Blind Rule-Following

1. Introduction

It is a great pleasure to be able to contribute to this Festschrift in honor of Crispin Wright, with whom I have enjoyed countless stimulating conversations about a host of fundamental philosophical issues over the past twenty years. It is especially appropriate that my contribution to this Festschrift concern the topic of rule-following for this topic has been central to both of our preoccupations and has dominated our discussions.

As anyone with the slightest familiarity with this subject will know, Wright has written several important and illuminating papers on the phenomenon of rule-following, papers that draw their inspiration from Wittgenstein's seminal discussion of that notion. In those papers, Wright lays out an original view both about what the rule-following problem is, or should be taken to be, and about its correct resolution. In both respects, his view differs from that of Saul Kripke's influential account.[1]

Kripke's well-known view is that there is an enormous skeptical problem seeing how rule-following is so much as possible.[2] Wright maintains that this problem is misguided, that Kripke's skeptical challenge can receive a relatively straightforward solution. According to Wright, we can follow rules by forming intentions to uphold certain patterns in our thought or behavior and by acting on those intentions. Call this the *Intention View* of rule-following.

The real problem posed by Wittgenstein's discussion, Wright continues, is not to explain the very possibility of rule-following but, rather, to explain how we can have the sort of privileged access to the contents of our intentions—and of our intentional states more generally—that we normally credit ourselves with.

I shall argue for three claims. First, Wright is correct to think that it is possible to respond to the specific challenge that Kripke poses with the Intention View. For (p.28) reasons that will emerge, this is not quite as straightforward as Wright seems to suppose, but it is ultimately defensible.

Second, there is little doubt that Wright's problem about self-knowledge of intentional states is a real problem, although I believe that his own proposed solution to that problem is unlikely to succeed.

Finally, though, I shall argue that while the Intention View may constitute an adequate response to the specific considerations adduced by Kripke, it cannot in the end be regarded a correct account of rule-following. In its most fundamental incarnation, I shall argue, rule-following *cannot* consist in some *intentional* fact.

In and of itself this result might be considered significant but not necessarily dispiriting. Matters begin to look significantly worse, however, when we combine this result with the powerful arguments to be found, in Kripke and elsewhere, that threaten to show that rule-following *cannot* consist in some *non-intentional* fact either.

And the full depth of our problem emerges when we realize that skepticism about following a rule is not ultimately a coherent option.

If all of this is right, then we face what might be called, following Kant, an antinomy of pure reason: we both must—and cannot—make sense of someone's following a rule. That is what I propose to argue for in this essay.

2. What is a Rule? What is Rule-Following?

We should begin with a very basic question: What are we talking about when we talk about someone following a rule? I don't mean: What is the right *theory* of rule-following? I mean to be asking in an intuitive and pre-theoretic manner: what phenomenon is at issue? It is surprising how often writers will launch into a discussion of this topic without pausing long enough to give it an intuitive characterization, especially since the answer turns out to be anything but obvious.

The question—What is it to follow a rule?—naturally breaks up into two. What is a rule? And what is to follow one?

The first question is surprisingly tricky, though limitations of space prevent me from discussing it in the requisite detail.[3]

Clearly, when we talk about people following rules we mean that they are somehow or other observing (or attempting to observe) certain *general* principles or standards. But there are at least two importantly different ways of conceiving such principles.

On the one hand, we can think of rules as "directions" or "instructions"—i.e. contents that are expressed by *imperatives* of the form: *If C, do A!* On this conception, which appears to be Kripke's, rules are contents that *prescribe* certain patterns of behavior under certain conditions. This is certainly a very common way of understanding the notion of a rule.

(p.29) However, not everything that we call a rule in ordinary language, or in the course of philosophical theorizing, conforms to this characterization. For example, we talk about the "rules of chess." One of these rules is:

> (Castle) If the configuration is C, you *may* castle.

This is not an imperative but, rather, what I would call a *normative proposition*. It is a norm of *permission.* It cannot be expressed by the imperative

> *If configuration is C, castle!*

because that would suggest that whenever the configuration is C, you must castle, whereas the rule for castling merely *permits* castling and does not require it. Indeed, I would argue, but won't do so here, that there is no good way to express a norm of permission in imperatival terms.[4]

So we have a distinction between an imperatival content and a normative proposition and we need to decide whether, when we talk about following rules, we are talking about following the one sort of content or the other.

For a variety of reasons, I am inclined to think that the more fundamental notion is that of a normative proposition and not that of a prescription or direction. But I won't argue for it here.[5] For the purposes of the present essay, I will take a rule to be *either* a general normative proposition or a general imperatival content.

Let me turn instead to asking about the more pressing question: What is it to *follow* a rule?

In answering this question, we should distinguish between a *personal-level* notion of rule-following and a *subpersonal* notion. We should not assume, at the outset, that our talk of a *person's* following a rule comes to exactly the same thing as our talk of, say, his brain's following a rule, or of his calculator's computing a function.

I propose to start with attempting to understand the personal-level notion, returning to the subpersonal

notion later. My view will be that there is a core concept that is common to both notions, but that the personal-level notion is richer in a particular respect that I shall describe below. Once we have a handle on the personal-level notion it will be easy to indicate the weakening that gets us the subpersonal notion.

Apropos of the personal-level notion, we certainly know this much: to say that S is following rule R is not the same as saying that S's behavior *conforms* to R. Conforming to R is neither necessary nor sufficient for following it.

It is not necessary because S may be following R even while he fails to conform to it. This can happen in one of two ways. Say that R is the instruction 'If C, do A!' S may fail to recognize that he is in circumstance C, and so fail to do A; yet it may still be true that S is following R. Or, he may correctly recognize that he is in C, but, as a result of a performance error, fail to do A, even though he tries.

(p.30) Conformity to R is not sufficient for S's following R because, for any behavior that S displays, there will be a rule—indeed, infinitely many rules—to which his behavior will conform. Yet it would be absurd to say that S is following all the rules to which his behavior conforms.

There is another possible gloss on our notion that we need to warn against. There is a persistent tendency in the literature to suggest that the claim that S is following rule R means something roughly like: R may correctly be used to *evaluate* S's behavior.

