

PAUL A. BOGHOSSIAN

## RULES, MEANING AND INTENTION – DISCUSSION

Wittgenstein doesn't ever outright say that the key to understanding many of the fundamental questions of philosophy lies in the phenomenon of rule-following – but then, again, when does he ever say anything outright? He does, however, tantalizingly, suggest it. And so, ever since the publication of *Philosophical Investigations*, various philosophers have tried to execute Wittgenstein's project (as I shall call it), attempting to show how a correct explanation of the possibility of rule-following conditions our understanding of other fundamental issues.

In recent times, the most notable attempt to carry out this project was made by Saul Kripke in his now justly famous *Wittgenstein on Rules and Private Language*. In Kripke's hands, the rough architecture of the project looks something like this. There is a skeptical problem about following rules: it is hard to see how there could actually be any such thing. The correct resolution of this intolerable skeptical problem shows us something deep about meaning and the conditions under which it is possible: first, that meaning is a matter of assertibility conditions and not of truth conditions; and, second, that meaning is only possible in a communal setting and not in isolation.

In *Rules, Reasons and Norms*, Philip Pettit develops his own interesting, ambitious and wide-ranging version of Wittgenstein's project. Although his version has much the same architecture as Kripke's, it differs from the latter in several significant respects. First, his rendition of the skeptical problem about rule-following is very different from Kripke's (although Pettit is fond of understating the difference, calling it more a change of “emphasis.”) Second, his preferred solution to that

problem is very different from where Kripke ends up, although similar-sounding lessons about anti-realism and communalism are drawn.

#### WHAT IS RULE-FOLLOWING

What exactly is the phenomenon of following a rule and what is supposed to be the problem with being able to pull it off?

Crispin Wright, who has written many illuminating pieces on the topic, introduces it thus:

Philosophical issues to do with rule-following arise for every normatively constrained area of human thought and activity, wherever there is better and worse opinion, correct and incorrect practice.<sup>1</sup>

This seems to suggest that wherever there is normative constraint there there must be rule-following. But if that is the suggestion, it seems wrong. Intuitively, and without the benefit of controversial assumptions, the phenomenon of rule-following does not arise *wherever* there is normative constraint, but only where there is a certain *kind* of normative constraint. Let me explain.

We are familiar with the idea that, having thought something, or having done something, our thought or behavior is evaluable according to a set of norms that are objectively applicable, applicable to our thought or behavior whether we like it or not. Having thought

It is snowing outside

I no longer have any choice about whether my thought is subject to assessment as true or false, or as justified or unjustified. These norms apply to my thought just by virtue of its being a thought with factual content.

Similarly, having tossed the UNICEF envelope in the trash without opening it, we have no choice about whether what we have done is subject to moral assessment, as either right, wrong or morally neutral.

What seems distinctive about rule-following, at least at first blush, is not that it gives rise to normative assessment as such, but that it gives rise to normative assessment relative to an *accepted* commitment – an accepted commitment to conform one’s behavior to a certain pattern.<sup>2</sup> From an objective standpoint, a rule that one may be following – gambler’s fallacy, for example, or affirming the consequent – may be quite a bad rule, leading to thoughts that are false, or to behavior that is irrational. But those very thoughts or behaviors may yet be entirely *correct relative to the rule that one is following*; and that is the sort of normative constraint that we need to understand if we are to understand the phenomenon of following a rule. The possibility of assessment relative to an *accepted* commitment is the phenomenon we are talking about; not the possibility of assessment as such.

#### PETTIT’S CONCEPTION OF RULE-FOLLOWING

How should we think of this “acceptance of commitments?” Pettit is helpfully explicit about this:

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.<sup>3</sup> (p. 27)

Pettit is thus building it in to what he *means* by rule-following that one only follows a rule if one is acting in accord with one’s *intentions* to conform to a rule. In terms of the terminology of “accepted commitments,” he is taking it that the only way to accept a commitment to conform one’s behavior to a certain pattern is through the formation of an intention. Call this the Intention Assumption.

For Kripke, by contrast, the idea that rule-following is always in this way intention-based is one of the candidate *theories* about rule-following, not something that provides the very definition

of the notion of following a rule. There is an important reason for this that it will be instructive to bring out.

One of the central ideas of Kripke's discussion of rules is that meaning itself is a matter of rule-following: both in the case of our public language and in the case of our "language of thought," our words come to have meaning by virtue of the fact that we use those words according to certain rules. Call this Kripke's Meaning Assumption. This Assumption is a linchpin of Kripke's discussion, crucial both to explaining the importance that Kripke attaches to the subject of rule-following and also to explaining how his skeptical argument proceeds.

