Wittgenstein, Kripke, and the Rule Following Paradox

Adam M. Croom
University of Pennsylvania

ABSTRACT: In §201 of *Philosophical Investigations*, Ludwig Wittgenstein puts forward his famous "rule-following paradox." The paradox is how can one follow in accord with a rule – the applications of which are potentially infinite – when the instances from which one learns the rule and the instances in which one displays that one has learned the rule are only finite? How can one be certain of rule-following at all? In *Wittgenstein: On Rules and Private Language*, Saul Kripke concedes the skeptical position that there are no *facts* that we follow a rule but that there still are conditions under which we are *warranted in asserting* of others that they are following a rule. In this paper, I explain why Kripke's solution to the rule-following paradox fails. I then offer an alternative.

Introducing the Rule-Following Paradox

At some point when I was a child, I learned how to add. Presumably, I was given examples such as "1 + 1 = 2," "2 + 1 = 3," "3 + 1 = 4" and so on. After being taught some finite number of examples, I was then able apply the rule "+ 1" to new addition-type-situations and obtain the correct results – the results that my teacher expected from me and the results that others obtained when they performed the same calculations. You could say that at some point I *grasped* the addition rule – the meaning of the '+' sign – and that my ability to solve other addition problems consists in the fact that I can follow in accord with this rule that I've grasped.

In general, children are taught to grasp the addition rule so that in future addition-type-situations, they will be able to answer any as-of-yet unperformed calculation correctly in virtue of following the addition rule that they have learned. We don't want children simply to *memorize* answers to the *previous* addition problems that they have already solved. We want them to grasp the addition rule so that they can apply that rule to solve new problems in *future* cases.

But since the child was only given a finite number of examples from which to learn the rule for addition, how do we know that the child will be able to apply the rule correctly to the infinite number of possible addition-type-problems he may come across? For instance, imagine that the child has thus far only solved addition problems involving numbers less than 57. Surely the child will be said to have produced correct solutions to these addition problems involving numbers smaller than 57 if the rule that he in fact learned was indeed the plus rule. But the child will *also* produce the correct solutions to addition problems in cases involving numbers smaller than 57 *even if* the rule he in fact learned was *not* plus but in fact *quus*, denoted by '$\oplus$' where $x \oplus y = x + y$, if $x, y < 57$, but $= 5$ otherwise (Kripke, 9). This is also the case even if the rule the child in fact learned was *not* plus *or* quus but in fact *guss*, denoted by '$\oplus^1$' where $x \oplus^1 y = x + y$, if $x, y < 57$, but $= 6$ otherwise. This is also the case even if the rule the child in fact learned was *not* plus *or* quus *or* guss but in fact *buss*, denoted by '$\oplus^2$' where $x \oplus^2 y = x + y$, if $x, y < 57$, but $= 7$ otherwise. Since, by hypothesis, the child has only previously solved addition problems involving numbers smaller than 57, it's clear from these examples that an *infinite* number of other rules compatible with the child's previous behavior can be generated. So how do we know that the child has not really learned some *other* rule that is *like* addition for calculations involving numbers less than or equal to 57 but differs from addition for calculations involving numbers equal to or greater than 57, which the child has not yet performed? Moreover, how can we

be sure that *anyone* has grasped this rule – or that there even are such fixed rules – given that rules are supposed to apply to infinitely many cases, whereas our learning and displaying of rules are necessarily delimited to a finite set of cases? Saul Kripke, in *Wittgenstein: On Rules and Private Language*, puts this skeptical point in the following way:

> The sceptic doubts whether any instructions I gave myself in the past compel (or justify) the answer '125' rather than '5' [to the addition problem '68 + 57']. He puts the challenge in terms of a sceptical hypothesis about a change in my usage. Perhaps when I used the term 'plus' in the *past*, I always meant quus [denoted by '⊕' where x ⊕ y = x + y, if x, y < 57, but = 5 otherwise]: by hypothesis I never gave myself any explicit directions that were incompatible with such a proposition. [13]

