The Reality of Rule-Following

Philip Pettit

Mind, New Series, Vol. 99, No. 393. (Jan., 1990), pp. 1-21.

Stable URL:
http://links.jstor.org/sici?sici=0026-4423%28199001%292%3A99%3A393%3C1%3ATROR%3E2.0.CO%3B2-O

Mind is currently published by Oxford University Press.

The Reality of Rule-Following

PHILIP PETTIT

Drawing on Wittgensteinian materials, Saul Kripke has raised a problem for anyone who thinks that we follow rules, say rules of meaning, in the ordinary sense of that phrase: the sense in which it suggests that rules are entities we can identify at a time and form the intention of trying to honour thereafter.[1] He has presented a sceptical challenge to the idea of rule-following, elaborating—if not wholly endorsing—arguments which purport to show that the idea is rooted in illusion.

I believe that this challenge is of the greatest importance in the philosophy of mind, though many practitioners seem to think that they can ignore it. I argue that the challenge can be met and the reality of rule-following vindicated. But I show that in order to meet it in this way, some quite dramatic shifts have to be made in the ways of conceiving mentality that have become standard among philosophers and psychologists.

The paper is in four parts. In the first I give a characterization of rules and rule-following, trying to show how central they are in our everyday thought about ourselves. In the second I present the sceptical challenge, drawing heavily on Kripke's work; I exercise some license here, since I do not aspire to be an exegete either of Kripke or of Wittgenstein. In the third section I offer my response to the challenge, outlining a non-sceptical conception of rules and rule-following. In the fourth section I look at three corollaries of the response. And then in the last section I buttress the response, showing how the non-sceptical conception can be extended to encompass public as well as private rules.

One word of warning, in case too much is expected. I formulate the problem of rule-following, and I propose a solution, on the assumption that a creature which does not follow rules in my sense, and is not therefore a speaker or thinker—see section 1—may yet be capable of having beliefs, desires, and intentions, including beliefs, desires, and intentions directed to others of its kind: it may yet be an intentional and even social subject. Those who want to ascribe intentionality only to speakers or thinkers will quarrel with this assumption but they will probably endorse a variant that would serve my purposes equally well: they will recognize the possibility that creatures who do not follow rules may yet display a suitably complex range of recognitional, behavioural, and information-processing skills. The important point is that rule-

[1] Saul Kripke, *Wittgenstein on Rules and Private Language*, Blackwells, Oxford, 1982; henceforth, *WRPL*. See also Robert Fogelin, *Wittgenstein*, 2nd edn, Routledge, London, 1987.

Mind, Vol. 99 . 393 . January 1990

following, as I understand it, does not encompass the full leap from protoplasm to personhood, only the transition from, as it were, sub-personal to personal mentality.[2]

1. *Rules and rule-following*

What the sceptical challenge puts in doubt is the fact, as it appears to us, that we follow rules. The notion of following a rule, as it is conceived here, involves an important element over and beyond that of conforming to a rule. The conformity must be intentional, being something that is achieved, at least in part, on the basis of belief and desire. To follow a rule is to conform to it but the act of conforming, or at least the act of trying to conform—if that is distinct—must be intentional. It must be explicable, in the appropriate way, by the agent's beliefs and desires.

But more than this is required to understand the appearance of rule-following which the sceptical challenge questions. We need to understand not just what following involves, but also what sorts of things the rules followed are supposed to be. From the viewpoint of the sceptical challenge, there are four important elements in the notion of rules; a further element will be identified later when we move to the discussion of public rules. I make no pretence at analysing the everyday notion of rules in distinguishing these elements. My analysis, which is influenced in great part by Kripke's discussion, is offered as a stipulative account.

The first and main element in the definition of rules is the stipulation that rules are normative constraints, in particular normative constraints which are relevant in an indefinitely large number of decision-types. That something is a normative constraint in a decision means that it identifies one option—or perhaps one subset of options—as more appropriate in some way than the others. The option may be the most polite, as with a rule of etiquette; the most becoming, as with a rule of fashion; the most just, as with a rule of fairness; or whatever. That a normative constraint is relevant in an indefinitely large number of decision-types means that the decisions which it is capable of constraining are not limited to sorts of situations which the rule-follower can use an effective procedure to specify independently in advance. Most familiar rules involve normative constraints which are relevant in an indefinitely large number of decision-types. There is no suitable specification of the types of situation where most rules of etiquette or ethics apply and certainly no such specification of the circumstances where the rules governing normal word-usage apply; the point is made vivid, if that is necessary, by Wittgenstein on family resemblance.

[2] I defend the assumption mentioned in this paragraph in a work under preparation—*The Common Mind: From Folk Psychology to Social and Political Theory*, Oxford University Press, forthcoming. I also mark the distinction there between the problem of rule-following and the problem of content that arises for unthinking intentional subjects.

This first element in our definition means that a rule is a function which can take an indefinite variety of decision-types as inputs and deliver in each case one option—or set of options—as output: this is the option that is identified as the most appropriate in some way. Consistently with meeting this condition a rule may be a total function over all decision-types—a function yielding an output in each case—or a partial function that only yields an outcome in some cases. An example of a suitable function would be the indefinitely large set of pairs, one for every decision-type, which each involve, first, a relevant decision-type and then the option appropriate for that type. We might refer to such a set as the rule-in-extension. Another example of a suitable function would be the abstract object which is conceived as having the property of identifying the appropriate option for every relevant decision-type to which it is applied. We call this the rule-in-intension, though it might be more familiar under a name like 'universal' or 'concept' or 'property'.

The other three elements in the definition of a rule can be derived from this first element, together with the assumption that the rule is capable of being followed. They are requirements that an indefinitely normative constraint—a rule in the objective sense—must satisfy, if it is to make sense for finite subjects like you and me to try to conform to it.

The first of the additional elements is the requirement that not only should a rule be normative over an indefinite variety of applications, it should be determinable or identifiable by a finite subject independently of any particular application: the prospective rule-follower should be in a position to identify the rule in such a manner that he can sensibly try to be faithful to it in any application. If the rule were identified by reference in part to how the subject responded in a given case, then the subject could not see the rule as something to which he should try to be faithful in that case. He could not see it as a normative constraint for him to try to respect there.