Now, it should be obvious that combining the Meaning Assumption with the Intention Assumption will lead rather quickly to the conclusion that rule-following, and with it mental content, are metaphysically impossible. Given the two assumptions, in order to follow rules, we would have to have thoughts (intentions) with content. For us to have thoughts with content, the expressions of our language of thought would have to have meaning. For those expressions to have meaning, we would have to use them according to rules. For us to use them according to rules, we would antecedently have to have thoughts with content. And so neither content nor rule-following would be able to get off the ground.

So, combining the two assumptions will lead rather quickly to the conclusion that meaning and rule-following are impossible. Although skepticism about those phenomena is supposed to be the target of Kripke's argument, it surely wasn't supposed to be that easy!

Clear-headedly, then, Pettit rejects the Meaning Assumption, at least with respect to the meanings of *mental* words. (He endorses it, plausibly enough, for the meanings of public language words.) On Pettit's view, mental content does not depend on rule-following and so can simply be conceded for the purposes of the present discussion:

I formulate the problem of rule-following, and I propose a solution, on the assumption that a creature that does not follow rules in my sense, and is not therefore a speaker or thinker...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. (p. 27)

For the reader who may be puzzled how someone could have beliefs and desires and yet not be a “thinker,” the answer is that Pettit is using “thinker” in a special demanding sense which requires not merely that the person have propositional attitudes but that he modify those attitudes with certain aims. More on this later.

#### PETTIT’S SKEPTICAL PROBLEM AND NON-SKEPTICAL SOLUTION

Given Pettit’s admirably clear understanding of what he means by “rule-following,” what skeptical problem for this phenomenon does he identify? There is a *prima facie* problem here. If we are granted full use of intentional notions, and we stipulate that rule-following consists in intentionally conforming to rules, what could possibly be the problem about rule-following?

In outlining his version of the skeptical problem, Pettit says that he can afford to be brief “since the challenge has been well elaborated by Kripke.” (p. 32) But as we have seen, Pettit’s problem can’t be Kripke’s problem, since (a) Kripke does not make the Intention Assumption and (b) a crucial part of Kripke’s skeptical argument relies on the Meaning Assumption that Pettit rejects.

Since, on Pettit’s view, we can help ourselves to intentional notions, we can presumably help ourselves to the intention to conform to a certain rule – for example, to the intention to conform to the rule: *For any two numbers  $x$  and  $y$ , whenever you are asked ‘ $x + y = ?$ ’ always respond with their sum!* And, since, given the Intention Assumption, following the rule for addition just means conforming one’s behavior with ‘+’ to such an intention, it’s hard to see what could stand in the way of our following such rules.

If I understand him correctly, Pettit seems to think that, these points notwithstanding, there is a problem about rule-following because there is a *special* problem seeing how we

could have intentions to conform to *rules*. It is one thing to concede the existence of intentions with ordinary contents – for example, the intention to go to the store – he must be thinking; it is quite another thing to concede the existence of intentions to conform to *rules*. (see p. 32)

Why should there be a special problem about the intention to conform to a rule? As Pettit explains it, such intentions are subject to two sorts of conditions, subjective and objective; and there is a *prima facie* problem seeing how anything could satisfy both sorts of conditions simultaneously.

On the objective side, if I have the intention to conform to the rule for addition, I undertake to conform to a set of normative constraints that apply over an indefinite number of decision types: for any pair of numbers, however large, the rule dictates a particular answer as the correct one, quite independently of what I am disposed to say. On the subjective side, if the rule is to be followable, then its requirements must be directly – albeit fallibly – readable by the thinker whose intention it is to follow the rule.

According to Pettit, the problem is that it is difficult to see how any one thing could simultaneously satisfy both of those sets of conditions. One way to satisfy the objective condition is to think of the rule for addition as involving the infinite set of triples that specifies the addition function. But how would such a set be directly readable by a thinker? Some of its entries would be too large to be surveyable by a finite mind.

On the other hand, says Pettit, we could try to conceive the rule for addition as the “rule-in-intension” – not as involving the addition function itself, but the *sense* of such a function, much as I described it above: *For any two numbers  $x$  and  $y$ , whenever you are asked ‘ $x + y = ?$ ’ always respond with their sum!* However, Pettit claims, “here the problem is to explain how we are able to get in contact with such an abstract object.” (p. 28)

But I don’t really understand what problem Pettit is gesturing at here, given the concession of intentional content. In the relevant sense, getting in contact with a rule is just thinking it, fixing it as the object of one of one’s attitudes, for example, as the object of one’s intentions. But conceding

that sort of “contact” with an abstract object is precisely what it is to concede the problem of intentionality. Why is it any harder to “get in contact” with the sorts of abstract object that rules are than it is to get in contact with any propositional content? A rule, after all, is just an imperatival content of the form:

*If C, do A!*<sup>4</sup>

If we are not questioning our ability to think thoughts of the form

*If C, A,*

on what grounds do we worry about our ability to think thoughts of the form

*If C, do A!?*

Consider, for example, the rule: “Answer any email immediately upon receipt!” That’s a rule just as much as that for addition. It can be expressed in the imperatival form: “If x is an email and you have just received it, answer it immediately!” If we are able to assume the notion of an intention, then we can assume that we can form the intention to answer any email immediately upon receipt. But then nothing further seems required to explain how I can follow that particular rule. I count as following the rule if I conform to it and do so precisely because it is my intention to do so.