I am not confident that this construal can be pinned on Wright; but it is suggested by the remarks with which he tends to introduce the topic of rule-following, for example:

> 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.[6]

Whether or not this can be attributed to Wright, it is worthwhile seeing what is wrong with it. Intuitively, and without the help of controversial assumptions, it looks as though there are many thoughts that S can have, and many activities that he can engage in, that are subject to assessment in terms of rule R even if there is no intuitive sense in which they involve S's *following* rule R.

Consider Nora playing roulette. She has a "hunch" that the next number will be '36' and she goes with it: she bets all her money on it. We need not suppose that, in going with her hunch, she was following any rule—perhaps this was just a one-time event. Still, it looks as though we can normatively criticize her belief as *irrational* since it was based on no good evidence.

Or consider Peter who has just tossed the UNICEF envelope in the trash without opening it. Once more, we need not suppose that Peter has a standing policy of tossing out charity envelopes without opening them and considering their merits. However, even if no rule was involved it can still be true that Peter's behavior was subject to normative assessment, that there are norms covering his behavior.

In both of these cases, then, a norm or rule applies to some thought or behavior even though there is no intuitive sense in which the agent in question was attempting to *observe* that norm or rule.

Of course, some philosophers—like Kripke's Wittgenstein—think that wherever there is *intentional content* there must be rule-following, since meaning itself is a matter of following rules. But that is not a suitably pre-theoretic fact about rule-following; and what we are after at the moment is just some intuitive characterization of the phenomenon. We will come back to the question whether meaning is a matter of following rules.

When we say that S is following a rule R in doing A, we mean neither that S conforms to R nor simply that R may be used to assess S's behavior, ruling it correct if he conforms and incorrect if he doesn't.

What, then, do we mean?

(p.31) Let us take a clear case of personal-level rule-following. Suppose I receive an email and that I answer it immediately. When would we say that this behavior was a case of following the:

> (Email Rule) Answer any email that calls for an answer immediately upon receipt!

as opposed to just being something that I did that happened to be in conformity with that rule?

I think it is clear that it would be correct to say that I was following the Email Rule in replying to the email, rather than just coincidentally conforming to it, when it is somehow or other *because* of the Email Rule that I reply immediately.

Equally clearly, the because here is not any old causal relation: if a malicious scientist (or an enterprising colleague) had programmed my brain to answer any email upon receipt (in some zombie-like way) because *he* accepted the rule that I should answer any email upon receipt, that would not count as *my* following the Email Rule. (It might count as my brain following the rule.) Rather, for me to be following the rule, the 'because' must be that of rational action explanation: I follow the Email Rule when that rule serves as *my reason* for replying immediately, when that rule *rationalizes* my behavior.

I want to suggest, then, that the minimal content of saying that person S follows rule R in doing A is that *R serves as S's reason for doing A*.

Now, R is just a content of some sort, either an imperative or a normative proposition, as previously discussed. How is it possible for a content to serve as S's reason for doing something? Obviously, by being *accepted* or *internalized* by him.

I shall typically refer to this as S's *acceptance* or *internalization* of the rule, though, clearly, it will be very important to understand this as neutrally as possible for now.[7]

However exactly it is understood, what is important is that, in any given case of rule-following, we have something with the following structure: a state that can play the role of rule-acceptance; and some non-deviant casual chain leading from that state to a piece of behavior that would allow us to say that the accepted rule explains and (in the personal-level case) rationalizes the behavior in question.

Occasionally, I will also describe the matter in terms of the language of commitment: In rule-following, I will say, there is, on the one hand, a *commitment*, on the part of the thinker, to uphold a certain pattern in his thought or behavior; and, on the other, some behavior that expresses that commitment, that is explained and rationalized by it.

I will leave it to the reader to discern whether I have construed these notions in a way that is illicit or question begging. For the moment, let me just note that this characterization coincides well with the way Kripke seems to be thinking about the phenomenon of rule-following. As he says apropos of following the rule for addition:

> (p.32) I learned—and internalized instructions for—a *rule*, which determines how addition is to be continued…This set of directions, I may suppose, I explicitly gave myself at some earlier time…It is this set of directions…that justifies and determines my present response.[8]

I think it was a mistake on Kripke's part to use the word "justify" in this passage, rather than the word "rationalize." In talking about rule-following, it is important to bear in mind that we might be following *bad* rules. The problem of rule-following arises no less for Modus Ponens than it does for Affirming the Consequent or Gambler's Fallacy. If I am following Gamblers' Fallacy, my betting big on black after a long string of reds at the roulette wheel wouldn't be *justified*; but it would be *rationalized* by the rule

that I am following. Given that I am committed to the fallacious rule, it makes sense that I would bet big on black at that point.

We may summarize our characterization of personal-level rule-following by the following four theses:

> (Acceptance) If S is following rule R ('If C, do A'), then S has somehow accepted R.
>
> (Correctness) If S is following rule R, then S acts correctly *relative to his acceptance* if it is the case that C and he does A.
>
> (Explanation) If S is following rule R by doing A, then S's acceptance of R *explains* S's doing A.
>
> (Rationalization) If S is following rule R by doing A, then S's acceptance of R rationalizes S's doing A.

With this characterization of the personal-level notion in place, it is possible, I think, to see the subpersonal notion of following a rule as involving the first three elements but not the fourth.

If I say of a calculator that it is adding, then I am saying that its 'internalization' of the rule for addition (via programming) explains why it gives the answers that it gives. But I am obviously not saying that the addition rule *rationalizes* the calculator's answers. The calculator doesn't act for reasons, much less general ones.[9]

3. Acceptance, Intention, and Wright's Problem

With these important preliminaries behind us, let us turn to asking why there is supposed to be a problem about following a rule. What, in particular, does Kripke find so mystifying about it?

Kripke's problem is focused on the personal-level notion and on the Acceptance condition for it. He is struck by two facts. First, by the fact that most of the rules we follow are rules with infinitary contents; and, second, by the fact that our rules are supposed to rationalize (he says "justify") the behavior that constitutes following them. (p.33) And he is mystified by how it might be possible for finite minds like ours to instantiate the sort of state that would have both features. What could rule acceptance be such that it could have both of these requisite features?

