. Cf. also Mounce 1986 pp. 195–96, Trigg 1991 p. 210 and Williams 1991 p. 113.

8. Compare Blackburn 1984 pp. 289–91.

9. Cf. also Malcolm 1989 pp. 22, 28, O'Hear 1991 pp. 47–51 and Trigg 1991 p. 211.

10. In criticising this argument, I am attacking only one possible basis on which a community view might be erected. That there may be other viable bases is not an issue I can assess here.

11. For Davidson's original statement of this view, cf. Davidson 1973 and 1974.

12. I would like to thank Stanley Cavell, Jane Heal and Avrum Stroll for their helpful comments on an earlier draft of this paper. This paper was written during a period of research made possible by a Herchel Smith Scholarship. I would like to express my appreciation to Dr. Herchel Smith and the Scholarship Committee.

REFERENCES

Baker, G. P. and Hacker, P. M. S. 1984 *Scepticism, Rules and Language*, Oxford: Blackwell.

Blackburn, S. 1984 *Spreading the Word*, Oxford: Clarendon Press.

Davidson, D. 1973 "Radical Interpretation," *Inquiries into Truth and Interpretation*, 1984, Oxford: Clarendon Press, 125–40.

____ 1974 "Belief and the Basis of Meaning," *Inquiries into Truth and Interpretation*, 1984, Oxford: Clarendon Press, 141–54.

Kripke, S. A. 1982 *Wittgenstein on Rules and Private Language*, Oxford: Blackwell.

Malcolm, N. 1989 "Wittgenstein on Language and Rules," *Philosophy* **64**, 5–28.

Mellor, D. H. 1978 "Conscious Belief," *Proceedings of the Aristotelian Society*, 87–101.

Mounce, H. O. 1986 "Following a Rule," *Philosophical Investigations* **9**, 187–98.

O'Hear, A. 1991 "Wittgenstein and the Transmission of Traditions." *Wittgenstein Centenary Essays*, ed. A. Phillips Griffiths, Cambridge: Cambridge University Press, 41–60.

Peacocke, C. 1981 "Rule-following: The Nature of Wittgenstein's Arguments," *Wittgenstein: To follow a Rule*, eds. S. H. Holtzman and C. M. Leich, London: Routledge and Kegan Paul, 72–95.

Rhees, R. 1954 "Could Language Be Invented by a Robinson Crusoe?," *The Private Language Argument*, ed. O. R. Jones, London: Macmillan, 1971, 61–75.

Robinson, G. 1992 "Language and the Society of Others," *Philosophy* **67**, 329–41.

Trigg, R. 1991 "Wittgenstein and Social Science," *Wittgenstein Centenary Essays*, ed. A. Phillips Griffiths, Cambridge: Cambridge University Press, 209–22.

Williams, M. 1991 "Blind Obedience: Rules, Community and the Individual," *Meaning Scepticism*, ed. K. Puhl, Berlin: de Gruyter, 93–125.

Wittgenstein, L. 1966 *Lectures and Conversations on Aesthetics, Psychology and Religious Belief*, ed. C. Barrett, Oxford: Blackwell.

____ 1967 *Philosophical Investigations*, ed. G. E. M. Anscombe and R. Rhees, trans. G. E. M. Anscombe, Oxford: Blackwell.

____ 1969 *The Blue and Brown Books*, Oxford: Blackwell.

____ 1971 "Notes for Lectures on 'Private Experience' and 'Sense Data'," ed. R. Rhees, *The Private Language Argument*, ed. O. R. Jones, London: Macmillan, 229–75.

____ 1977 *Remarks on Colour*, ed. G. E. M. Anscombe, trans. L. L. McAlister and M. Schättle, Oxford: Blackwell.

Wright, C. 1980 *Wittgenstein on the Foundations of Mathematics*, London: Duckworth.