It’s hard to see, therefore, exactly what problem about rule-following Pettit has identified, given his understanding of what the phenomenon intuitively consists in.

I also have difficulties making sense of Pettit’s solution to the problem how we can intend to conform to rules. This is how Pettit describes his solution:

We have taken three steps in developing our response to Kripke’s challenge. We have argued, first, 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 that 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 that the inclination operates under favourable conditions. (p. 40)

If we apply what Pettit is offering to the case of addition, we get the following: What determines what rule someone is following in his use of the ‘+’ sign? The rule he is following is the rule to which his inclinations to use the sign conform under favorable conditions.

This very much looks like the sort of dispositional account that Kripke criticizes at length and against which he develops two important objections. First, our dispositions are finite: with respect to some very large numbers, it may well not be true to say that we are disposed to answer with one number as opposed to another, when asked about their sum. However, the rule for addition determines a unique answer for an indefinite number of inputs. Second, it is hard to specify the “favorable” conditions in substantive terms, so that they avoid assuming a vacuous “whatever it takes” character.

Pettit does not address the first of these criticisms. But he devotes an entire essay to developing a theory of favorable conditions, one that is designed to avoid the charge of vacuity. He writes:

What factors, then, do people’s practices make it right for them to treat as unfavourable; and, relatedly, what conditions do they make it right for them to treat as unfavourable? Suppose, as their practices commit people to supposing, that there is a property or other entity there for a term like ‘red’ or ‘regular’ to designate, and that other people are reliable detectors of that property in the absence and only in the absence of certain perturbing or limiting factors. The answer, then, is straightforward. Unfavorable factors will be those factors such that, if people identify them as perturbances and limitations that undermine detection, then that would maximize expected, long-term convergence among individuals in the use of ‘red’ or ‘regular’: specifically, in the use of ‘red’ or ‘regular’ to ascribe the property it currently ascribes. Or at least it would maximize such convergence among individuals who are not separated by relevant differences of background belief. (pp. 147–148)

If I understand Pettit’s view here – and I am not fully confident that I do – it is difficult to see how it escapes the charge of vacuity or problematic circularity. The trouble is that it

identifies the favorable conditions as those that it would be *right* for the users of a word to treat as favorable, not as those that that, as a matter of fact, they *do* treat as favorable. The latter answer would be no good, of course, because thinkers cannot be guaranteed to know which conditions are most conducive to the correct application of their words.

But Pettit's preferred answer is problematic, too, because the question which conditions it would be *right* for the users of a word to treat as favorable cannot be settled independently of what they mean by that word, what rule they intend to be conforming to by their use of that word. One and the same set of conditions will be optimal for detecting the property denoted by one word and limiting for the detection of another: for example, the shape of a large object is best detected by someone at a certain distance from it, whereas its texture is best ascertained from up close.

But if the equation of  
the rule being followed with respect to *w*

with

the rule to which one is disposed to conform in one's  
use of *w* under favorable

is to be substantive, then the account of "favorable conditions" cannot assume prior appeal to facts about the rule with which *w* is being used. Otherwise, to know whether a particular set of conditions is favorable we will need to know what *w* means, and to know what *w* means we will need to know whether we are in favorable conditions, neither notion giving us any purchase on the other.

#### MEANING, INTENTION AND RULES

Let me return to the question about how the issue about rule-following is best set up. If the phenomenon of rule-following is not to be metaphysically impossible, we cannot combine the Meaning and Intention Assumptions. So the question arises which, if either, we should keep.

My own view, articulated in my 1989 paper, used to be the same as Pettit's.<sup>5</sup> I used to think it was constitutive of rule-following that it be intentional action, and I used to think that it was implausible that we come to mean things by our mental words because we follow rules in respect of them.

I am now inclined to believe that this gets matters exactly wrong, and that it is Kripke's picture, which combines Meaning with the rejection of Intention, that is correct (although my reasons for thinking this differ from Kripke's).

One relatively minor reason for rejecting Intention is that we are prone to think that we follow many rules that we cannot state even approximately correctly, even after considerable reflection. Linguists have been trying to figure out the rules of syntax for years; but to this day no one has succeeded in stating those rules in a way that is free of counterexamples. But as Pettit says, if we are to take seriously the idea that rules are the contents of intentions, then those rules had better be rather directly readable: it had better be relatively easy to say what the contents of those intentions are.