Kripke illustrates his problem by considering the case of the symbol '+'. In using this symbol, he supposes, I may be taken to be following the rule for addition. In what does this fact consist? There look to be two serious candidates: either in some fact about my dispositions to use the symbol, or in some sort of intentional fact about me. Kripke finds both candidates wanting.

Now, Wright agrees with Kripke's rejection of dispositionalism. He maintains, however, that the *intentional* suggestion emerges unscathed from Kripke's skeptical considerations. As he puts it:

> so far from finding any mystery in the matter, we habitually assign just these characteristics [the characteristics constitutive of the acceptance of a rule] to the ordinary notion of *intention*… intentions may be general, and so may possess, in the intuitively relevant sense, potentially infinite content.[10]

This is the position that I earlier called the Intention View.[11]

The Intention View, of course, is just a special version of a more general class of views according to which rule acceptance consists in some intentional state or other, even if it is not identified specifically

with an intention. Call this more general view the Intentional View of rule acceptance. Although I will follow Wright in focusing on the Intention View, most everything I say will apply to the less committal Intentional View.

Wright considers the Intention View to be a perfectly adequate response to Kripke's question, as far as it goes. He believes the real difficulty lies not in explaining how it is possible for there to be such infinitary commitments, but rather in explaining how it is possible for a subject to have the sort of authoritative self-knowledge about them that we seem to have. For Wright takes it to be characteristic of the intuitive notion of intention, that

> it is a state of mind, alongside mood, thought, desire, sensation, etc., for which, in at least a very large class of cases, subjects have special authority and whose epistemology is first/third person asymmetric.[12]

And he takes the difficult question to be how states with such epistemologies are possible.

(p.34) I think there is no doubt about two points. First, that there are many instances of rule-following that are well captured by the Intention View, the email example outlined above being one of them. One adopts the email rule by forming a general intention to answer any email upon receipt and by having this intention subsequently inform and control one's behavior.

The second point on which we can all agree is that Wright's worry is a genuine one, in as much as it is very difficult to explain how we are able to have authoritative first-personal knowledge of our intentional states. There are at least two problems here (externalist conceptions of content would add a third).

First, intentional states typically lack an individuative phenomenology. How, then, are we able to introspect them so reliably? And why are they first-person/third-person asymmetric?

Second, intentional attributions often behave as though what is getting attributed is a *capacity* or a *disposition* to do something, and not some state of mind which could be wholly present to the subject's consciousness at a given moment in time.[13] Thus, we allow that Doron can authoritatively avow at time t that he wants to play chess, even if he did not explicitly think at t of all the rules that chess consists in. But this avowal will be defeated if Doron's subsequent performance falls short of the standard implicit in the attribution—if for example, he goes on to show that he doesn't understand the rules of chess. How can we reconcile these two facets of the attribution of an intentional state?

By way of solving these problems, Wright experiments with treating avowals as fact-constituting rather than fact-detecting judgments. Given the difficulty of accounting for these judgments in classically Cartesian terms, this is clearly a line of research that is worth pursuing.

But it is hard to see how to pull it off: it's hard to see how it could in general be true that facts about mental content are constituted by our judgments, since, it would appear that, by the terms of the theory itself, facts about the contents of the putatively fact-constituting judgments would have to be constituted independently of such judgments.[14] Still, I agree with Wright that this consideration doesn't decisively refute such views and that there are many interesting questions about them that remain to be investigated.

4. Problems for the Intention View

But what concerns me in this essay is whether it is true that we can solve Kripke's problem as easily as all that? Can we really just appeal to intentions with infinitary contents to explain how the Acceptance condition on rule-following gets fulfilled? I can think of three reasons that may be found in the

literature why someone might resist Wright's Intention View.

(p.35) The first reason would be provided by the assumption of Naturalism. A Naturalist would insist that intentions be shown to be naturalistically reducible before they could legitimately be appealed to in solving Kripke's problem. However, it is none too clear how such a reduction is to be pulled off and Kripke's book provides a battery of arguments against its feasibility (more on this below).

Second, and even if we were to put the Naturalist Assumption to one side, there look to be straightforward counterexamples to the Intention View: not everything that we would intuitively count as rule-following is intentional action in the sense that it specifies.

To be sure, and as I have already emphasized, there are many cases of rule-following that are captured very well by the Intention View, the email rule discussed above being one of them. But there also seem to be several significant cases of rule-following in which it would be implausible to say that there was a general *intention* to conform to a pattern involved.

The trouble comes in part (and ironically) from a feature of intentions that Wright himself emphasizes —namely, that their contents are typically thought of as highly accessible to their owners. The problem is that we appear to follow many rules that we aren't able to specify at all precisely.

Take, for example, the rules by which we regulate our moral conduct. Are we Kantians or consequentialists? It is, of course, a highly controversial normative matter which rules are the morally correct ones. But even just as a matter of descriptive fact, it is controversial whether we employ deontological principles, consequentialist ones, or some confused mixture of the two. If these moral rules were the contents of intentions of ours, wouldn't we expect to know what they are with a much higher degree of precision and clarity? A similar point could be made using principles of etiquette or the epistemic rules by which we update our beliefs.

A third type of consideration against the Intention View is provided by an assumption that is crucial to Kripke's thinking about rule-following. Kripke sets up the rule-following problem by asking what determines whether I am using the '+' sign according to the rule of addition as opposed to the rule for quaddition, where quaddition is a function just like addition, except that it diverges from it for numbers larger than we are able to compute. He considers saying that what determines that rule-following fact is some general intention I formed to use the symbol according to the one rule rather than the other:

> What was the rule? Well, say, to take it in its most primitive form: suppose we wish to add x and y. Take a huge bunch of marbles. First count out x marbles in one heap. Then count out y marbles in another. Put the two heaps together and count out the number of marbles in the union thus formed. The result is x+y. This set of directions, I may suppose, I explicitly gave myself at some earlier time. It is engraved on my mind as on a slate. It is incompatible with the hypothesis that I meant quus. It is this set of directions, not the finite list of particular additions I performed in the past that justifies and determines my present response.