The other two additional elements in our account of rules require respectively that a rule must be directly readable and fallibly readable. That a rule is directly readable means that the competent rule-follower can tell straightaway what it apparently requires or, if he tells what it requires by applying other rules, that these are ultimately rules whose apparent requirements he can tell straightaway. That a rule is fallibly readable means that no matter how directly the rule speaks to him, no matter how quickly he can tell what it apparently requires, that fact alone does not provide the rule-follower with an epistemic guarantee that he has got the requirement of the rule right. He can understand properly the situation on hand, seeing clearly the options before him, and yet for all that shows, fail to read the rule properly. Thus, in one sense at least, he is not an infallible authority; there is an epistemic possibility of his going wrong.[3]

[3] In another sense he may be: he may be designed so that he always reads the rule right as a matter of fact.

It is clear that an infinitely normative constraint must be identifiable independently of any application, if it is to make sense for a finite subject to try intentionally to conform to it. But the fulfilment of that condition also requires that the rule be directly and fallibly readable. The rule-follower must be able to tell straight off what the rule apparently requires or to tell what it requires by applying rules such that ultimately he tells straight off what they apparently require. How else could he intentionally try to conform? And the rule-follower must be able to tell only fallibly what a rule requires. Otherwise the notion of intentional action, in particular the notion of trying, would be out of place.

This account of rules suffices, I hope, to give some sense of the apparent fact about ourselves which the sceptical challenge is designed to put in doubt. The fact under challenge is that we intentionally try to conform to rules: that we intentionally try to conform to indefinitely normative constraints that are independently determinable, directly and fallibly readable. Before going to the sceptical challenge however it will be useful to consider two respects in which we are required, it seems, to be capable of following rules: first, so far as we are speakers, and secondly—this is more contentious—so far as we are thinkers.

The case of speech is the one that is most commonly mentioned. The situation, as acknowledged on all sides, is that when I grasp the meaning of a word, say by examples of its usage, I put myself in touch with a rule which I am then in a position to intend to honour in future cases. The meaning normatively constrains usage over an indefinite variety of cases. It is determinable independently of any particular case. And from the point of view of someone like me who has just grasped it, the meaning is directly but fallibly readable. Thus consider the case on which Kripke focuses, in which a few examples of addition enable me, or so I feel assured, to grasp the meaning of 'plus'. 'I feel confident that there is something in my mind—the meaning I attach to the "plus" sign—that instructs me what I ought to do in all future cases. I do not *predict* what I *will* do . . . but instruct myself what I ought to do to conform to the meaning.'[4]

Although it is generally conceded that we are required to be able to follow rules so far as we speak, it is not always recognized that we also seem required to be able to follow rules if we are to think.[5] Thinking requires more than just the having of intentional attitudes: attitudes, as most people are prepared to describe them, of belief and desire. A system has such attitudes, I am prepared to say, so far as its behaviour is non-redundantly explained by their presence.[6] The behaviour is intentional, being produced in each case—and produced in the right way—by the

[4] Kripke, *WRPL*, p. 22.

[5] But see Colin McGinn, *Wittgenstein on Meaning*, Blackwells, Oxford, 1984, pp. 144–6.

[6] See Frank Jackson and Philip Pettit 'Functionalism and Broad Content', *Mind*, 1987, pp. 381–400 and 'In Defence of Folk Psychology', *Philosophical Studies*, forthcoming.

desire for a certain state of affairs and the belief that doing this or that offers the best promise of desire-satisfaction. But a system that is intentional in this sense need not be cogitative or thoughtful.

Thinking involves not just having intentional attitudes, but intentionally shaping those attitudes: say, shaping them with a view to having beliefs that are adequate for certain projects, or beliefs that are true. The thinker must be able to wonder whether something is so, to institute tests to see whether it is so or not, to accept in the light of those tests that it probably is, and so on. He may or may not conduct these activities very explicitly of course and the only sign that he is thoughtful, one to which Donald Davidson draws attention in another context, may be that he shows surprise at the appearance of this or that piece of evidence.[7] That a subject shows such surprise means, plausibly, that he had a belief whose content was that the potential content of another belief is likely to be true, is supported by the evidence so far, or whatever; previously he believed that it is likely that *p* and he is surprised because it turns out that not *p*.[8] If a subject has beliefs about contents in this way, as distinct just from beliefs with contents, then he is able to wonder about whether such contents—such propositions—are true, able to desire that he have beliefs with contents that are true, and the like. In short he is able, in our sense, to think. He is a cogitative system, not just an intentional one.

Thinking in the sense involved here seems to require, like speech, the capacity to follow rules. This is not surprising, since in this sense thinking conforms to the shape of what has traditionally been regarded as inner speech, discourse with oneself. The thinker who wonders what is the sum of two numbers is in exactly the position of the speaker who sets out to apply the word 'plus' properly. He is required, it seems, to have identified something that serves as a normative constraint in the determination of this, and an indefinite variety of other sums; something determinable in advance of any particular application; and something that he can directly, if only fallibly, read. His problem is to identify in the case on hand the answer that is fixed by the concept of addition; that concept constitutes a rule and his problem is to remain faithful to it in performing the computation.

If speech and thought involve rule-following, then the magnitude of the challenge discussed in the next section can hardly be overstated. Deny that there are such things as rules, deny that there is anything that counts strictly as rule-following, and you put in jeopardy some of our most central notions about ourselves. More than that, you also put in jeopardy our

[7] 'Rational Animals', in Ernest Le Pore and Brian McLaughlin (eds), *Action and Events*, Blackwells, Oxford, 1985. Davidson wishes to make the possibility of surprise a criterion, not of thought, but of belief.

[8] If surprise at evidence does not require such a belief about '*p*', then of course it is not a sign of the capacity to think.

notion of the world as requiring us, given our words and concepts, to describe it this way rather than that; you undermine our conception of objective characterization. There is no extant philosophical challenge that compares on the scale of iconoclasm with the sceptical challenge to rule-following.