What is being called into question here is the criterion of "sameness" when continuing to follow a rule in the "same" way, and the problem is that, given any finite number of examples or instances displaying a rule, those finite instances are compatible with an infinite number of other possible rules. So the child may mean any number of functions by the '+' sign when he first learnt it. And similarly, he may *now* mean any number of functions by that same sign in new instances. But what follows from this is that there seems to be no fact about the rule for addition – what '+' means – that is to determine what counts as following in accordance with it, as continuing on in the same way.

Kripke's Skeptical Solution

According to Kripke, there are two possible ways to answer Wittgenstein's "skeptical paradox." The first way is to offer a "straight solution," which attempts to show that "on closer examination the skepticism proves to be unwarranted" (66). The second is to offer a "sceptical solution," which begins "by conceding that the sceptic's negative assertions are unanswerable" (66).

Kripke rejects the straight solution and turns to the skeptical solution because he thinks that since any instruction I've given myself for *plus* could be misinterpreted as one for *quus*, such instructions cannot provide us with facts about whether I meant *plus* rather than *quus* (13). And since any instruction I've given myself cannot constitute a fact as to what I meant, presumably there is no fact of the matter as to what I meant that we could use to refute the skeptic. Moreover, Kripke claims that this is Wittgenstein's position: "Wittgenstein's solution to his [sceptical] problem is a sceptical one. He does not give a 'straight' solution [...]. In fact, he agrees with his own hypothetical sceptic" (69).

However, Kripke thinks that we can still find a skeptical solution. Instead of looking for the factual conditions under which the instructions for following a rule are true, we are to look for the conditions under which we are warranted in asserting of an individual that she is following a rule. The shift of focus is from truth-conditions to assertability-conditions, a shift made from Wittgenstein's *Tractatus Logico-Philosophicus* to his *Philosophical Investigations*.

In the *Tractatus*, the meaning of a sentence is expressed by its truth-conditions (TLP §2.11, 2.18, 2.222), and a sentence is true just in case it corresponds to a fact that obtains in the world (TLP §2.222, 2.223). Wittgenstein's view of "correspondence-to-facts" in the *Tractatus* is expressed in terms of propositions picturing states of affairs. For a proposition to picture a state of affairs, the structure or "logical form" of the elements within the proposition (i.e., terms) must mirror the structure or "logical form" of the elements within the state of affairs (i.e., objects) such that the former depicts the latter (TLP §2.18, 2.2, 3.21). Thus the proposition "corresponds" to the fact.

In the *Investigations*, Wittgenstein shifts his focus to assertability-conditions (Kripke, 74). With his new focus on conditions of assertion, Wittgenstein is interested in how language is used and the circumstances under which assertions are made. Once we have made the move from truth-conditions to assertability-conditions, what we require of a person to be a rule-follower is that she meets the conditions under which we can *assert* of that person that she is following a rule, *not* the truth-conditional satisfaction of some instructions, mental or otherwise. In response to the skeptical question, "What constitutes my following the same rule at all?" the answer is not "Some fact." Rather, the skeptical answer Kripke suggests is "our community can assert of any individual that he follows a rule if he passes the tests for rule following applied to any member of the community" (110) and that "if an individual passes enough tests, the community [...] accepts him as a rule follower" (109).

Rejecting Kripke's Skeptical Solution

Is the skeptical solution that Kripke offers really a solution to the problem of rule-following *as such*, the problem as Wittgenstein raises it in §201? I am not convinced. For consider again the nature of Wittgenstein's problem:

> This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with a rule. The answer was: if everything can be made out to accord with the rule, then it can also be made out to conflict with it. And so there would be neither accord nor conflict here. [*PI* §201]

Notice that when Wittgenstein says, "*no* course of action could be determined by a rule, because *every* course of action can be made out to accord with a rule" (emphasis added), he does not qualify the skeptical problem to only *a particular kind* of acting in accord with a rule, such

as acting in accord with the rule for addition. Wittgenstein's inquiry is a *philosophical* one, not a particularly pedagogical one critiquing, for instance, the viability of mathematical education.[1] What Wittgenstein's skeptic is calling into question is rule-following *as such*. Therefore, an adequate answer to Wittgenstein's skeptic needs to address the *general* problem of rule-following.