Kripke continues:

> (p.36) Despite the initial plausibility of this objection, the sceptic's response is all too obvious: True, if 'count' as I used the word in the past, referred to the act of counting (and my other words are correctly interpreted in the standard way) then 'plus' must have stood for addition. But I applied 'count' like 'plus' to only finitely many past cases. Thus the sceptic can question my present interpretation of my past usage of 'count' as he did with 'plus'.[15]

How should we understand this passage? On one way of reading it, Kripke would be assuming that the contents of mental states are derived from the contents of linguistic expressions. But if that's the assumption, it is vulnerable: most philosophers think that the relation between mind and language is in fact the other way round, that linguistic meaning derives from mental content.

On another way of reading it, Kripke would be assuming not some controversial view of the relation between mind and language, but rather just the familiar 'language of thought' picture of thought—that thoughts themselves involve the tokenings of expressions (of mentalese)—and claiming that those expressions, too, get their meaning by our following rules in respect of them.

I think this latter assumption is clearly what Kripke had in mind. Let's call it Kripke's *Meaning Assumption* and let's go along with it for now.

Now, it should be obvious that combining the Meaning Assumption with the Intention View will lead rather quickly to the conclusion that rule-following, and with it mental content, are metaphysically impossible. For given the two assumptions, we would be able to reason as follows. In order to follow rules, we would antecedently have to have intentions. To have intentions, 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 intentions. And so neither content nor rule-following would be able to get off the ground.

Since Kripke regards the Meaning Assumption as non-optional, he rejects the Intention View. The problem then becomes to find a way in which someone could be said to have committed himself to a certain pattern of use for a symbol without this being the result of his forming an *intention* (or other intentional state) to uphold that pattern. And that is why so much attention is focused on the dispositional view.

5. Are there Solutions to these Problems for the Intention View?

Let us take stock. If the second of the three considerations we have outlined is correct, then there must be a species of rule-following that is non-intentional.

And, if either the first or the third of our three considerations is correct, then not only must there be a species of rule-following that is non-intentional, *all* rule-following (p.37) must be at bottom non-intentional, because even *intentional* forms of rule-following will presuppose the *non-intentional* kind.

Now, since we know that it is going to be extremely difficult to make sense of rule-following in purely non-intentional, dispositional terms, we should ask whether there is any way around these considerations. Can they be rebutted?

To the first objection, one might respond by saying that Naturalism is not obviously correct and so can hardly be used to constrain the acceptability of an otherwise intuitively compelling theory of rule-following. After all, it continues to prove difficult to account for consciousness within a naturalistic setting. Kripke tried to argue that there would be something irredeemably queer about an intentional state that is simply taken unreduced, without hope of reconstruction in terms of materials that would be naturalistically acceptable. But I agree with Wright that his argument here is not successful and doesn't rise above begging the question against the anti-reductionist suggestion.

To the objection based on cases of alleged non-intentional rule-following, one could try responding by introducing the notion of a *tacit* intention, an intention to do something that is not explicitly articulated in someone's consciousness but which the thinker could be said to have implicitly or tacitly. Specifying such a notion in a satisfactory way would obviously be a huge undertaking and has not yet been shown to be feasible. But it is not clearly hopeless.

However, even if the foregoing responses were accepted, I hope it is clear that we would still be stuck with a huge problem for the Intention View, if Kripke's Meaning Assumption is left in place. The problem, of course, is that even unreduced, tacit intentions are contentful states and so it would still not be possible to combine the Intention View with the Meaning Assumption. But can the Meaning Assumption be plausibly discarded?

Let's distinguish between the question whether *public language* expressions get their meaning through rule-following and the question whether the *expressions of the language of thought* do.

Is the Meaning Assumption correct at least when it comes to the words of public language? Is it right to say that the words of English, for example, get their meaning as a result of our following rules in respect of them?

Well, a word is just an inscription, a mark on paper. Something has got to be done to it by its user for it to get a meaning. That much is clear.

It is also clear that meaningful words have conditions of *correct application*. Thus, the word 'tiger' is correctly applied only to tigers and the word 'red' only to red things.

But it doesn't follow from these obvious truths that the way the word 'tiger' comes to mean what it does for a given speaker S—the way it comes to have the correctness conditions that it has in S's idiolect—is by S committing himself to using it according to the rule: Apply the word 'tiger' only to tigers![16]

(p.38) For meaning to be a matter of rule-following in the sense presupposed by the Meaning Assumption, it must be true not only that words *have* satisfaction conditions but that they *get* their satisfaction conditions by their users committing themselves to using them according to certain patterns.

Still, it does look as though one can make a strong case for the Meaning Assumption as applied to public language expressions.[17] When I apply the word 'tiger' to a newly encountered animal, it is very natural to think that my application of the word is guided and rationalized by my understanding of its meaning, an understanding that is general and which determines what the word does and does not apply to.

However one may feel about the relation between public language and rules, though, there looks to be very little prospect that the Meaning Assumption applies at the level of *mental* expressions.

Since we are dealing with a personal-level notion of rule-following, it makes very little sense to say that we follow rules in respect of our mental expressions, expressions that we have no access to and which, for all that the ordinary person knows, may not even exist. (Whether we should regard their meaning as generated by subpersonal rule-following is a question that I shall come back to.)

So, here, then, is the problem for Kripke's Meaning Assumption that I alluded to at the beginning. Kripke is clearly working with a personal-level notion of rule-following. That is why he can confidently claim that when someone is following a rule that rule *justifies* (rationalizes) his behavior. But it can hardly be true that all meaning is a matter of rule-following in this sense. In particular, it can hardly be true that the expressions of mentalese get their meaning by our following rules in respect of them in this sense.

So, it looks as though we are free to reject Kripke's Meaning Assumption, at least as it applies to mental expressions. And with that we seem to have answered the third of the three objections we had posed for the Intention View.

If we reject the Meaning Assumption, we give up on the claim that mental expressions get their

meaning by our following rules in respect of them. How, then, do they get their meaning?