2. *The sceptical challenge*

I can be brief in stating the sceptical challenge to rules and rule-following, since the challenge has been well elaborated by Kripke.[9] Only a difference in emphasis separates my version of the challenge from his. He tends to ask after what fact about a person could constitute his following a rule whereas I shall ask after what sort of thing could constitute a rule that the person might follow.[10] But this shift of emphasis does not beg the question against any possible resolution of Kripke's problem; thus I remain open to the possibility that there is something to constitute rule-following without there being anything to constitute a rule. The shift of emphasis is designed to link up smoothly with our discussion in the last section.

Among the elements invoked in the definition of a rule, there is a salient distinction between the first element and the other three. The first tells us about what objectively, so to speak, a rule is. It is a constraint that is normative over an indefinite variety of cases; in effect, or so it would seem, it is a rule-in-extension or rule-in-intension. The other three elements tell us what an objective rule must be to engage subjectively with potential followers. It must be identifiable independently of any particular application, it must be directly readable, and it must be fallibly readable.

The sceptical challenge to rules is best presented as a challenge to identify anything that could simultaneously satisfy the objective and subjective elements in the definition of a rule. What sort of thing could be indefinitely normative and engage in the manner required with finite minds like ours? Putting the question the other way around, among the things that engage appropriately with our minds, what sort could serve as an indefinitely normative constraint?

Take the sorts of entities which we know to satisfy the objective condition: the rule-in-extension and the rule-in-intension. The rule-in-extension does not seem capable of satisfying the subjective conditions, because it is liable, as in the case of 'plus', to be an infinitely large set. 'The infinitely many cases of the table are not in my mind for my future self to consult.'[11] There is no way that I could get in touch appropriately with such an infinite object. Or so it certainly seems.

[9] *WRPL*, ch. 2.

[10] We may, in doing this, be sticking more closely to Wittgenstein. See Marie McGinn, 'Kripke on Wittgenstein's Sceptical Problem', *Ratio*, 1984, pp. 19–32.

[11] Kripke, *WRPL*, p. 22.

What of the rule-in-intension? What, for example, of the addition function, as Frege would conceive of it, which determines the correct option in any decision about the sum of two numbers?[12] What is there against the idea that this abstract object might be able to satisfy the subjective conditions, engaging our minds appropriately? Here the problem is to explain how we are able to get in contact with such an abstract object. It does not affect our senses like a physical object and so we are not causally connected with it in the ordinary way. So how then does it become present to our minds? The obvious sort of answer is to say, like Frege, that it does so via an idea—or some such entity—that we can contemplate. But then the suggestion boils down to one that we consider in a moment and find wanting.[13]

Moving from the entities which can clearly satisfy the objective condition on a rule to entities that look more likely to be able to satisfy the subjective conditions, the question here is whether such entities can be objectively satisfactory: whether they can serve as normative constraints over an indefinite variety of cases. Kripke mentions two main candidates for entities of this kind: first, actual or possible examples of the application of the rule in question, such as examples of addition; and secondly, introspectible states of consciousness, as for instance the sort of *quale* which might be thought to be associated with adding numbers together. There is a special problem with the second candidate, which is that often no plausible *quale* is available.[14] But, more importantly, there is an objection that applies equally to both candidates, so Kripke argues, and indeed to any finite object that is proposed for the role in question.[15]

The objection, and this is clearly derived from Wittgensteinian materials, is that no finite object contemplated by the mind can unambiguously identify a constraint that is normative over an indefinite variety of cases. Consider a series of examples of addition: $1+1=2$, $1+2=3$, $2+2=4$, and the like. For all that any such finite object can determine, the right way to go with a novel case remains open. 'Plus', as we understand it, forces us to say that $68+57=125$ but the examples given do nothing to identify the plus-rule as distinct from, say, the quus-rule, where this says that the answer in the case of 68 and 57 is 5. The fact is that any finite set of examples, mathematical or otherwise, can be extrapolated in an infinite number of ways; equivalently, any finite set of examples instantiates an infinite number of rules.

It appears then that I cannot be put in touch with a particular rule just on the basis of finite examples.

When I respond in one way rather than another to such a problem as '$68+57$', I can have no justification for one response rather than another. Since the sceptic

[12] Kripke, *WRPL*, p. 53. [13] Kripke, *WRPL*, p. 54. [14] Kripke, *WRPL*, p. 43.
[15] Kripke, *WRPL*, p. 43.

who supposes that I meant quus cannot be answered, there is no fact about me that distinguishes between my meaning plus and my meaning quus. Indeed, there is no fact about me that distinguishes between my meaning a definite function by 'plus' (which determines my responses in new cases) and my meaning nothing at all.[16]

The problem raised extends to *qualia*. 'No internal impression, with a *quale*, could possibly tell me in itself how it is to be applied in future cases.'[17] If the impression has a bearing on future cases, say on the application of 'plus', it will be capable of being extrapolated in any of an infinite number of ways. How then am I supposed to grasp a particular rule in contemplating the impression? How is the impression supposed to make salient just one of the infinite number of rules that it might be held to illustrate? The question extends from qualitative impressions to all mental objects of contemplation, including the sort of idea postulated by Frege. 'The idea in my mind is a finite object: can it not be interpreted as determining a quus function, rather than a plus function?'[18]

The upshot of these considerations is that rules are, at the least, extremely mysterious. They are required to satisfy two sets of conditions, objective and subjective, which no familiar sort of entity seems to be capable of simultaneously satisfying. A number of responses are possible at this point. One is to go sceptical and deny that there are rules. A second is to go dogmatic and, insisting that of course there are rules, argue that they are *sui generis*.[19] Such responses are not attractive however and so we shall look again in the next section for some way around the challenge.

