

## EPISTEMIC RULES\*

According to a very natural picture of rational belief, we aim to believe only what is true. However, as Bernard Williams used to say, the world does not just inscribe itself onto our minds. Rather, we have to try to figure out what is true from the evidence available to us.

To do this, we rely on a set of epistemic *rules* that tell us in some general way what it would be most *rational* to believe under various epistemic circumstances. We *reason* about what to believe; and we do so by relying on a set of rules.<sup>1</sup>

Although there is some controversy about exactly how these rules are to be formulated, we take ourselves to know roughly what they are. For example, we have a rule linking visual appearances to beliefs:

(Observation) If it visually seems to you that *p*, then you are *prima facie* rationally permitted to believe that *p*.

We have some sort of *inductive* rule linking beliefs about the observed to beliefs about the unobserved, an example of which might be:

(Induction) For appropriate *F*s and *G*s, if you have observed *n* (for some sufficiently large *n*) *F*s and they have all been *G*s, then you are *prima facie* rationally permitted to believe that all *F*s are *G*s.

We also have *deductive* rules, such as:

(Modus Ponens): If you are rationally permitted to believe both that *p* and that 'If *p*, then *q*', then, you are *prima facie* rationally permitted to believe that *q*.<sup>2</sup>

\*This paper has been in the works for quite a long time. A very early version of some of its arguments appeared as "Meaning, Rules and Intention," *Philosophical Studies*, cxxxiv, 2 (2005): 185–97, as part of a symposium on Philip Pettit's *Rules, Reasons and Norms* (New York: Oxford, 2002). I have benefited greatly from feedback over the intervening years from various audiences—at various seminars at New York University, the Graduate Conference at the University of Warwick, the Workshop on Epistemic Normativity at Chapel Hill, University of California in Los Angeles, Stony Brook, Rutgers, Princeton, and the Transcendental Philosophy Network Workshop in London, to name just those that come to mind. I am also grateful to Shamik Dasgupta, Sinan Dogramaci, Paul Horwich, Matthew Kotzen, Christopher Peacocke, James Pryor, Josh Schechter, and Stephen Schiffer for valuable comments on earlier drafts.

<sup>1</sup>We could put everything in terms of partial belief, but that will not matter for our purposes.

<sup>2</sup>Of course, this is not quite the rule that is labeled Modus Ponens in logic textbooks. It is actually quite mysterious what the logic textbook rule is supposed to be, but I cannot go into that here.

These rules, and others like them, constitute what we may call our *epistemic system*. They represent our conception of how it would be most rational for a thinker to form beliefs under different epistemic circumstances.

Let us call this the *rule-following picture of rational belief*. It is a very familiar picture and has tempted many. As I shall try to explain later, its roots run very deep.<sup>3</sup>

Because we accept this picture, we take seriously a number of questions that it seems to entrain.

For example, we recognize that, in addition to the rules that we actually use, there are other rules, different from and incompatible with ours, which we might have used instead. And this seems to raise the question: Are our rules the right ones? Are they the ones that deliver genuinely justified belief?

These questions in turn raise a more fundamental one: In what sense could there be a fact of the matter as to what the right epistemic rules are? And if there is such a fact of the matter, how do we find out what it is? And what, in any case, entitles us to operate with the rules that we actually operate with?

None of these familiar and compelling questions would make much sense in the absence of the rule-following picture of rational belief. Each of them presupposes that we rely on rules in forming rational beliefs.

I find the rule-following picture, along with the questions that it entrains, as natural and as compelling as the next person. However, I have also come to worry about its ultimate intelligibility, a worry that I find myself unable to lay to rest. In this paper, I aim to explain the considerations that give rise to this worry.

I have been talking about the rule-governed picture of rational belief. But rational belief is hardly the only domain in which rule-following has been thought to play a prominent role. The sort of generalist picture I have been sketching for epistemology has of course always loomed large in ethics. We find it very natural to think that, in our moral judgments, we are guided by a set of general *moral* principles that tell us what we have most reason to do under various practical conditions.

<sup>3</sup>For explicit endorsements of the view, see, among many others, John Pollock and Joseph Cruz, *Contemporary Theories of Knowledge* (Lanham, MD: Rowman and Littlefield, 1999), chapter 5; Christopher Peacocke, *The Realm of Reason* (New York: Oxford, 2004); Ralph Wedgwood, "Internalism Explained," *Philosophy and Phenomenological Research*, LXV (2002): 349–69; and Hartly Field, "Apriority as an Evaluative Notion," in Boghossian and Peacocke, eds., *New Essays on the A Priori* (New York: Oxford, 2000), pp. 117–49.