Kripke's discussion may be seen as containing a battery of effective arguments against *reductive* accounts of meaning facts. But as I have already mentioned above, and argued at length elsewhere, his arguments against *anti-reductionist* accounts are rebuttable.[18]

If we adopt such an anti-reductionist conception of mental content, doesn't that mean that we are now free to adopt the Intention View of rule-following?

(p.39) 6. Can the Intention View be Saved?

Not quite. For what I now want to argue is that even if *all* of these responses were to pan out, that still wouldn't suffice to salvage the Intention View. The Intention View suffers from a further and seemingly fatal flaw. It concerns not, as on Kripke's view, Acceptance, but rather, Explanation or Rationalization.

To see what it is, let us waive Naturalism; let us ignore the examples of putatively non-intentional forms of rule-following; and let us reject the Meaning Assumption. And let us simply help ourselves to an anti-reductionist view of mental content.

Once such contentful thoughts are available, they can be used to frame intentions—and so, it would seem, to account for our acceptance of rules. If something like this picture could be sustained, would that imply that there is nothing left of the rule-following problem?

In a passage whose import I believe many commentators have missed, Wittgenstein seems to indicate the answer is No—even if we could simply help ourselves to the full use of intentional resources, Wittgenstein appears to be saying, there would *still* be a problem about how rule-following is possible.

The passage I have in mind is at *Philosophical Investigations* 219. In it Wittgenstein considers the temptation to say that when we commit ourselves to some rule, that rule determines how we are to act in indefinitely many future cases:

> "All the steps are really already taken," means: I no longer have any choice. The rule, once stamped with a particular meaning, traces the lines along which it is to be followed through the whole of space.

If we were reading this with Kripke's eyes, what would we expect Wittgenstein to say in reply? Something along the following lines (with absolutely no aspiration to capturing Wittgenstein's literary tone):

> And how did you get to stamp the rule with a particular meaning so that it traces the lines along which it is to be followed through the whole of space? To do that you would need to be able to think, to frame intentions. But that assumes that we have figured out how we manage to follow rules in respect of mental expressions. And that is something that we have not yet done.

But what Wittgenstein says in reply is rather this:

> But if something of this sort really were the case, how would it help?

Even if we were to grant that we could somehow imbue the rule with a meaning that would determine how it applies in indefinitely many cases in the future, Wittgenstein seems to be saying, it would still not help us understand the phenomenon of rule-following.

How mystifying this must seem from a Kripkean point of view. How would it help? How could it not

help? We wanted an answer to the question: By virtue of what is it true that I use the '+' sign according to the rule for addition and not some other rule? (p.40) According to the picture currently under consideration, one of our options is to say that it is by virtue of the fact that I use the '+' sign with the intention that its use conform to the rule for addition and where it is understood that the availability of such intentions is not itself a function of our following rules in respect of them. Under the terms of the picture in place, what would be left over?

How should we understand what Wittgenstein is saying here? It is, of course, always hard to be confident of any particular interpretation of this philosopher's cryptic remarks; but here is a suggestion that seems of independent philosophical interest.

Let us revert to our email example. Suppose I have adopted the rule: Answer any email (that calls for an answer) immediately upon receipt. And let us construe my adoption of this rule as involving an explicit intention on my part to conform to the instruction:

Intention: For all x, if x is an email and you have just received x, answer it immediately!

Now, how should we imagine my following this rule? How should we imagine its guiding, or explaining, the conduct that constitutes my following it?

To act on this intention, it would seem, I am going to have to think, even if very fleetingly and not very consciously, that its antecedent is satisfied. The rule itself, after all, has a conditional content. It doesn't call on me to just do something, but to always perform some action, if I am in a particular kind of circumstance. And it is very hard to see how such a conditional intention could guide my action without my coming to have the belief that its antecedent is satisfied. So, let us imagine, then, that I think to myself:

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

in order to draw the

Conclusion: Answer it immediately!

At least in this case, then, rule-following, on the Intention model, requires *inference*: it requires the rule-follower to infer what the rule calls for in the circumstances in which he finds himself.

At least in this case, then, rule-following, on the Intention model, requires some sort of inference.

In this regard, though, the email case is hardly special. Since *any* rule has *general* content, if our acceptance of a rule is pictured as involving its representation by a mental state of ours, an inference will always be required to determine what action the rule calls for in any particular circumstance. On the Intention View, then, applying a rule will always involve inference.

Inference, however, is an example of rule-following *par excellence*. In the email case, in moving from the intention, via the premise about the antecedent, to the conclusion, I am relying on a general rule that says that from any such premises I am (p.41) entitled to draw such-and-so conclusion. Since, as I have set up the example, I have construed the email rule as an imperative, this isn't quite Modus Ponens, of course, but it is something very similar:

(MP*) From 'If C, do A' and C, conclude 'do A'!

But now: If on the Intention View, rule-following always requires inference; and if inference is itself always a form of rule-following, then the Intention View would look to be hopeless: under its terms, following any rule requires embarking upon a vicious infinite regress in which we succeed in following

no rule.

To see this explicitly, let us go back to the email case. On the Intention View, applying the Email Rule requires, as we have seen, having an intention with the rule as its content and inferring from it a certain course of action. However, inference, we have said, involves following a rule, in this case, MP*. Now, if the Intention View is correct, then following the rule MP* itself requires having an intention with MP* as its content and inferring from it a certain course of action. And now we would be off on a vicious regress: inference rules whose operation cannot be captured by the intention-based model are presupposed by that model itself.[19]

This argument bears an obvious similarity to Lewis Carroll's famous argument in "What the Tortoise Said to Achilles."[20] The Carrollian argument, however, is meant to raise a problem for the *justification* of our rules of inference—how can we justify our belief, for example, that Modus Ponens is a good rule of inference?

The argument I am putting forward, though, raises an even more basic problem for how it is possible to follow an inference rule of any kind, good or bad, justified or unjustified. Even if it were Affirming the Consequent that was at issue, the problem I am pointing to would still arise.

It would seem, then, that there would still be a problem with the Intention View even if we somehow managed to resolve all the other difficulties that we outlined for it. The mere combination of the Intention View and a Rulish of inference are sufficient for generating a problem.