Yet Kripke's solution does not seem to be a solution to the general problem of rule-following. According to Kripke, we now have assertability-conditions under which one is warranted in saying of someone that she can follow a rule, and those conditions are met by community test. But at this point Kripke faces two challenges:

Challenge 1: What makes it the case that what the community in fact does when it conducts a test in each new instance counts as testing in the *same* way? In the case of conducting tests – in particular, tests for ascribing rule-following to other individuals – what is the rule or criterion for "sameness" such that we can continue to conduct tests in the same way? The problem is that for a test to be reliable, we require that it be consistent, but consistency requires that we conduct our tests in the same way. However, this continuing-on-in-the-same-way is precisely what the skeptical paradox calls into question.

Challenge 2: By what criteria are people to be identified as members belonging to the *same* community, and what makes it the case that a community in fact *continues* to remain the same over time? If people are to be considered members of the same community – in particular, the community that is conducting tests for ascribing rule-following to other individuals and the community one is accepted into when one has sufficiently passed these tests – what is the rule or criterion for "sameness" in virtue of which this community is to count as the *same*? The problem is this: for members *a* and *b* to be members of the same community *c*, we require some feature (behavior, beliefs, etc.) of *a* and *b* to be

the *same*, since this feature is that which identifies them as members of the same *c*. Moreover, insofar as *a* and *b* are to *remain* members of the same *c*, they must *continue* to share this same feature. However, this continuing-on-in-the-same-way and continuing-to-share-the-same-feature is precisely what the skeptical paradox calls into question.

Yet surely if we cannot continue to test in the same way, and communities cannot continue on in the same way such that they remain the communities that they are, then the notions of *test* and *community* have slipped away along with our notion of following in accord with a rule. But it is within the framework of these notions of "test" and "community" that assertability-conditions function. So, unfortunately, since Kripke cannot answer Wittgenstein's skeptical paradox by means of assertability-conditions without presupposing rule-following at the outset, Kripke's skeptical "solution" is not an adequate solution to the skeptical paradox at all.

The Straight Solution: The Skeptic's Dilemma

I think that the rule-following skeptic – namely, Kripke – faces a dilemma and that this dilemma gives us our straight solution by dissolving the rule-following paradox. I formulate the skeptical dilemma as follows: if the statements constituting the skeptical argument are meaningful, then its conclusion cannot be true, and if the skeptical argument's conclusion is true, then its constituent statements cannot be meaningful. First, suppose that statements constituting Kripke's argument are meaningful and that we take them to be communicating something in particular to us: the apparent paradox of rule-following. However, the argument is meaningful in virtue of its communicative properties. These include that it consists of strings of terms combined in a well-formed way, extending for some temporal duration. Kripke's argument is book-length, but we'll focus

on just a simple sentence for the sake of simplicity. Still, notice that if the statements are meaningful, we thereby grant that the temporally distinct instances of terms out of which they are constituted retain sameness of use. That is, Kripke's skeptical argument is only meaningful because his use of the term "term" in its *first* occurring instance in the argument means the same thing as the term "term" in its *second* occurring instance of that same argument (this holds true for the other terms involved as well, such as "same," "past," "present," "doubt," "rules," "addition," and so on). For clearly you cannot present the temporally extended statement that "my use of a 'term' in the past does not conform to my use of a 'term' in the present" without presupposing at the outset that you mean the *same thing* by your terms in both *temporally distinct instances*.