Recent writers have complained about this generalist picture in ethics. They say that moral reasons are too *holistic* for there to be general principles that can tell us what it would be morally correct to do under varying practical conditions.<sup>4</sup>

That is not the sort of problem I have in mind for the generalist picture of rational belief. Rather, I will develop two other types of difficulty.

The first concerns how to understand the notion of a “rule” as it is used in the rule-following picture. What exactly is it that we are being said to follow, when we are said to follow epistemic rules?

The second difficulty concerns what it is to *follow* a rule regardless of how exactly a rule is construed. My worry here is closely related to the famous discussion of following a rule that was inaugurated by Wittgenstein and brilliantly expounded by Saul Kripke.<sup>5</sup> Like Kripke, I think that there really is a skeptical problem about rule-following that can be derived from Wittgenstein’s discussion.

But my problem is not Kripke’s. Unlike Kripke’s problem, my problem arises in an especially virulent form for *epistemic rules*, as opposed to rules of other kinds. And it cannot be solved, as Kripke’s problem can, by our helping ourselves to various forms of anti-reductionist conceptions of meaning or content.

All of this is what I propose to explain in what follows.

#### I. WHAT DO WE FOLLOW: IMPERATIVES OR NORMATIVE PROPOSITIONS?

*Imperatives versus Norms.* We talk interchangeably about epistemic *rules* and about epistemic *norms*. Are these the same sorts of thing or are there important differences between them? This is an area in which our language is sloppy and we are not often very explicit about what we mean.

Take the word ‘rule’. By and large, when Kripke talks about “rules” he is talking about general imperatival contents of the form:

If *C*, do *A*!

where ‘*C*’ names a type of situation and ‘*A*’ a type of action. On this construal, rules are general contents that *prescribe* certain types of behavior under certain kinds of condition.

However, not everything that we call a rule in ordinary language 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.

<sup>4</sup> See, for example, Jonathan Dancy, *Ethics without Principles* (New York: Oxford, 2006).

<sup>5</sup> See Wittgenstein, *Philosophical Investigations* (Cambridge, UK: Blackwell, 1953); Kripke, *Wittgenstein on Rules and Private Language* (Cambridge: Harvard, 1982).

This does not look like an imperative. Unlike an imperative, it seems truth-evaluable. It looks more like something we should call a *normative proposition* or norm, for short. It is a norm of *permission*. In addition to the permissive norms, of course, there are norms of *requirement*:

(First Move) At the beginning of the game, White *must* make the first move.

*Arguments for the Propositional Construal.* We need, then, to recognize a distinction between two different kinds of content—the imperatival and the propositional; and we need to clarify whether, in talking about epistemic rules, we are talking about contents of the one type or the other.

When I gave a rough characterization of these rules above, I gave them a normative propositional formulation. There are at least three considerations that favor this construal.

To begin with, epistemic justification is a normative notion. We would expect, therefore, that the contents that encode our conception of it would be normative contents. However, imperatives are not normative in any way. They are merely commands or instructions.<sup>6</sup> If such commands or instructions do play a role in our epistemic systems, it is natural to think of them as having a derivative status—a status derived from the more fundamental normative propositions that encode our conception of epistemic justification.

The second reason for favoring a propositional construal has to do with our need to distinguish between different kinds of action-guiding or belief-guiding rules. Thus, there are epistemic rules, prudential rules, aesthetic rules, moral rules, and so forth. It is easy to distinguish among these types of rule in propositional terms, by building their identity into their propositional content. Thus, an epistemic rule would be a normative proposition of the following kind:

If *C*, then *S* is epistemically permitted to believe that *p*.

A prudential rule, on the other hand, would involve the concept of a prudential permission; and so forth.

By contrast, it is hard to see how to get this differentiation on an imperatival picture. The trouble is that all imperatives are alike—they all assume the form

If *C*, do *A*!

<sup>6</sup> A point emphasized to me by Derek Parfit.

And so the mere content of an imperative is incapable of telling us whether it is an epistemic, prudential, or moral imperative.<sup>7</sup>

A third reason for favoring a propositional construal of epistemic rules has to do with the need to capture not only requirements but permissions as well. The trouble, however, is that there looks to be a real difficulty capturing a norm of *permission* in imperatival terms.

The difficulty, in a nutshell, is this: An imperative, by definition, tells you to do something, if a certain condition is satisfied. However, a norm of permission does not call on you to do anything; it just says that, if a certain condition were satisfied, then performing some particular action would be alright.