7. Intentions and Intentional States

How should we proceed? I have been talking about the Intention View, but, of course, everything I've been saying will apply to any Intentional View. So let me restate our problem in full generality exposing as many of our assumptions as possible.

(p.42) The claim is that the following five propositions form an inconsistent set.

1. 1. Rule-following is possible.
2. 2. Following a rule consists in acting on one's acceptance (or internalization) of a rule.
3. 3. Accepting a rule consists in an intentional state with general (prescriptive or normative) content.
4. 4. Acting under particular circumstances on an intentional state with general (prescriptive or normative) content involves some sort of inference to what the content calls for under the circumstances.
5. 5. Inference involves following a rule.

If my argument is correct, then one of these claims has to go.[21] The question is which one.

Giving up (1) would give us rule-following skepticism. (2) seems to be the minimal content of saying that someone is following a rule. (3) is the Intentional View. (4) seems virtually platitudinous. (How could a general conditional content of the form 'Whenever C, do A' serve as your reason for doing A, unless you inferred that doing A was called for from the belief that the circumstances are C?) (5) seems analytic of the very idea of deductive inference (more on this below).

When we review our options, the only plausible non-skeptical option would appear to be to give up (3), the Intentional View. To rescue the possibility of rule-following, it seems, we must find a way of understanding the notion of *accepting* or *internalizing* a rule that does not consist in our having some *intentional* state in which that rule's requirements are explicitly represented. Wittgenstein can be read

as having arrived at the same conclusion.

The full passage from *Investigations* 219 reads as follows:

> "All the steps are really already taken," means: I no longer have any choice. The rule once stamped with a particular meaning, traces the lines along which it is to be followed through the whole of space. But if something of this sort really were the case, how would it help?
>
> No; my description only made if it was understood symbolically.—I should have said: *This is how it strikes me.*
>
> When I obey a rule, I do not choose.
>
> I obey the rule *blindly.*

The drift of the considerations I have been presenting seems to capture the intended point behind this passage.

(p.43) Even without assuming Naturalism as an a priori constraint on the acceptability of a solution to the rule-following problem, and without assuming that mental content itself must be engendered by rule-following, it would seem that we have shown that, in its most fundamental incarnation, rule acceptance cannot consist in the formation of a propositional attitude in which the requirements of the rule are explicitly encoded.

Such a picture would be one according to which rule-following is always fully *sighted*, always fully informed by some recognition of the requirements of the rule being followed. And the point that Wittgenstein seems to be making is that, in its most fundamental incarnation, not all rule-following can be like that—some rule-following must simply be *blind*. The argument I have presented supports this conclusion.

8. Rule-Following without Intentionality: Dispositions

The question is how rule-following *could be* blind. How can someone commit himself to a certain pattern in his thought or behavior, which can then rationalize what he does, without this consisting in the formation of some appropriate kind of intentional state?

The only option that seems to be available to us is the one that Kripke considers at length, that we should somehow succeed in understanding what it is for someone to accept a given rule just by invoking his or her *dispositions* to conform to the rule. If we were able to do that, we could explain how it is possible to act on a rule without inference because the relation between a disposition and its exercise is, of course, non-inferential.

Now, Kripke, as we all know, gives an extended critique of the dispositional view. However, that critique is not generally thought to be very effective; many writers have rejected it.[22] So perhaps there is hope for rule-following after all along the lines of a dispositional understanding.

My own view, by contrast, is that Kripke's critique is extremely effective, although even I underestimated the force of what I now take to be its most telling strand. And so I think that it can't offer us any refuge, if we abandon the Intentional View.

The core idea of a dispositional account is that what it is for me to accept the rule Modus Ponens is, roughly, for me to be disposed, for any p and q, upon believing both p and 'if p, then q,' to conclude q.

Kripke pointed out that any such dispositional view runs into two problems. First, a person's

dispositions to apply a rule are bound to contain performance errors; so one can't simply read off his dispositions which rule is at work.

Second, the rule Modus Ponens, for example, is defined over an infinite number of pairs of propositions. For any p and q, p and 'if p, then q,' entail q. However, a person's dispositions are finite: it is not true that I have a disposition to answer q when asked (p.44) what follows from any two propositions of the form p and 'if p, then q,' no matter how large.

To get around these problems, the dispositionalist would have to specify ideal conditions under which (a) I would not be capable of any performance errors and (b) I would in fact be disposed to infer q from *any* two propositions of the form p and 'if p, then q.'

But it is very hard to see that there are conditions under which I would be metaphysically incapable of performance errors.

And whatever one thinks about that, it's certainly very hard to see that there are ideal conditions under which I would in fact be disposed to infer q from any two propositions of the form p and 'if p, then q' no matter how long or complex. As Kripke says, for most propositions, it would be more correct to say that my disposition is to die before I am even able to grasp which propositions are at issue.

Along with many other commentators, I used to underestimate the force of this point. The following response to it seemed compelling. A glass can have infinitary dispositions. Thus, a glass can be disposed to break when struck here, or when struck there; when struck at this angle or at that one, when struck at this location, or at that one. And so on. But if a mere glass can have infinitary dispositions, why couldn't a human being?[23]

There is an important difference between the two cases. In the case of the glass, the existence of the infinite number of inputs—the different places, angles, and locations—just follows from the nature of the glass qua physical object. No idealization is required.

But a capacity to grasp infinitely long propositions—the inputs in the rule-following case—does *not* follow from our nature as thinking beings, and certainly not from our nature as physical beings. In fact, it seems pretty clear that we do not have that capacity and could not have it, no matter how liberally we apply the notion of idealization.

These, then, are Kripke's central arguments against a dispositional account of rule-following, and although it would take much more elaboration to completely nail these arguments down, I believe that such an elaboration can be given.

But both before and after he gives those arguments, Kripke several times suggests that the whole exercise is pointless, that it should simply be *obvious* that the dispositional account is no good. Thus, he says:

> To a good extent this [dispositional] reply ought to appear to be misdirected, off target. For the skeptic created an air of puzzlement as to my *justification* for responding '125' rather than '5' to the addition problem…he thinks my response is no better than a stab in the dark. Does the suggested reply advance matters? How does it *justify* my choice of '125'? What it says is "'125' is the response you are disposed to give…" Well and good, I know that '125' is the response I am (p.45) disposed to give (I am actually giving it!)…How does any of this indicate that…'125' was an answer *justified* in terms of instructions I gave myself, rather than a mere jack-in-the-box unjustified response?