This consideration already speaks quite strongly against Intention. However, my main reason for rejecting Intention derives from another source: from the appeal of the idea that being justified in having the beliefs that we have consists in our following the right epistemic rules. Call this the Epistemic Assumption.

One example of the sort of epistemic rule I have in mind might be given by

Observation: For any observable proposition  $p$ , if it visually seems to you that  $p$  and circumstantial conditions  $D$  obtain, then believe  $p$ .

It is not easy to be precise about which propositional contents are observable in this sense, but our commitment to the existence of some such distinction is clear enough (propositions about the shapes of middle-sized objects count, whereas those about subatomic particles don't). Another example is the rule we may call

Modus Ponens: *If both  $\langle p \rangle$  and  $\langle \text{if } p, \text{ then } \rangle$  then, other things being equal, believe  $\langle q \rangle$ .*<sup>6</sup>

Pettit evinces considerable sympathy for the Epistemic Assumption when he says that “thinking in the sense involved here [meaning thinking that is aimed at truth or justification] seems to require, like speech, the capacity to follow rules.” (p. 31)

But the Epistemic Assumption isn’t any more consistent with Intention than Meaning was: just as we found the combination of Meaning and Intention to pose an insuperable problem for rule-following, so, I will argue, the combination of the Intention and Epistemic Assumptions leads to a similar impossibility result.

To see why, let’s go back to the example of answering every email upon receipt. Let us construe the adoption of that rule, very plausibly, I think, as involving an explicit intention to conform to the instruction: *If  $x$  is an email and you have just received  $x$ , answer it immediately!*

How should we imagine my conforming to this explicit intention? Well, I form the intention:

Intention: If  $x$  is an email and you have just received  $x$ , answer it immediately!

Now, to act on this intention, I am going to have to discharge the antecedent: I am going to have to think, even if very fleetingly and not terribly consciously:

Premise: This is an email that I have just received.

in order to draw the

Conclusion: I will answer it immediately.

Notice, though, that in acting on this intention to conform to a given rule, I have had to *use* another rule, namely Modus Ponens. The explicitly formulated intention cannot help me in deciding what to do unless I use Modus Ponens to reason my way to a conclusion about what the rule demands under the circumstances.

If the intention-based model is to capture all instances of rule-following, then it had better also capture my use of the

Modus Ponens rule. But, as any reader of Lewis Carroll's famous note "What the Tortoise said to Achilles" will realize, it is impossible to see how that might be done. If we tried, we would get something like the following:

Intention: If C, do A

Premise: C

Modus Ponens: Whenever both 'If, C, do A' and C' do A

Conclusion: Do A

And now we would be off on a vicious regress. Epistemic rules whose operation cannot be captured by the intention-based model are presupposed by that model itself.

It would seem, therefore, that even without the Meaning Assumption, and given only the Epistemic Assumption, it is easy to show that Intention must go.

However, once we get rid of Intention, it is very tempting to suppose that, at least in the important case of the mental words that express the logical constants, their meanings are grasped by virtue of our using them according to certain epistemic rules and not others. And so we get Meaning – or at least a special case of it.

Hence, I am now inclined to think that the correct issue about rule-following is: How could we be said to operate according to certain rules – especially according to certain epistemic rules – given that we could not be said to have accepted those rules by forming the *intention* to conform to them.

#### NOTES

<sup>1</sup> "Wittgenstein on Following a Rule: Five Themes," unpublished ms, p. 1. There is a slightly weaker formulation in Wright's *Rails to Infinity* (Cambridge: Harvard University Press, 2003), p. 1:

The principal philosophical issues to do with rule-following impinge on every normatively constrained area of human thought and activity: on every institution where there is right and wrong opinion, correct and incorrect practice.

<sup>2</sup> In the sense in which I mean it here, acceptance need not be a conscious mental process: in learning the rules of syntax of his language, a child comes to accept those rules.

<sup>3</sup> All references to Pettit are to *Rules, Reasons and Norms* and will be incorporated into the text.

<sup>4</sup> Not everything that we call rules in ordinary discourse has exactly this form. For example, we often talk of there being a rule for castling in chess. But this rule only *permits* castling under certain conditions; it never *requires* it. These subtleties won't matter for the point in the text.

<sup>5</sup> 'The Rule-Following Considerations', *Mind*, 1989.

<sup>6</sup> There are many delicate issues about exactly how these epistemic rules should be formulated, many of them raised fruitfully by Gilbert Harman. However, they won't matter for present purposes so I shall ignore them.

*Department of Philosophy*  
*New York University*  
*New York NY 10003*  
*USA*  
*E-mail: pb@nyu.edu*