Thus, obviously, the norm of permission (Castle) cannot be expressed in terms of the imperative

If configuration is *C*, Castle!

because that would suggest that whenever the configuration is *C* one must castle, whereas the norm merely permits castling and does not require it. Could we perhaps express (Castle) as:

Castle!, only if *C*.

But this seems to want to embed an imperative in the antecedent of a conditional:

If Castle!, then *C*

and I do not know what that means.

Gideon Rosen has suggested another strategy for the imperativalist—using complex imperatives with disjunctive consequents.<sup>8</sup> Thus, he suggests that the imperative that corresponds to an epistemic norm of permission of the form:

- (4) If for some *e*,  $f(e, h)$ , then it is rationally permissible to believe *h*  
(on the basis of *e*)

would be something more like this:

- (5) If for some *e*,  $f(e, h)$ , then either believe *h* (on the basis of *e*), or suspend judgment about *h*.

<sup>7</sup> One idea about how to remedy this would be to look at a thinker's *grounds* for accepting any given imperative—the idea being to try to distinguish between an epistemic imperative and a prudential one not in terms of their overt grounds but in terms of the characteristic grounds on which they are accepted. But this is a difficult program to execute because it depends on the not obviously correct idea that, corresponding to each type of norm, there exists an individuating type of ground on which a thinker accepts it.

<sup>8</sup> See his "The Case against Epistemic Relativism: Reflections on Chapter 6 of *Fear of Knowledge*," *Episteme*, iv, 1 (2007): 49–65.

Now, I take it that “suspending judgment” about  $h$  is not simply: not believing  $h$ . If it were, then the imperative at (5) would amount to saying:

If  $e$ , then either believe  $h$  or do not believe  $h$ !

which does not say much of anything. Suspending judgment, then, requires something active—considering whether  $h$  and then rejecting taking a view on the matter.

If that is right, though, (5) now seems to call for you to do things that go well beyond what (4) says. According to (4), if a certain kind of evidence is available, then, *if* you believed  $h$  on its basis, that belief would be justified. (4) does not say that you should believe  $h$ ; it does not say that you should consider whether  $h$ ; it does not say that you should *do* anything.

In other words, (5) is most naturally seen not as the imperatival counterpart of the norm of permission formulated in (4) but as the imperatival counterpart of the norm of *requirement* formulated in

(6) If for some  $e, f(e, h)$ , then you are required either to believe  $h$  (on the basis of  $e$ ) or to suspend judgment on  $h$ .

Would we do better with something more along the lines of (7) rather than (5)?

(7) If for some  $e, f(e, h)$ , then either believe  $h$  (on the basis of  $e$ ) or do not do anything (on the basis of  $e$ )!

But this does not seem right, either. Even without going into the details of what it might mean for someone to “not do something on the basis of  $e$ ,” I hope it is clear that, whatever exactly it means, if, in response to  $e$ , I scratched my nose on the basis of  $e$ , I would not have done anything that is in violation of the norm of permission issued by (4).

There are, no doubt, many other proposals that could be considered, but I hope it is clear that there really is a problem capturing a norm of permission in imperatival terms. An imperative, however disjunctive its consequent, will require you to do something, or to refrain from doing something; but a norm of permission does not say anything about anyone’s doing anything, or refraining from doing anything. It just says that, under the appropriate conditions, *if* one were to do something, doing that thing would be alright.<sup>9</sup>

<sup>9</sup> There are a number of other proposals that we could consider, but I cannot go into them more here. Probably the most promising is the one employed by Allan Gibbard: think of accepting a rule of permission as consisting in the *rejection* of a rule of requirement. So accepting the permissibility of castling under  $C$  would consist in rejecting the rule: If  $C$ , don’t castle! But we are now owed an account of what it is to reject an imperative. See Gibbard, *Thinking How to Live* (Cambridge: Harvard, 2003).

*Arguments for the Imperative Construal.* These, then, are some of the considerations that push one in favor of a propositional view of rules. On the other hand, there is the following argument that pushes one in the opposite direction.

Recall that the picture we are working with says that it is necessary and sufficient for a belief to be rational that it be held in accordance with the correct epistemic rules. In other words, we are working with:

(RuleRatBel) *S*'s belief that *p* is rationally permitted if and only if *S* arrived at the belief that *p* by following the correct rule *N*.

Now, suppose we take *N* to be an epistemic normative proposition of the form:

(EpNorm) If *C*, then *S* is rationally permitted to believe that *p*.