But before leaving this section we must mention the response to his challenge on which Kripke spends most time. This response says nothing on what rules are but still insists that there is such a thing as rule-following. It identifies following a rule with displaying a disposition to go on after a certain pattern, say a pattern in applying the word 'plus' to new cases. I will not delay over this theory since, while it attracts a variety of criticisms from Kripke, the basic flaw is already crippling. The theory does nothing to explain how in following a rule I am directly but fallibly guided by something which determines the right response in advance. A disposition may determine what I do but it cannot provide this sort of guidance. 'As a candidate for a "fact" that determines what I mean, it fails to satisfy the basic condition on such a candidate, . . . that it should *tell* me what I ought to do in each new instance. Ultimately, almost all objections to these dispositional accounts boil down to this one.'[20]

[16] Kripke, *WRPL*, p. 21. [17] Kripke, *WRPL*, p. 43. [18] Kripke, *WRPL*, p. 54.

[19] See for example Warren Goldfarb, 'Kripke on Wittgenstein on Rules', *Journal of Philosophy*, 1985. A third response, to which Peter Menzies has drawn my attention, would be to argue that there are two distinct conceptions of rules corresponding to the two sorts of conditions.

[20] Kripke, *WRPL*, p. 24.

3. *A non-sceptical response*

Any non-sceptical response to the challenge about rules has to vindicate the idea that we intentionally try to conform to entities that satisfy the objective condition: constraints that are normative over an indefinite variety of cases. Let us assume then that if we follow a rule we are indeed put in touch with an entity of this kind. We can think of it as a rule-in-extension or a rule-in-intension.

The question that arises under this assumption is how a rule-in-extension or rule-in-intension—henceforth I shall simply say, a rule—can satisfy the subjective conditions, being independently identifiable, directly readable, and fallibly readable. This is a question, at base, about how a rule can be suitably represented to a human subject, since there is no possibility of a rule presenting itself immediately: there is no possibility of the subject's 'mainlining' the rule. Let us concentrate then on this representational issue. In exploring the issue, we shall have in mind rules of a kind that can be identified and read without the application of other rules. If the issue can be solved for such simple rules, as we may call them, it can be solved for more complex ones.[21]

What material, material directly accessible to the human subject, could serve to represent a rule, in particular a simple rule? The outstanding candidate is: examples of its application. The plus rule might be represented then as the (1, 1, 2)–(1, 2, 3)–(2, 2, 4) rule, the rule for chair as the (X)–(Y)–(Z) rule, where X, Y, and Z are all chairs, and so on. It appears however that this candidate has already been ruled out. Any finite set of examples instantiates an indefinite number of rules, as we saw in the last section. And does not that mean that no set of examples can represent a determinate rule for an agent?

The first step towards the proposal I wish to develop here is to recognize that no, it does not necessarily mean this. Instantiation is a two-place relationship between a set of examples and a rule and it certainly has the feature of being a one–many relationship: one finite set of examples instantiates many rules. But the relationship that is of concern to us when we ask whether a finite set of examples can represent a determinate rule is not instantiation but exemplification. Exemplification is a three-place relationship, not a two-place one: it involves not just a set of examples and a rule but also a person for whom the examples are supposed to exemplify the rule.[22] Although any finite set of examples instantiates an indefinite number of rules, for a particular agent the set may exemplify just one rule. Nothing has been said at least to disallow this possibility.

The second step in developing the proposal I wish to defend is to see

[21] I do not assume that the simple–complex distinction is invariant across persons or for a single person across times. Division of linguistic labour argues against the first sort of invariance, conceptual development against the second.

[22] See Nelson Goodman, *Languages of Art*, Oxford University Press, 1969.

how that possibility might be realized. Suppose that on being presented with a set of examples, an agent develops an independent disposition or inclination to extrapolate in a certain way to other cases: an inclination of which he may or may not be aware. That set of examples will continue to instantiate many rules but the rule it will then exemplify for the agent will certainly be a rule associated suitably—we come back to this in the next step—with the inclination generated by the examples. If she uses the examples to pick out a rule for herself—if she refers to that rule, the one that goes (1, 1, 2), (1, 2, 3), (2, 2, 4), and so on—she will certainly have in mind that rule among the rules instantiated by the examples which her inclination makes salient. We know of course, and indeed we recognized this in the last section, that human agents who claim to pick up rules by ostension, by the use of examples, certainly develop independent inclinations to carry on in a particular way when they are exposed to such examples. Thus we now see that there really is a possibility that a finite set of examples can exemplify a determinate rule for a human agent: it can exemplify the rule that is suitably associated with the inclination generated by the examples.

It is commonly recognized that the inclination involved in following any rule plays a role in prompting the agent's case by case responses. I do not reject that observation, though I did argue in the last section that following a rule must involve more than just indulging such a disposition: otherwise there would be no question of taking one's guidance directly but fallibly from something that determines the right response in advance. What we have now been led to see however is that the inclination involved in following a rule may have a dual function, serving not only to prompt the agent's responses, but also to make salient the rule she intends to follow: the rule which, given the inclination they engender, a certain set of examples can exemplify.

But it is important to stress one aspect of the proposal. This is that it does not require a rule-follower to have any awareness of the inclination generated by the examples that exemplify a rule, let alone to attend to that inclination in herself. I speak of the inclination making salient one of the rules instantiated by the examples, and of the agent representing the rule—via the examples—on the basis of the inclination. But none of this is meant to suggest that the rule-follower focuses on the inclination. She will focus simply on the examples and—in them, as it were—on the rule they manifest to her. The inclination explains how the examples exemplify or manifest a particular rule but it does this without having to feature in consciousness.

Perhaps the best way of casting the proposal is with the help of a familiar analogy. When I look at a physical object, all that is in one sense presented to me is a sequence of profiles: now this profile, now that, as I move around the object. Yet in experiencing those profiles I see the object

itself in the perfectly ordinary sense of that verb. I see it, as we might say, *in* the profiles. Indeed I scarcely notice the profiles, focusing as I do on the object they manifest to me. What explains how the profiles manifest *this* sort of object, conforming to the ordinary image of the middle-sized spatio-temporal continuant: *this* object, rather than any of the many ontological inventions that are strictly consistent with the sequence of profiles? Presumably something about my psychology, a disposition that I share with others of my species. This disposition may lend itself to psychological investigation but it will not be something of which I am necessarily aware.