Yet what the skeptical argument calls into question is precisely whether my use of a term in the *present* is the *same* as my use of the term in the *past* – that I continue to use my terms in the same way in accord with a rule. But if we grant that the constituent statements of the skeptical argument are meaningful, and so grant that the temporally distinct instances of terms retains sameness of use here, then we have already denied what the paradox was calling into question: that my use of a term might have changed from one instance to another. And if we grant that in *this* case we can follow in accord with a rule for using our terms in the same way, then we grant that we cannot doubt rule-following *as such*. Therefore, if the skeptical argument is meaningful, then the skeptical argument serves as a counterexample to the truth of the conclusion of the skeptical paradox.

Consider our second option. Let's assume that the conclusion of the skeptical paradox is true. As we have seen, the meaningfulness of the skeptical argument is due to the fact that it consists of strings of terms combined in a well-formed way, extending for some temporal duration, and that these terms retain sameness of

use throughout their temporally distinct occurrences. For instance, the statement that "my use of a 'term' in the past does not conform to my use of a 'term' in the present" presupposes for its meaningfulness that the term 'term' retains sameness of use throughout the statement. However, if the conclusion of the skeptical paradox *were true*, then temporally distinct instances of terms *could not* retain sameness of use. But if the temporally distinct instances of terms could not retain sameness of use, then the constituent terms in the skeptical argument could not be taken as retaining the same meaning. But if this were the case, then it would be impossible for the statement that "my use of a 'term' in the past does not conform to my use of a 'term' in the present" to present a problem of *sameness of terms*. There would be no stable term of which to say that *it* remains the same or not and no stable term that we could *use* to try to say this. Therefore, if the conclusion of the skeptical paradox were true, then you would not be able to understand the meaning of the argument at all. But since you do understand the meaning of the argument, then the skeptical conclusion must be false.

We see from this analysis that Kripke's rule-following skepticism is untenable and that there is no justification for supposing, along with Kripke, that "there can be no such thing as meaning anything by a word [... since] any present intention could be interpreted so as to accord with anything we choose to do" (55). Kripke's skeptic is guilty of putting the cart before the horse, because we must *first* learn the rules for the appropriate use of words such as "term," "doubt," "rules," "addition," etc., and then only *after* we have learned how to use these words can we *then use these words* to "doubt" the "rules" of "addition" in the way that Kripke has done. And surely this is the point that Wittgenstein is advocating, contra Kripke, when he claims that "if you tried to doubt everything you would not get as far as doubting anything. The game of

doubting itself presupposes certainty" (*OC* §115). This is because "the child learns by believing the adult. Doubt comes *after* belief" (*OC* §160).

The Limits of Paradox and What We Can Doubt

I think that Kripke misunderstands Wittgenstein in §201 of the *Investigations* partly because Kripke's view of how to account for what we call "following in accord with a rule" differs fundamentally from Wittgenstein's. When Kripke claims that "the important problem for Wittgenstein is that my present mental state does not appear to determine what I *ought* to do in the future" (56), Kripke is framing Wittgenstein's rule-following paradox as a problem concerning how we can infer that our present states legitimately determine our future ones. That is, Kripke is posing, and correlatively attempting to solve, the rule-following problem in terms of *induction*. However, this is clearly not Wittgenstein's approach. For as Wittgenstein claims, "The squirrel does not infer by induction that it is going to need stores next winter as well. And no more do we need a law of induction to justify our actions or our predictions" (*OC* §286). So Wittgenstein does not pose, and correlatively does not attempt to *solve*, the rule-following problem in terms of induction. Rather, it is more appropriate to see Wittgenstein as *dissolving* the rule-following problem by way of an analysis concerning a kind of *presupposition*, or more precisely, "an ungrounded way of acting" (*OC* §110).