Now, EpNorm—the norm we are said to be following—says that it is sufficient for my being rationally permitted to believe that *p* that condition *C* obtains.

However, the rule-following picture of rational belief (RuleRatBel) implies that it is not sufficient for my being rationally permitted to believe that *p* that *C* obtains—in addition, *I need to have followed the rule EpNorm*.

If we put these two facts together, we get the following peculiar result: The only way to implement the rule-following picture of rational belief, with the rules construed as normative propositions, is to accept that the normative propositions that we are required to follow, in order to acquire rational belief, must be *false* epistemic propositions! To have rationally permitted beliefs a thinker is required to follow *false* epistemic normative propositions.

And that is surely very odd. How could it be that, in order to arrive at genuinely rationally permitted beliefs, I must be armed with, and guided by, a set of false epistemic propositions about the conditions under which a belief would be genuinely epistemically justified?

It is important to note two points. First, the problem here is structural. Whatever proposition we replace (EpNorm) with, we will face some version or other of this false rules problem, because the rule-following picture will always insist on imposing a *further* necessary condition on rational belief beyond that recognized by the proposition that is said to constitute an epistemic rule—namely, the condition that that rule be followed.<sup>10</sup>

<sup>10</sup> It might be thought that some self-referential device might meet this problem. Perhaps we should think of the epistemic rules as consisting in propositions of the following form:

(EpNorm\*) If *C*, then if *S* were to believe that *p* on the basis of this very norm, he would be rationally permitted to believe that *p*.

Second, this problem of false rules would *not* arise on the imperatival picture of epistemic rules, on which the rules are of the form:

If *C*, believe that *p*!

Since, on this conception, the rules themselves do not make any claims, they can hardly *conflict* with the claims being made by the rule-following picture of rational belief (RuleRatBel) about the conditions necessary for rational belief.

That constitutes a significant argument in favor of an imperatival construal of epistemic rules. The trouble is that, on such a construal, we would face all the other problems outlined above.

This, then, is the first difficulty I wanted to raise for the rule-following picture of rational belief: it is very unclear what satisfactory answer we can give to the question: What sort of content can a rule be such that following it is necessary for a belief to be rational?<sup>11</sup>

## II. HOW CAN WE FOLLOW RULES?

*The Intuitive Notion.* Let us assume, though, for the purposes of argument, that we have a satisfactory solution to this problem. Let us now turn to asking how it is possible for someone to *follow* a rule. For the purpose of posing this question it will not much matter whether we construe rules in imperatival or propositional terms.

Before proceeding we should clarify what we mean to be asking about. What intuitive phenomenon is at issue when we talk about someone *following* a rule?

In answering this question, we should distinguish between a *personal-level* notion of rule-following and a *sub-personal* 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.

We should also recognize that, *prima facie*, anyway, it is the personal-level notion that is involved in the generalist, rule-following picture of rational belief with which we are concerned. *I* reason about what to believe, not a part of my brain.

I propose, therefore, to start with attempting to understand the personal-level notion, returning to the sub-personal notion later. My view will be that there is a core concept that is common to both no-

---

This suggestion is worth exploring, although, for obvious reasons, I am always leery of self-referential devices and am not sure I understand them.

<sup>11</sup> Limitations of space prevent me from considering various ways of responding to this difficulty for the propositional construal. For further discussion, see my *Rules and Intentionality* (in preparation).

tions, 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 sub-personal notion.

A propos 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 happens to conform to *R*. Conforming to *R* is neither necessary nor sufficient for following *R*.

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 conditions *C*, but, as a result of a performance error, fail to do *A*, even though he tries.

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.

Crispin Wright, for example, often introduces the topic of rule-following with something like the following remark:

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>12</sup>

The suggestion seems to be that rule-following and normative constraint come to much the same thing. Or, if not quite that, that rule-following on *S*'s part is necessary for *S*'s behavior to be subject to normative assessment.

But this seems wrong. 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

<sup>12</sup> Wright, *Rails to Infinity* (Cambridge: Harvard, 2001), p. 1.

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, norms or rules *apply* to some thought or behavior even though there is no intuitive sense in which the agent in question was attempting to observe those norms or follow those rules himself.

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 does not. What, then, do we mean?

Let us take a clear case. 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 happened to do that was in conformity with that rule?

Clearly, the answer is 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 *because* of some appropriate relation that I bear to the Email Rule that I reply immediately.