The relevance of the analogy should be clear. As I see a particular sort of object in these profiles, so I see a particular rule manifested in such and such examples. As the profiles efface themselves in my attention, yielding centre stage to the object, so the examples command less attention than the rule they exemplify. And as the disposition which explains why I see a certain sort of object is something of which I may not be aware, so the inclination which explains why I am directed to a particular rule need not figure in my consciousness either. This analogy may be the best way of grasping the sort of proposal I am trying to develop.

We are now in a position to move to the third and most crucial step in developing the proposal. We have to identify a relationship between an inclination and a rule which would serve to save the appearance of rule-following, vindicating the claim that a finite set of examples can exemplify a determinate rule for an agent and can put her in a position to read the rule directly but fallibly. What relationship would be suitable? In order to approach an answer, notice that the sort of inclination in question serves like a description of the rule, so far as it gives putative information about the rule: the putative information that the rule requires those reponses, those ways of going on, which the inclination supports. Given that the inclination has the status of a description, we can taxonomize the salient ways in which it may relate to the rule. It may or may not be a priori true to the rule. And it may or may not be necessarily true to the rule.

The inclination will be a priori true to the rule, if the rule is this: whatever rule dictates the responses which the inclination supports. But if inclination and rule are related in this way, then the proposal must fail. Rule-following will become a matter of intentionally trying to conform to that rule, whatever it is, which is revealed by my inclination, instance by instance. It will become an enterprise in which I cannot fail, and cannot see myself as failing, contrary to the assumption that rules are fallibly readable. The question then is whether there is a suitably a posteriori relationship that might be postulated between inclination and rule. Happily there is.

If the inclination is a priori connected with the rule, then it correlates with that rule which fits it exactly: the rule correctly applied in the

responses it supports. If the inclination is to be a posteriori connected, then it must connect with a rule which is related to it in some other way, a rule which may not exactly fit it. What other way is there for a rule to relate to my inclination? It can only relate as that rule which fits my inclination but only so far as certain favourable conditions are fulfilled: in particular favourable conditions such that I can discover that in some cases they are not fulfilled, and that I got the rule wrong. The rule associated with the inclination will be that rule, the one that satisfies this inclination, provided the inclination fires under the conditions identified.

It is important to be clear about what exactly this proposal means for the first person point of view. As emphasized before, there is no suggestion that I as rule-follower am reflective about the inclination generated by the cases exemplifying the rule: I may scarcely have recognized that I have such an inclination. All that I need be aware of is that here are some examples that, so far as I am concerned, exemplify a particular rule. Which rule? *That* rule, I say, gesturing at the original examples and perhaps some others. The rule is fixed by what goes in favourable conditions with my inclination but I do not think of it in that way. So how then do favourable conditions enter my consciousness? In this way: that I will be able to admit that I may have got the rule wrong in a particular application, so far as I find that conditions were not favourable there.

In order to see that this suggestion may have something going for it, we need to recognize that the favourable conditions required do not have to be identified in advance by the subject. If they had to be, then that would make the suggestion implausible from the start. All that is necessary however is that I be in a position such that I may have to recognize after following the inclination in a given case that the response was vitiated by some perturbing conditions and was not in conformity with the rule which I represent to myself on the basis of the inclination. If I am in such a position then the inclination can serve to represent a rule with which it is associated other than by invariably supporting responses that conform to the rule.

We are pushed on by this observation to ask about how I might come to occupy a position of this kind. One obvious way, and perhaps the only conceivable way, is this. I might be committed to the principle that intertemporal or interpersonal differences in how the inclination generated by certain examples goes are a sign that perturbing influences are at play and I might generally be able to identify such influences and provide an *ex post* explanation of any difference. The inclination on the basis of which I represent a rule to myself leads me at one time to respond in one way to a certain type of decision, at another time in another.[23] Or the inclination

[23] See Simon Blackburn, 'The Individual Strikes Back', *Synthèse*, 1984, p. 294, and following on this intrapersonal case. For a critical perspective see Crispin Wright, 'A Cogent Argument against Private Language?', in Philip Pettit and John McDowell (eds), *Subject, Thought and Context*, Oxford University Press, 1986.

leads me to respond in one way, while the counterpart inclination—associated with the same generative examples—leads you to respond in another. Happily however I am able to explain the difference—I am able to find it intelligible—recognizing that a factor which is generally explanatory of differences—say, intoxication or inattention—affected me at one of the times in question, or affected one of the two of us in the interpersonal case.

Let us suppose then, in developing our proposal, that the inclination involved in rule-following connects in this a posteriori fashion to the rule it enables the agent to identify. The other question, given that the inclination has the status of a description, is whether it connects with it necessarily or contingently. It will connect necessarily if the rule is the rule which the inclination corresponds with in favourable conditions, whatever the possible world in question. In this case there will be no possibility that the inclination could fail under favourable conditions to correspond to the rule. The inclination will connect contingently with the rule on the other hand if the rule is that rule which the inclination corresponds with under favourable conditions in the actual world. This will allow for the possibility of inclination and rule coming apart, even under such conditions. There will be possible worlds where the inclination corresponds with quite different rules from that involved in the actual world.

This question is not as pressing as the issue about a priori and a posteriori status. There is no conflict between either reading of the inclination-rule relationship and the constraints on rules. But I prefer the contingent reading to the necessary one, at least in the general case. Consider that possible world where our counterparts are led by a counterpart inclination to claim that $68 + 57 = 5$. We would hardly want to say that they were being faithful to the plus-rule and yet that is what the necessity reading would entail. Under the contingent reading there is no such problem. Our counterparts are not faithful to the rule with which the inclination corresponds in the actual world and so they are simply miscounting. There may be cases where the necessary reading is less implausible, for example with colour-rules. We might accept that counterparts whose inclination led them to group green things with red were not misclassifying those things. But even here there is an intuition that after all that may be the least Pickwickian thing to say. Hence I shall generally assume that inclination relates contingently to rule, the rule being that rule with which the inclination corresponds under favourable circumstances in the actual world.