This passage can seem puzzling and unconvincing when it is read, as Kripke seems to have intended it, as directed against dispositional accounts of mental content. After all, one of the most influential views

of mental content nowadays is that expressions of mentalese get their meaning by virtue of their having a certain causal role in reasoning. Could it really be that this view is so obviously false that it's not worth discussing, as Kripke suggests? And is it really plausible that the facts by virtue of which my *mentalese* symbol '+' means what it does have to *justify* me when I use it one way rather than another?

But if we see the passage as directed not at dispositional accounts of mental content but rather at dispositional accounts of personal-level rule-following, and if we substitute "rationalize" for "justify," then its points seem correct. It should be puzzling that anyone was inclined to take a dispositional account of rule-following seriously. We can see why in two stages (this is a different argument than the one Kripke gives).

First, and as I have been emphasizing, if I am following the rule Modus Ponens, then my following that rule explains and rationalizes my concluding q from p and 'if p, then q,' (just as it would be true that, if I were following the rule of Affirming the Consequent, then my following that rule would explain and rationalize my inferring q from p and 'if q, then p').

Second, if I am following the rule Modus Ponens, then not only is my *actually* inferring q explained and rationalized by my accepting that rule, but so, too, is my being *disposed* to infer q. Suppose I consider a particular MP inference, find myself disposed to draw the conclusion, but, for whatever reason, fail to do so. That disposition to draw the conclusion would itself be explained and rationalized by my acceptance of the MP rule.

However, it is, I take it, independently plausible that something can neither be explained by itself, nor rationalized by itself. So, following rule R and being disposed to conform to it cannot be the same thing.

Here we see, once again, how Kripke's Meaning Assumption gets in the way of his argument: a good point about rule-following comes out looking false when it is extended to mental content.

9. Is Going Subpersonal the Solution?

It might be thought a crucial assumption of the preceding argument is that all rule-following is personal-level rule-following and that its moral, therefore, should be that at least some rule-following is *subpersonal*. In particular, would our problem about rule-following disappear if we construed the inferences that mediate between intentions and actions in some subpersonal way?

(p.46) This suggestion resonates with a number of discussions of the Intentional View that may be found in the literature. These discussions tend to accuse the Intentional View of being 'overly intellectualized' and recommend substituting a subpersonal notion in its place.[24] It isn't very often made clear exactly what that is supposed to amount to. The preceding discussion should help us see that this is not a particularly useful suggestion.

In the present context, going subpersonal presumably means identifying rule acceptance or internalization not with some person-level state, such as an intention, but with some subpersonal state. Such a state will either be an intentional state or some non-intentional state.

Let us say that it is some intentional state in which the rule's requirements are explicitly represented. Then, once again, it would appear that some inference (now, subpersonal) will be required to figure out what the rule calls for under the circumstances. And at this point the regress problem will recur. (That is what I meant by saying earlier that the structure of the regress problem seems to be indifferent as to whether the states of rule acceptance are personal or subpersonal.)

On the other hand, we could try identifying rule internalization with some non-intentional state. Indeed, even if the state of rule internalization is initially identified with a subpersonal intentional state, it will

ultimately, I take it, have to be identified with some sort of non-intentional state.

But then what we would have on our hands would be some version or other of a dispositional view (with the dispositions now understood subpersonally). And although we would no longer face the rationalization problem—because subpersonal mechanisms are not called upon to rationalize their outputs—we would still face the enormous problems posed by the error and finitude objections.

In consequence, I don't believe that going subpersonal offers any sort of panacea for our problems.

10. Conclusion

In my 1989 paper, "The Rule-Following Considerations," I was concerned to explicate and assess Kripke's arguments for his rule-following skepticism. However, since at that time something like Wright's Intention View about rule-following struck me as correct, I thought that the interesting issues had to center on the notion of mental content rather than on rule-following in particular. Rule-following, like any other intentional process, I thought, came along for the ride.

Viewed in that light, Kripke's arguments were most effectively seen as arguments against various attempts to naturalize intentional content. And I believed then, and continue to believe now, that his arguments are very effective against that particular target.

(p.47) The upshot, however, was that the so-called "rule-following considerations" had very little to do with rule-following *per se* and a great deal to do with meaning and content—as I noted at the time.

I now believe that this was a mistake, induced by the illusory plausibility of the Intention View. In its most fundamental incarnation, rule acceptance cannot consist in an intentional state. For if it did, rule-following would have to be inference; and we know that, whatever else it may be, it cannot be *that*.

In its most fundamental incarnation, rule-following is something that must be done blindly, without the benefit of some intentional encoding of the rule's requirements. The question is whether that is so much as possible.

The only way to make sense of it, it would seem, is if rule acceptance could be understood dispositionally. But there seem to be powerful objections to any such understanding.

As a result, it is hard to explain how rule-following is so much as possible, and this difficulty arises even without our having to assume (in the way that Kripke effectively does) that intentional states need to be given a naturalistic reduction.

What are we to do?

Perhaps we should embrace rule-skepticism, denying that our reasoning is under the influence of general rules?

The trouble is that this seems not only false about reasoning in general, but also unintelligible in connection with deductive inference. It is of the essence of deductive inference that the reasons I have for moving from certain premises to certain conclusions are general ones.

So what we are contemplating, when we contemplate giving up on the Rulish picture of deductive inference, is not so much giving up on a Rulish *construal* of deductive inference as giving up on deductive inference itself.

But that is surely not stable a resting point—didn't we arrive at the present conclusion through the application of several instances of deductive inference?

Hence we have what I have somewhat grandly called an "antinomy of pure reason": we both must—and cannot—make sense of the notion of someone's following a rule.

The only non-skeptical option that seems open to us is to try taking the notion of following—or applying—a rule as primitive, effectively a rejection of proposition (4) above.