I shall refer to this relation 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.<sup>13</sup>

<sup>13</sup> 'Internalization' is Kripke's preferred word, as we see below; it is probably more neutral than 'acceptance'.

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 my acceptance of that rule serves as my *reason* for replying immediately, when that acceptance *rationalizes* my behavior.

However exactly the notion of acceptance or internalization 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 nondeviant causal chain leading from that state to a piece of behavior that would allow us to say that one’s acceptance of the rule explains and rationalizes the behavior.

Occasionally, I will also describe the matter in terms of the language of commitment: In rule-following 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.

It will be up to the reader to discern whether I have loaded these notions in a way that is illicit. 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 a propos of following the rule for addition:

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 (*op. cit.*, p. 16).

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 Affirming the Consequent or Gambler’s Fallacy than it does for Modus Ponens. If I am following Gamblers’ Fallacy, my betting the house on black after a long string of reds at the roulette wheel would not 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 the house on black.

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*; incorrectly otherwise.

(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*.

Against the backdrop of this characterization of the personal-level notion, we can see the sub-personal 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 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 does not act for reasons, much less general ones.

*Following Epistemic Rules.* If we apply this analysis to the rule-following picture of rational belief with which we began, we arrive at the view that our internalization of general epistemic rules—like Modus Ponens and Induction—explain and rationalize why we form the beliefs that we form. And that seems intuitively correct.

As in the case of our linguistic and conceptual abilities, our ability to form rational beliefs is *productive*: on the basis of finite learning, we are able to form rational beliefs under a potential infinity of novel circumstances. The only plausible explanation for this is that we have, somehow, internalized a rule that tells us, in some general way, what it would be rational to believe under varying epistemic circumstances.

Furthermore, we form beliefs for reasons. As Kripke likes to say, when we form the belief that  $68+57=125$ , that does not feel like a stab in the dark, a result that is spat out by some subpersonal mechanism that we find ourselves giving and which, to our surprise, turns out to be reliable.

Rather, the processes by which we fix beliefs are personal-level processes, processes of which we are, in some appropriate sense, aware. In that appropriate sense, we know why, on any given occasion, we are inclined to believe what we believe, what our grounds are.

Combining these two natural thoughts gives us the personal-level rule-following picture of rational belief. And a very natural picture it is. The picture is perhaps most obviously at work in the case of deductive reasoning; but it applies equally to inductive reasoning, arithmetical reasoning and moral reasoning. Let us take a somewhat closer look at the deductive case.

Suppose someone asks me to accept that

Mitochondria are mitochondria.

Even if I knew very little about what mitochondria are, I would be very confident that I should accept this proposition. What could be the reason for my confidence if not that I have accepted the general principle:

Accept any proposition of the form 'All  $F$ s are  $F$ s'.

Or consider the inference:

If  $x$  is a Malament-Hogarth space-time, then it has no Cauchy surface.

$x$  is a Malament-Hogarth space-time.

Therefore,

$x$  doesn't have a Cauchy surface.

Once again, I may know very little about the ingredient concepts. But I can be very confident that, if I were justified in believing the premises, I would be justified in believing the conclusion. Once more, the only plausible explanation is that I have internalized (or accepted) a general Modus Ponens rule.

*Acceptance and Intention.* Let us turn now to asking why there is supposed to be a problem about rule-following. Why, in particular, does Kripke's Wittgenstein maintain that it is not possible for us to follow rules?

Kripke's problem is focused on *Acceptance*. He is struck by the fact that the patterns to which we are said to be able to commit ourselves are *infinitary* patterns. Thus, we claim to follow the rule of inference Modus Ponens:

(Modus Ponens): If you are rationally permitted to believe both that  $p$  and that 'If  $p$ , then  $q$ ', then, you are *prima facie* rationally permitted to believe that  $q$ .

MP, however, is defined over an infinite number of possible propositions. How is it possible, Kripke asks, for a thinker to commit himself to uphold this potentially infinitary pattern? Kripke despairs of answering this challenge head-on.

As we all know, Kripke's argument proceeds by elimination. There look to be only two serious candidates for constituting the state of rule acceptance: either it consists in some *intentional state* of a thinker, or it consists in his *dispositions*, very broadly understood, to use that symbol in certain ways. And he finds fault with both options.

Let us go along for now with the rejection of the dispositional suggestion. Still, what could possibly be wrong with invoking some intentional notion, as Crispin Wright has done? As Wright 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 (*op. cit.*, pp. 125–26).

Let us call this the *Intention View* of rule acceptance.<sup>14</sup>

The Intention View is itself 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.