We have taken three steps in developing our response to Kripke's challenge. We have argued, firstly, that the fact that any finite set of examples instantiates an indefinite number of rules does not mean that it cannot exemplify a determinate rule for a given agent; secondly, that the set of examples can exemplify such a rule if the examples generate an inclination in the agent to go on in a certain way: the rule exemplified will

be one which is suitably associated with the inclination; and thirdly, that a suitable association between inclination and rule is this: that the rule is that rule to which the inclination corresponds in the actual world, provided the inclination operates under favourable conditions.

We know that in picking up rules from examples, human beings develop inclinations of the kind which this proposal requires. Thus the materials required for the proposal are certainly available and there is nothing to be said against the claim that it may be sound. But whether we assert that it is sound or not will depend on whether it has explanatory value: on whether, in particular, it can explain how human beings can identify a determinate rule independently of any particular application and can then read the rule directly but fallibly.

A rule will be identifiable independently of any particular application provided two conditions are fulfilled. The first, and it is surely plausible, is that no particular application has to figure among the instances that exemplify the rule for the agent. The second, which requires a little more commentary, is that there is only one rule exemplified by such examples. This will be fulfilled so long as the inclination generated by those examples is associated suitably, after standardization for favourable circumstances, with just one rule. I hold that this condition too is plausible.

The inclination invoked in any case is a currently determinate object, however standardized by the reference to favourable conditions, and it can serve in principle therefore to make it determinate which rule is the one identified by the individual subject. The rule is that which, other things being equal, the standardized inclination would identify, instance by instance.[24] True, we have to wait on the operation of the standardized inclination to see how the rule goes in new instances. But that means only that at any time we may be uncertain as to what the rule requires in new cases, not that there is an objective indeterminacy about the requirement before the case comes up for resolution. So far as there is no objective indeterminacy, the inclination enables the individual to identify a particular rule in advance of any particular application.

The other subjective conditions on a rule are that it should be directly and fallibly readable. If the rule is identified by inclination then of course there is no difficulty about how it can be directly readable. The inclination serves on our proposal, not just to identify the rule, but also to prompt the agent's responses: it has a dual function. The individual will read off the requirement of the rule in a new case by letting her inclination lead her, as with the simple rule, or by applying other rules whose requirements she ultimately reads off in that way. No mode of reading a rule could be more

[24] Does the inclination stretch to an infinite number of instances? Under idealization, yes. *Pace* WRPL, p. 27, it is not necessary to have a story about what in fact would happen if we had the unbounded memory required. Jerry Fodor makes a related point in 'A Theory of Content', Part 2, mimeo. See also Simon Blackburn 'The Individual Strikes Back', pp. 289–91.

direct. But if the rule is read under the assumption that conditions are favourable, then equally there is no difficulty, even with a simple rule, about how it comes to be fallibly readable. The individual will have to recognize in any instance of reading the rule that for all she knows she may be forced *ex post* to judge that she got it wrong.[25]

The upshot is a cheering one. It begins to seem that the sceptical challenge can be met after all. I can intentionally conform my behaviour to a rule exemplified for me by certain examples, given that those examples generate a certain inclination in me. I can identify such a rule independently of any particular application; I can read off what it requires directly; and yet in any instance of applying the rule I have to admit that I may be mistaken. The phenomenology of rule-following, as it is described in the first section, can be saved.

In conclusion, a methodological comment. Kripke is sometimes accused of putting a tendentious challenge: the challenge to identify rule-following reductively with this or that independent and familiar sort of psychological fact.[26] This challenge would be tendentious, so far as it assumes that rule-following is not a *sui generis* psychological fact. In responding however to the challenge posed in section 2, I have assumed that it takes a different form. I have taken the challenge to be that of explaining in familiar psychological terms how rule-following is possible, given the different and apparently conflicting constraints, objective and subjective, on rules. To explain rule-following in this sense need not be to identify it reductively with any independent psychological fact; it need not be to analyse rule-following in some other terms.[27]

A noteworthy feature of the account offered here is that while it seeks to explain how rule-following is possible, it does nothing to identify or analyse rule-following in reductive terms. Rule-following is possible, I argue, under two conditions. The first is that on being presented with certain examples the rule-follower develops an inclination to carry on in a particular fashion, an inclination in virtue of which the examples exemplify a particular rule for the agent. The second condition is that the agent is able to explain any intertemporal or interpersonal discrepancies in spontaneous application by appeal to perturbing factors, so that the rule exemplified—though she will not think of it this way—is the rule which dictates those responses that the corrected or standardized inclination supports, not the inclination neat. This explanation of how rule-following is possible—of how the objective and subjective constraints on rules can be simultaneously satisfied—nowhere says what rule-following is, reductively

[25] The account also makes room for a different sort of fallibility: not fallibility in applying a rule but fallibility in picking it up. Circumstances may miscue me so that I judge later that I went wrong about the rule which certain examples exemplified.

[26] See Goldfarb, 'Kripke on Wittgenstein on Rules'.

[27] See Huw Price, *Facts and the Function of Truth*, Blackwells, Oxford, 1989, for the distinction between explanation and analysis.

characterized. It tells a story about how rule-following might get going; it offers a genealogy of rule-following on a par with Hume's genealogy of causal talk or, more notoriously, Nietzsche's genealogy of morals. But it does not analyse in reductive terms what it means to say that this or that is a rule, that this or that is what it means for a rule to require something, and so on. That the agent follows such and such a rule will be supervenient in a suitable way on the facts about her inclination and context but it will not be identifiable with any such fact.

This abstention from analysis has one important result that we should mention in particular. The proposal which Kripke spends most time in demolishing, the proposal that rule-following reduces to indulging a disposition to go on in a certain way, is open to the following criticism: that the disposition mentioned in this analysis must be subject to the qualification of operating in the right way and that there is no reductive way of expressing this; to operate in the right way is just to operate in accord with the rule.[28] Our proposal, by contrast, is not vulnerable to this style of criticism. Since we do not try to analyse rule-following in reductive terms, we face no such problems. We attempt to give an explanation of how a rule-follower may see herself as having made a mistake and an explanation therefore of how we may see her inclination as having misfired. But this does not involve an assumption that there is a reductive account available of what it is for the inclination to fire correctly or incorrectly.