Notice that this goes well beyond the sort of anti-reductionist response to Kripke's arguments that I was already inclined to favor—an anti-reductionism about mental content.

It would involve a primitivism about rule-following or rule-application itself: we would have to take as primitive a *general (often conditional) content serving as the reason for which one believes something*, without this being mediated by inference of any kind. It is not obvious that we can make sense of this, but the matter clearly deserves greater consideration.

(p.48) Note

This essay draws heavily on my 2008. A very early version of some of its arguments appeared in Boghossian 2005, as part of a symposium on Philip Pettit's 2002 book. I have benefited greatly from feedback over the intervening years from various audiences—at various seminars at NYU in 2001, 2003, and 2006, the Graduate Conference at the University of Warwick, the Workshop on Epistemic Normativity at Chapel Hill, UCLA, Stony Brook, Princeton, and the Transcendental Philosophy Network Workshop in London, and the Workshop on Wittgenstein in Tokyo, to name just those that come to mind. I am also grateful to, Sinan Dogramaci, Paul Horwich, Matthew Kotzen, Christopher Peacocke, Josh Schechter, and Stephen Schiffer for valuable comments on earlier drafts.

Bibliography

Bibliography references:

Boghossian, P. A. 1989 "The Rule-Following Considerations," *Mind* 98/392, pp. 507–49.

Boghossian, P. A. 2005 "Meaning, Rules and Intention," *Philosophical Studies* 124/2, pp. 185–97.

Boghossian, P. A. 2008 "Epistemic Rules," *Journal of Philosophy* CV, 9, pp. 472–500.

Carroll, L. 1895 "What the Tortoise Said to Achilles," *Mind* 4/14, pp. 278–80.

Cruz, J. and Pollock, J. 1999 *Contemporary Theories of Knowledge*, Oxford, Rowman and Littlefield.

Fodor, J. 1990 *A Theory of Content and Other Essays*, Cambridge, Mass., MIT Press.

Horwich, P. 1998 *Meaning*, Oxford, Oxford University Press.

Kripke, S. 1982 *Wittgenstein on Rules and Private Language*, Cambridge, Mass., Harvard University Press.

Pettit, P. 2002 *Rules, Reasons and Norms*, Oxford, Oxford University Press.

Quine, W. V. O. 1936 "Truth by Convention," in W. V. O. Quine 1976 *The Ways of Paradox and Other Essays*, Cambridge, Mass., Harvard University Press, pp. 77–106.

Soames, S. 1998 "Skepticism About Meaning: Indeterminacy, Normativity and the Rule Following Paradox," *Canadian Journal of Philosophy* 23, pp. 211–50.

Wittgenstein, L. 1953 *Philosophical Investigations*, ed. G. E. M. Anscombe and R. Rhees, trans. G. E. M. Anscombe, Oxford, Blackwell.

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

Wright, C. 2001 *Rails to Infinity: Essays on Themes from Wittgenstein's* Philosophical Investigations, Cambridge, Mass., Harvard University Press.

Wright, C. 2007 "Rule-Following Without Reasons: Wittgenstein's Quietism and the Constitutive

Question," in J. Preston (ed.) *Wittgenstein and Reason, Ratio* 20/4, pp. 481–502.

Notes:

([1]) Wright's papers are collected in Wright 2001. For Kripke's view, see Kripke 1982. Wright's view of rule-following in *Rails* is substantially different from the view he presented in Wright 1980. In more recent work, Wright 2007, Wright independently explores ideas that are much closer in spirit to the conclusions of this essay than his earlier views.

([2]) Caveat: Kripke presents this as an exposition of Wittgenstein's view rather than his own.

([3]) For more discussion see Boghossian 2008.

([4]) For further discussion, see ibid.

([5]) Again, see ibid.

([6]) Wright 2001, p. 1.

([7]) "Internalization" is Kripke's preferred word, as we shall see below; it is probably more neutral than "acceptance."

([8]) Kripke 1982, p. 16.

([9]) This account summarizes a large number of considerations that I do not have the space to describe in detail here.

([10]) Wright 2001, pp. 125–6.

([11]) See also Pettit 2002, p. 27: "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."

([12]) Wright 2001, p. 125.

([13]) Wright 2001, pp. 134 ff.

([14]) For more discussion, see Boghossian 1989, pp. 544–8.

([15]) Kripke 1982, p. 16.

([16]) Jerry Fodor may have been the first to appreciate this clearly; see Fodor 1990, pp. 135–6. I don't believe that any of the main arguments of Boghossian 1989 would be affected by paying greater heed to this distinction, although I am sure I wasn't as clear about it in that paper as I should have been.

([17]) I have gone back and forth about the plausibility of the Meaning Assumption as applied to public language expressions. In my NYU seminar of Spring 2006, I defended it, but in an earlier version of this essay, I retreated to saying that it was not settled. I thank Christopher Peacocke for rightly insisting to me that it met my characterization of person-level rule-following.

([18]) See Boghossian 1989.

([19]) This, I believe, is the correct interpretation of Wittgenstein's remarks about needing a rule to interpret a rule. In the Kripkean framework, this is read as supposing that a rule can only be given to

you as an inert sign whose meaning you would then have to divine. And this sets off an infinite regress of interpretations. However, a different way of reading Wittgenstein here is to see him as concerned not with the question: "How could an inert sign guide us, if not through the use of further rules?" But rather with the question: "How could a general content guide us, if not through the use of further rules?"

([20]) See Carroll 1895. There is also a similarity to Quine's arguments in Quine 1936.

([21]) Notice that this argument is not only neutral on whether what is at issue are intentions as opposed to other sorts of intentional state, but also on whether what is at issue are *personal-level* intentional states as opposed to *subpersonal* content-bearing states. So long as you think that the acceptance of a rule consists in some sort of intentional state with general content and that, as a result, inference will be required to act on that state, there will be a problem—it doesn't matter whether this is thought of as occurring at the personal or the subpersonal level—more on this below.

([22]) See, for example, Soames 1998, Horwich 1998.

([23]) See the discussion in Boghossian 1989.

([24]) See, for example, Cruz and Pollock 1999, chapter 5.