I will focus my discussion on the Intention View but most everything I say will apply equally to the less committal Intentional View.

*Some Problems for the Intention View.* Why not just accept the Intention View? What, if anything, is wrong with this flat-footed response to the rule-following challenge?

The problem with the Intention View cannot be that there are no cases that are accurately described by it, for there clearly are. If I now adopt a policy of always answering any email that I receive immediately upon receipt and if, on some future occasion, I answer an email immediately upon receipt precisely because it is my policy to do so, then all this would be very well captured by the Intention View.

The question can only be whether, on the one hand, the Intention View is a sufficiently *fundamental* account, and, on the other, whether it is a sufficiently *general* account of rule-following, so that all relevant cases can be said to fall under it.

A reductive Naturalist would have reason to think of it as insufficiently fundamental. Such a Naturalist would insist that intentional states be shown to be naturalistically reducible before they may legitimately be appealed to in solving the rule acceptance problem. However, it is none too clear how such a reduction of the intentional to the naturalistic is to be pulled off (and Kripke's own discussion may be seen to provide a battery of arguments against its feasibility—more on this below).

Second, and even if we were to put reductive Naturalism to one side, there look to be two severe difficulties with taking the Intention View to be a sufficiently *general* account: not everything that we would intuitively count as rule-following looks like a case of acting on an intention.

<sup>14</sup> See also Philip Pettit, *Rules, Reasons and Norms* (New York: Oxford, 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."

One problem is posed by the fact that we typically think of ourselves as having quite good—indeed, especially privileged—access to our own intentions: we know without empirical investigation what they are. Yet, although we are able to give some rough indication of what our epistemic rules are, there continues to be some controversy about their precise formulation (are we dogmatists or conservatives about perception, for example?).<sup>15</sup> If they 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 than we seem capable of?

A second type of consideration against the generality of 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:

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' (*op. cit.*, p. 16).

How should we understand this passage? On one way of reading it, Kripke would be assuming that the contents of mental states are

<sup>15</sup> For the distinction between dogmatists and conservatives about perception, see James Pryor, "The Sceptic and the Dogmatist," *Noûs*, xxxiv (2000): 517–49.

derived from the contents of public language linguistic expressions. But if that is 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 thought and language, but rather that thoughts themselves involve the tokenings of expressions (of mentalese) and 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 us call it Kripke's *Meaning Assumption* and let us 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 on, ad infinitum. If we combine the Meaning Assumption with the Intention View, neither mental content nor rule-following would be able to get off the ground and rather obviously so.

Since Kripke regards the Meaning Assumption as nonoptional, 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.

*Some Solutions to These Problems for the Intention View.* If the second of the three objections to the Intention View that we have outlined, the one based on the relative opacity of our rules, is correct, then there must be a species of rule-acceptance that is nonintentional. And, if either the first or the third of our three considerations is correct—that is either the one based on reductive Naturalism or the one based on the Meaning Assumption—then not only must there be a species of rule-acceptance that is nonintentional, *all* rule-acceptance must at bottom *be* nonintentional, because even *intentional* forms of rule-acceptance will presuppose the *nonintentional* kind.

Now, since we know that it is going to be extremely difficult to pull off a nonintentional, dispositional account of rule-acceptance, we

should ask whether there is any way around these considerations. How strong are they? Can they be answered?

To the first objection, one might respond by saying that reductive Naturalism is not obviously correct and so can hardly be used to constrain the acceptability of an otherwise intuitively compelling account of rule-following. After all, it continues to prove difficult to account for other important phenomena, such as consciousness, within a reductive naturalistic setting.

To the second objection one could try responding by appealing to the notion of a *tacit* intention, an intention to do something that is not explicitly articulated in someone's consciousness but which he could be said to have implicitly or tacitly. The idea would be that the mental states by which rules are accepted or internalized are tacit intentions, rather than the sorts of explicit intention with which we are familiar in ordinary action.

Specifying such a notion in a satisfactory way has defied many serious attempts. But it is not clearly hopeless. And if we could explain what it is for someone to have an intention to do something in a way that is not explicitly articulated in some conscious state of his, that might then be used to explain why we do not have the sort of super sharp access to our rules as we do to our ordinary intentions.

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 were left in place.

The problem, of course, is that even unreduced, tacit intentions are contentful states. As a result, it would still not be possible to combine the Intention View with the Meaning Assumption. But can the Meaning Assumption be plausibly rejected?