4. *Some corollaries*

Some philosophers will not be enthusiastic about the picture we have developed. While it saves the phenomenology of rule-following, the picture has corollaries which they will find repugnant. I shall mention three.

A first corollary we may describe as the precariousness of rule-following. Suppose that for some relevant decision-type, the standardized inclination goes awry; now it dictates this response, now that, without any evidence of perturbing influences. In that case I will have to conclude that the decision-type is not relevant or that there never was a unique rule on which I was targetted. The latter possibility is the threatening one and it remains ever present, so far as I cannot at any time be sure that there will not be a future breakdown of the kind envisaged. In order to aspire to follow a rule I must assume that the standardized inclination picks out a unique rule for me to follow. But I can never redeem that assumption fully. The enterprise of rule-following, and all that goes with it, then, is precarious. It rests on the contingency that certain responses can be corrected so as reliably to yield convergence.[29]

[28] Kripke, *WRPL*, p. 28.

[29] On this topic see John McDowell, 'Wittgenstein on Rule-Following', *Synthèse*, 1984, pp. 326–63.

A second corollary of our story is that not only is rule-following precarious, it is also in a certain sense interactive. It requires that the rule-following subject be in a position to interact with other bearers of the inclination—or a counterpart—at work in her: her self at later times or other persons. Without such interaction there cannot be a relationship between the inclination and the rule other than one of exact fit: specifically, there cannot be a suitable relationship of fit under favourable conditions. The subject would not be in a position to identify favourable conditions, even *ex post*. This means that the isolated *doppelgänger* of a rule-follower at any time *t*, the *doppelgänger* without history or company, cannot itself follow a rule. It may avail itself of certain inclinations to refer to *this* or *that* rule, as exemplified by certain examples, but it will not be fallible with respect to any rule identified and so it will not follow such a rule. Rule-following, like keeping your balance, is essentially an interactive enterprise. It makes requirements on the context of the rule-follower as well as on what happens in her head.

A third corollary, besides the precariousness of rule-following and its interactive character, is the relativity of rules. The story we have told means that it is a priori that if under favourable conditions there is appropriate convergence on respone *r* in situation *s*, then the rule in question requires that *r* in *s*. This is not to say that the person or even the total community can ever be certain—infallibly certain—that *r* is the correct response, for they can never rule out the possibility that later divergence will reveal that the conditions did in fact involve perturbations; they can never be sure that existing conditions are indeed favourable. Still, even if the a priori connection does not raise the spectre of infallibility, it does introduce a relativity to our species, perhaps even our culture, which many philosophers will find repugnant. It means that while I may struggle fallibly to be faithful to an objective rule in the enterprise of rule-following, which rule I am tracking is determined in a certain sense by my nature. Someone who lacked that nature, someone who lacked a suitable counterpart to the inclination operative in me, would have no capacity to tell what rule I was following or even that I was following a rule.

Of the three corollaries this last one will probably be found the most troubling. Consider how it bears on properties. Properties are rules-in-intension, so far as they normatively constrain predications over an indefinite variety of cases; they are thought of as determinable independently of any particular predication; and they are regarded generally as directly and fallibly accessible. The third corollary means that properties are in a certain sense relative to our kind. Each property may be independent of us, in the sense of being something in the world to which we each have only fallible access. But the extension of any property we engage with is determined in such a way that only someone who shares our inclinations can identify it.

This means, it will be said, that on our approach all the properties with which we engage fit a condition which many think of as a mark of secondary properties only. I agree but insist that a number of qualifications should be borne in mind. First, the secondary properties in any area are all of equal stature—for example properties of colour—whereas on our account some properties may well be identifiable only via other properties. Secondly, secondary properties have the characteristic that they are primarily associated with one sense only, whereas the inclination that goes by our account with any sort of property may operate on the basis of information from a number of senses. Thirdly, and perhaps most significantly, if I agree that secondary properties typify properties generally, that is only as far as I endorse a distinctively objectivist understanding of such properties. On that understanding the secondary property is realized in things perceived and is subjective only in the sense that which property is discerned in any perception is fixed relative to our kind: it is that property which is picked out in the actual world by such and such a sensation—and the associated inclination to go on—provided that conditions are favourable.[30]

Since the last corollary will still be found troubling, here is one further remark which may help to reconcile people to it. Where a property *P* is associated with human responses, such as judgements that it applies here or there, the following question picks up an important issue of objectivity: is something *P* because it is judged to be so or is it judged to be so because it is *P*.[31] The interesting feature of the account of properties inherent in our story about rules is that on that account this question naturally attracts the objectivist answer. Something is judged to be *P* because it is *P*; something commands a convergence in the *P*-response because of how it is, not because of collusion or whatever. Its being *P* is not exhausted then by its being subject to suitable judgements. Its being *P* ensures that under favourable conditions it will elicit suitable judgements.[32]

[30] We assume that favourable conditions cannot be identified in advance here any more than elsewhere. If they could be so identified, the secondary properties would cease altogether to be typical. See Crispin Wright, 'Moral Values, Projection and Secondary Qualities', *Proceedings of the Aristotelian Society*, 1988, pp. 1–22, for the sort of view I assume false.

[31] This is like the *Euthyphro* question as to whether something is right because the gods will it or whether the gods will it because it is right. It is akin to what Wright calls the order of determination test in 'Moral Values, Projection and Secondary Qualities'. For a similar test applied to truth and consensus see Philip Pettit, 'Habermas on Truth and Justice', in G. H. R. Parkinson, *Marx and Marxisms*, Cambridge University Press, 1982.