According to Wittgenstein, our "certainty" in *r* does not consist in *r*'s being a sort of "guaranteed fact" (*OC* §11). Rather, certainty can be understood in terms of what is unintelligible to doubt, what "there cannot be any doubt about" (*OC* §219). In §215 of *On Certainty*, Wittgenstein considers a case where one questions whether she in fact has hands or not. Can the fact that you have hands really be doubted? Wittgenstein's point is

that it is unintelligible to doubt this in normal cases. For if you don't know something as primitive as whether or not you have hands, how could you find out? And how could you be sure of your method for finding out? Could you justify your belief that you have hands by simply looking at them? Surely not, for if you can't even be sure of your belief that you have hands, then what basis do you have to believe that your *eyes* are valid instruments for checking your *hands*? Why not use your *hands* as the basis to check whether or not your *eyes* are deceiving instead? The point is that at root we must take for granted *some* standard with which to distinguish truth from falsity – there must be some "inherited background against which I distinguish between true and false" (*OC* §94). Without this "inherited background," one could not even be *wrong* about any beliefs, since being wrong presupposes the possibility of a *correct* alternative. A man who could not even trust his senses has no basis for trusting anything, and *a fortiori* the legitimacy for holding a skeptical doubt in the first place. In such a case when a man cannot even trust his senses, "we should regard him as demented," (*OC* §155) not wrong or incorrect.

The general validity of our senses is certain, not because we prove that they are correct but because they are the *basis* from which we prove correctness in other cases. We must assume as a matter of course that they are generally correct because they are the basis from which we construct the rest of our knowledge and the means by which we validate our knowledge. Even to prove that your senses have deceived you in a particular case, you must rely on their general correctness to provide you with evidential support of that particular error. It is in this sense that it is unintelligible to doubt them, and this justifies us in considering them certain.

Can Rule-Following Be Called into Question? Concluding Remarks

It is in this sense of "certainty" that we are to consider rule-following as unintelligible to doubt, for "the game of doubting itself presupposes certainty" (*OC* §115). With regard to rule-following, my response to Kripke is the following: you cannot offer a proof for rule-following because rule-following is what *enables* you to prove. You cannot doubt rule-following, because rule-following is what *enables* you to formulate, confirm, and disconfirm your doubts. What we call our "rule-following ability" is part of our nature – our "inherited background" – as bio-cultural entities and so cannot be intelligibly doubted. And it is in this sense that we are to consider rule-following as a certainty.

To conclude: there is no warrant in looking for a skeptical solution, because there is no skeptical "solution" to the rule-following problem to be found. Finding such a solution would already presuppose rule-following. Rather, since what we call our "rule-following ability" holds the place of a certainty in our lives, we see that "on closer examination the skepticism proves to be unwarranted." That is, the fact that rule-following is a certainty gives us our "straight solution." It is precisely because we cannot in fact doubt rule-following that community members can "continue on in the same way" to test individuals for acceptance as fellow "rule-followers" in their community. Because rule-following cannot be doubted, community membership is possible and those communities can continue to follow the same tests for ascribing "rule-following" to other individuals. Even the satisfaction of assertability-conditions and the framing of skeptical paradoxes requires that we are able to "continue on in the same way" – to follow a rule – at some more basic, primitive level.

"We do calculate according to a rule, and that is enough" (*OC* §46).

Note

[1]Wittgenstein does not question that we can actually learn mathematics: "We do calculate according to a rule, and that is enough" (*OC* §46).

Works Cited

Kripke, Saul. *Wittgenstein: On Rules and Private Language*. Cambridge: Harvard University Press, 1982.

Wittgenstein, Ludwig. *On Certainty*. Ed. G. E. M. Anscombe. New York: Harper Press, 1972.

———. *Philosophical Investigations*. Tr. G. E. M. Anscombe. Oxford: Blackwell Publishers, 1953.

———. *Tractatus Logico-Philosophicus*. Tr. D. F. Pears and B. F. McGuiness. New York: Routledge, 2001.