Let us 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 does not follow from these obvious truths that the way the word 'tiger' comes to mean what it does for a given speaker *S*—that

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!<sup>16</sup>

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.<sup>17</sup> 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 rule-like in its generality.

What about the expressions of our language of thought? Is it similarly compelling to say that they get their meaning by our following rules in respect of them?

Here things may look quite different, especially if we emphasize that we are dealing with a personal-level notion of rule-following according to which it is a person who follows a rule and not just his brain.

At a personal level it appears to make very little sense to say that we follow rules in respect of our mental expressions, expressions to which the ordinary person has no access and which, for all that such a person knows, may not even exist.

Kripke is clearly working with a person-level notion of rule-following. That is why he can confidently claim that when someone is following a rule that rule *justifies* (or as I would prefer to say, 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.

It looks, then, as though, at least as far as personal-level rule-following is concerned, we are free to reject Kripke’s Meaning Assumption, at

<sup>16</sup> Jerry Fodor may have been the first to appreciate this clearly; see his *A Theory of Content and Other Essays* (Cambridge: MIT, 1990), pp. 135–36. I do not believe that any of the main arguments of my “The Rule-Following Considerations,” *Mind*, xcvi (1989); 507–49, are affected by paying greater heed to this distinction, although I am sure I was not as clear about it in that paper as I should have been.

<sup>17</sup> I have gone back and forth about the plausibility of the Meaning Assumption as applied to public language expressions. In my New York University seminar of Spring 2006, I defended it, but in an earlier version of this paper 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.

least as it applies to mental expressions. And with that observation we seem to have answered the third of the three objections we had posed for the Intention View.<sup>18</sup>

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 arguments against *reductive* accounts of mental content facts, accounts such as those provided by dispositional or functionalist or informational theories. And I am inclined to think that these arguments, along with others that may be found in the literature, are very persuasive.<sup>19</sup>

Even if we concede all this, however, that still appears to leave anti-reductionist conceptions of mental content untouched. Kripke tries to undermine such conceptions of content as well, of course; but, as I have argued at length elsewhere, those arguments seem to me to be answerable (*ibid.*).

If we were to adopt such an anti-reductionist conception of mental content, would not that mean that we would now be free to adopt the Intention View of rule-following?

*The Real Problem for the Intention View.* Not quite. For what I now want to argue is that even if *all* of these responses were to pan out, that still would not suffice to salvage the Intention View. The Intention View suffers from a further and seemingly fatal flaw. It concerns not, as Kripke alleges, the Acceptance aspect of rule-following, but rather, the aspects that I earlier labeled Explanation and Rationalization.

To see what this problem is, let us waive Naturalism; let us ignore the examples of putatively nonintentional forms of rule-acceptance; 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, wouldn't 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 that the answer to this question is 'No'—even if we could simply help ourselves to the full use of

<sup>18</sup>For all that we have said, of course, it remains possible that we need to think of mental meaning as generated by sub-personal rule-following and that this will cause problems of its own. I shall come back to this question towards the end of the paper.

<sup>19</sup>For discussion and references, see my "The Rule Following Considerations." More on this below.

intentional contents, 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? Something along the following lines (with absolutely no aspiration to capturing Wittgenstein’s literary style):

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 how rule-following is possible.

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? 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 does not 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.

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, as we have already seen above, is a form 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 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 is not 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.<sup>20</sup>

This argument bears an obvious similarity to Lewis Carroll's famous argument in "What the Tortoise Said to Achilles."<sup>21</sup>

The Carrollian argument, however, is meant to raise a problem for the *justification* of our rules of inference—How can we justify our belief that Modus Ponens, for example, 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 we were talking about the rule Affirming the Consequent, the problem I am pointing to would still arise.

It would seem, then, that there would 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 the Rule-Following picture of inference are sufficient for generating a problem.

*Intentions and Intentional States.* How should we proceed? I have been talking about the Intention View, but, of course, everything I have 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.

<sup>20</sup> 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?"

<sup>21</sup> See L. Carroll, "What the Tortoise Said to Achilles," *Mind*, iv (1895): 278–80. There is also a similarity to Quine's arguments in his "Truth by Convention," in his *The Ways of Paradox and Other Essays* (Cambridge: Harvard, 1966), pp. 70–99.

The claim is that the following five propositions form an inconsistent set.

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

If my argument is correct, then one of these claims has to go.<sup>22</sup> 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. For how could, say, 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*? (I shall come back to this.) (5) seems analytic of the very idea of deductive inference (more on this below).