[32] Thus its being *P* will explain why it is judged to be *P*. The sort of explanation relevant is the program style of explanation distinguished in Frank Jackson and Philip Pettit, 'Functionalism and Broad Content' *Mind*, 1988, pp. 381–400. It is important with the Euthypro question to distinguish the causal—strictly, the causally programmatic—sense of 'because' from the evidential. An eraser bends because (causal) it is elastic, yet it is elastic because (evidential) it bends. Consistently with thinking that something is judged to be *P* because (causal) it is *P*, a theorist may think it is *P* because (evidential) it is judged to be *P*—by suitable subjects in suitable cirumstances. Indeed the theorist may even think that the evidential claim has a certain a priori support. In maintaining the causal claim as well as the evidential, the theorist will be distinguishing the property of being *P* from pseudo-

5. *Public rules*

There is a condition that is commonly imposed on the notion of a rule, other than the five distinguished in section 1. This is that the rule should be public in roughly the Wittgensteinian sense. 'Grasp of a rule must be manifest in what is interpersonally accessible—i.e. to others as well as to oneself—so that there can be no such thing as intrinsically *unknowable* (by another) rule-following.'[33] The question which we raise in this final section is whether this extra condition would force us to tell a more specific story about rule-following than that which is offered in the last section. I argue that it does, in particular that it requires rule-following to be interpersonally interactive. This means that any rule that it is possible for another to know someone is following is a rule identified by reference, not just to that person's own responses, but also to the responses of certain actual other people.

Suppose I believe that another person has identified a particular rule. Under the story of the last section, that means that I must take him to be representing the rule by certain examples. He will be doing this on the basis of an inclination that is intertemporally or interpersonally standardized: it is that rule which fits the inclination under favourable conditions, favourable conditions being judged on the basis of the assumption that intertemporal or interpersonal differences are explicable by perturbations. But suppose that the person identifies the rule on the basis of an inclination that is only intertemporally standardized; he has no expectation that others will display convergent responses. In such a circumstance it turns out that while I may *believe* that the person has identified such and such a rule—a rule I represent to myself via my intertemporally standardized inclination—I am not in a position to *know* that he has done so.

I am in no position to know what rule he has identified, because I do not meet a weak condition on knowledge. I cannot reliably tell that he is following one rule rather than any other. I have no reliable means of telling that the rule he is representing by such and such examples is the rule which requires this rather than that response on an example hitherto unencountered. Were our responses to come apart, he might remain quite content with his own response to the example. Using myself as a prosthetic device I may guess that it is this rule rather than that which he is following. But that is all I can do: guess. For all I know in any strict sense,

properties like that of being 'U' rather than 'non-U' (where saying 'lavatory' is 'U', saying 'toilet' is 'non-U' and so on). For a different perspective—and for the useful idea of a response-dependent concept—see Mark Johnston, 'Dispositional Theories of Value', *Proceedings of the Aristotelian Society*, supp. vol. 63, 1989, especially the last section. I read Johnston's paper while my own was going to press.

[33] Colin McGinn, *Wittgenstein on Meaning*, p. 192. See too Crispin Wright, 'A Cogent Argument Against Private Language?', in Philip Pettit and John McDowell (eds), *Subject, Thought and Context*, Oxford University Press, 1986, pp. 209–10.

his inclination may differ in a manner which means that he has a quite divergent rule in mind.[34]

This negative result means that it is only if the person identifies the rule on the basis of an interpersonally as well as intertemporally standardized inclination that I can know which rule he is following. But of course it remains to establish the corresponding positive result, showing that the fulfilment of this extra condition is probably sufficient as well as necessary to make such knowledge available to me. Suppose that I regard the person, and regard him rightly, as following a rule such that he expects convergence between us; he represents the rule on the basis of an intertemporally and interpersonally standardized inclination. Suppose that I get in step with him, developing the appropriately generated counterpart inclination: both of us survey some examples and we each develop the inclination to go on with which the rule is associated. The question is whether I am then in a position to know what rule he is following.

I am, for the following reasons. If there is a rule he is intentionally following, it is a rule exemplified in certain examples on the basis of an inclination we share. If there is a rule exemplified in certain examples on the basis of an inclination we share, then I am in a position to know what it is; I may actually get it wrong but I have at hand materials for reliably identifying the rule. Therefore if there is a rule he is intentionally following, I am in a position to know what it is. The condition under which I do not know what rule he is following—where our responses come irremediably apart—is a condition under which the rule he aspires to follow—the rule represented by the intertemporally and interpersonally standardized inclination—is an illusion; there is no such rule there to be identified.

There remains an assumption which has to be redeemed. This is the assumption that the other person follows a rule such that he expects convergence between us: a rule represented by an interpersonally standardized inclination. How can I—or we as analysts—have reason to think this and so to claim that I can know what rule he follows? The only way is *ambulando*, by finding in practice that the assumption works out, fitting with a disposition in the person to seek out an explanation of any difference between us. In that practice, as Wittgenstein would say, we hit bedrock. Here is where our spade turns.[35]

The upshot is that if rule-following is to be public then the rule-followers must interact with one another as well as with their earlier and later selves. Here we see a sort of vindication for the allegedly Wittgen-

[34] If Donald Davidson is right then on pain of dismissing the hypothesis that he is a rule-follower, I will have to interpret him as following rules familiar to me at some level. See his 'On the Very Idea of a Conceptual Scheme', reprinted in his *Truth and Interpretation*, Oxford University Press, 1984. But interpretation in this sense may still be just guesswork.

[35] See Edward Craig's discussion of the assumption of uniformity in 'Meaning, Use and Privacy', *Mind*, 1982, pp. 341–64.

steinian view that rule-following is possible only in a communal context. Rule-following as such requires interaction, or so the story of the last section has it. But that interaction can be provided in principle by oneself at other times as well as by other persons. Interaction with other persons only gets to be required if the rule is to be public: if it is to be a rule which another person can know you follow.[36]

Research School of Social Sciences
The Australian National University
Canberra, ACT, 2601
Australia

PHILIP PETTIT

[36] I am greatly indebted to discussions of this topic, some over years, with Simon Blackburn, Paul Boghossian, Frank Jackson, Peter Menzies, Karen Neander, Huw Price, Jack Smart, Neil Tennant, and Michael Tooley. I was also helped by comments received when versions of the paper were presented at the University of Sydney and at the Australian National University. The line in the paper may have been particularly influenced by Huw Price. Certainly it fits well with some strands of argument in his book *Facts and the Function of Truth*.