When we review our options, the only plausible nonskeptical option seems to be to give up (3), the Intentional View. To rescue the possibility of rule-following, it seems, we must find a way of *accepting* 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 *Philosophical 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 sense if it was understood symbolically. — I should have said: *This is how it strikes me.*

<sup>22</sup> 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 *sub-personal* 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 does not matter whether this is thought of as occurring at the personal or the sub-personal level—more on this below.

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.

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.

*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 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, noninferential.

Now, Kripke, as we know, gives an extended critique of the dispositional view. However, that critique has not generally been thought to be very effective; many writers have rejected it.<sup>23</sup> So perhaps there is hope for rule-following after all, in the form of a dispositional account.

My own view, by contrast with received opinion, 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 cannot offer us any refuge after all, if we abandon the Intentional View.

The core idea of a dispositional account is that what it is for someone to accept the rule Modus Ponens is, roughly, for him to be dis-

<sup>23</sup> See, for example, Scott Soames, "Skepticism about Meaning: Indeterminacy, Normativity and the Rule Following Paradox," *Canadian Journal of Philosophy*, Supp. Vol. xxiii (1998): 211–49; Paul Horwich, *Meaning* (New York: Oxford, 1998).

posed, 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 cannot simply read off his dispositions which rule is at work.

Second, the rule Modus Ponens is defined over an infinite number of pairs of propositions. However, a person’s dispositions are finite: it is not true that I have a disposition to answer  $q$  when asked 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) a thinker would not be capable of any performance errors and (b) he 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 is 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; so how come a human cannot? 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. If a mere glass can have infinitary dispositions, why could not a human being?<sup>24</sup>

There is a 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.

<sup>24</sup> See the discussion in “The Rule Following Considerations.”

These, then, are Kripke's central arguments against a dispositional account of rule-following, and although it would take much more elaboration to nail these arguments down, I believe that such an elaboration can be given.<sup>25</sup>

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 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 is 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.

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

<sup>25</sup> See my *Rules and Intentionality*, in preparation.

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.

*Is Going Sub-Personal the Solution?* I emphasized from the very beginning that the notion of rule-following that appears to underwrite the rule-following picture of rational belief is a personal-level notion. I reason about what to believe, not a part of my brain. As a result, it is the personal-level notion with which I have been most concerned in this paper.

Someone may therefore be tempted to think that perhaps the moral of the preceding discussion is precisely that it cannot be the personal-level notion that is at work in the rule-following picture, that the solution to the difficulties we have been outlining is to go sub-personal.

This suggestion resonates with what has been a robust tendency in the literature on rule-following. There are many discussions of the Intentional View that accuse it of being "overly intellectualized" and which recommend substituting a sub-personal notion in its place.<sup>26</sup> It is not very often made clear exactly what that is supposed to amount to. The preceding discussion should help us see that this is not a very useful suggestion.

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

<sup>26</sup> See, for example, Pollock and Cruz, chapter 5.

On the other hand, we could try identifying rule-internalization with some nonintentional state. Indeed, even if the state of rule internalization is initially identified with a sub-personal 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 sub-personally). And although we would no longer face the rationalization problem—because, presumably, sub-personal mechanisms are not called upon to rationalize their outputs—we would still face the enormous problems posed by the error and finitude objections.<sup>27</sup>

In consequence, I do not believe that going sub-personal offers a satisfying solution to the problems for the notion of rule-following that we have been describing.

### III. CONCLUSION

We think of our reasoning as governed by rules. We worry about whether our rules are the right ones, the ones that really deliver justified belief. We worry about how we might establish that they are the right ones; and about whether there can be a fact of the matter about that.

This entire way of looking at matters, though, depends on our being able to vindicate its fundamental assumption, that our reasoning is governed by rules.

If the preceding arguments are correct, there is a real problem about this.

First, it is hard to give a satisfactory answer to the question: What is a rule such that following it is necessary for rational belief?

Second, it is hard to explain how rule-following is so much as possible, and this difficulty arises even without our assuming either that rule-following or intentionality needs 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 Rule-Following picture of deductive inference, is not so much giv-

<sup>27</sup> As I say, I am unable to go into these objections in detail here—they are discussed at length in my *Rules and Intentionality*.

ing up on a Rule-Following *construal* of deductive inference as giving up on deductive inference itself.

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

The only other option with respect to our second problem (I do not at the moment know what to say about the first) 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.<sup>28</sup>

PAUL A. BOGHOSSIAN

New York University

<sup>28</sup> Tyler Burge urged this primitivist suggestion on me in conversation.