

Rule Following — A Pedestrian Approach

Masahiro Yamada^{*}

This is the peer reviewed version of the following article: “Rule Following—A Pedestrian Approach”, *Philosophy and Phenomenological Research*, 80 (2): 283–311, which has been published in final form at <https://doi.org/10.1111/j.1933-1592.2010.00336.x>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

1. Introduction

Whenever an agent acts, her action, A, conforms to rules that require her to do A if C where C is any condition that is true. Of course, this by itself is not interesting. Her actions will conform to infinitely many rules since C can be any condition that is true at the time of her A-ing. But sometimes a rule the agent conforms to plays an important role in the explanations of her actions. For instance, we might say that Oscar changes his password for his email account with his employer every three months because he follows the company rule to change the password if it has

^{*}masahiro.yamada@cgu.edu

been three months since the last time he changed his password. Lucinda shakes hands with people she meets because she obeys the social norm to shake hands when she meets someone. Rudi lets in the man in a black coat to his boss's office because he has been instructed to let the man in when he arrives. Donna puts some change into a jar because she is in the habit of dumping whatever change she has into the jar whenever she gets near it. There are, of course, important differences between these cases but there is also a common thread. All of these cases involve an agent whose behavior is guided by a rule. Let us say that an agent follows a rule when his behavior is guided by a rule in the ways typified by the examples just given. Under what conditions is an agent following a rule? In this paper I will be developing an account of necessary conditions for an agent's following a rule.

I am assuming that there is a common core in all these cases such that it is possible to give an account of rule-following which will be a common factor in all of them. My justification for this assumption will come in the form of an account that fits the bill. Some remarks about the scope and limitations of such an account are in order. Because the account is designed to cover cases of habit as in Donna's case above as well as cases of social norms as in Lucinda's case it will not be able to shed light on the question as to what, if anything, distinguishes a social norm from mere habits agents might have. The account is also not capable of distinguishing between someone intentionally trying to conform to a rule and someone conforming to a rule without any such trying as in the case of habits and rules we follow because of our up-binging.

However, my account is designed to shed light on an issue of fundamental importance in answering the questions in the above paragraph. The issue is: what is the difference between, on the one hand, an agent's conforming to the rule to

sweat when hot which is not a case of following a rule — it is not a case of obeying a social norm, nor is it a matter of being in the habit of sweating when hot nor a case of intentionally conforming to the rule to sweat when hot — and, on the other hand, the examples above where an agent’s conformity to a rule is explained by her following a rule? To paraphrase Wittgenstein, the question is: what is left when the fact that my arm goes up every five seconds is subtracted from the fact that I am following the rule to raise my arm every five seconds? Rule-following is a type of action and an understanding of this type of action will presumably play an important role in understanding, for instance, what it is for a rule to be a social norm: perhaps, for a rule to be a social norm, not only must agents follow the rule but also be disposed to hold certain attitudes toward the rule in question.¹ Similarly, to be intentionally conforming to a rule seems to be a case of rule-following where this stands in an appropriate relation to an intention to follow the rule. To fully understand such accounts, one must first of all understand what it is to follow a rule in the sense I have in mind.

But is an account of rule-following even possible? In a discussion of Wittgenstein’s arguments concerning rule-following and private language, Saul Kripke (1982) has argued that there can be no adequate account of rule-following. If Kripke is right, my project is doomed from the start. Let me first take a closer look at Kripke’s skeptical arguments.

2. Kripke and Dispositional Accounts of Rule-Following

The discussion by Kripke (1982) is framed within the theory of meaning. The argument proceeds by arguing that there can be no adequate account of rule-following

¹Stueber (2005) can be interpreted as offering such an account.

and that this spells trouble for a realist theory of meaning. The precise connection of the difficulties for an account of rule-following and a theory of meaning do not concern me here.² Here, I am interested in the alleged difficulties for understanding the phenomenon of rule-following for the difficulties he raises are independent of any possible connections to a theory of meaning.

Kripke spends most of his time criticizing a dispositional account of rule-following. On a dispositional account, the state of someone's following a rule is to be analyzed in terms of the subject's having certain dispositions. Kripke contends that no such view has any chance of succeeding and much of the subsequent literature responding to him are defenses of dispositional accounts. The sketch that I will be providing falls within this class of dispositional accounts.

Kripke's discussion is centered around the example of rules governing the word 'plus.' 'x plus y' denotes the sum of x and y. We can think of a different word 'quus.' 'x quus y' denotes the sum of x and y if x and y are both less than 57, otherwise it denotes 5. Clearly, following the rules for 'plus' is different from following the rules for 'quus.' Kripke's challenge is to explain what the difference consists in. If no adequate explanation is forthcoming, we must give up the idea that following the rules for 'plus' is different from following the rules for 'quus.' Since the points do not depend on the choice of these particular rules, it would follow that there is no such thing as rule-following. So the argument form is that of a *reductio ad*

²One possibility is that the assumption is that the only promising candidate for a realist theory of meaning is one that treats possession of concepts as cases of rule-following. Another possibility is that the crucial point for the theory of meaning is not so much whether or not there is a coherent account of rule-following but the specific point concerning the alleged difficulty of specifying correctness conditions that can cover infinitely many cases (e.g. Boghossian (1989, 517), Blackburn (1984, 281–2)). Perhaps, there are other ways of seeing a connection.

absurdum: suppose there is such a thing as following the rule R; if this supposition is true, there must be an account of what following the rule R consists in; but there cannot be such an account; thus, the supposition that there is such a thing as following the rule R must be false; since the argument does not depend on the choice of R, there is no such thing as rule-following. A rebuttal of this argument requires a rebuttal of the allegation that there cannot be an account of rule following. And in offering the rebuttal, we can safely go along with the skeptic in supposing that we can understand and sometimes also follow any given rule that we can formulate. What we must do is show that this supposition does not land us into the kinds of trouble that the skeptic thinks we will find ourselves in. Since I am going to provide a dispositional account of rule-following, let me focus on his reasons for thinking that there cannot be such a dispositional account.

Kripke's reasons for thinking that there cannot be a dispositional account of rule-following can be divided into three challenges that any adequate account must meet. First, a rule has infinitely many applications. That is, there are infinitely many circumstances for which the rule tells the agent to do something. For instance, the rule *turn right at the next deli!* tells you to do something when you are standing on the corner of 59th and Lexington Avenue, when you are standing on the corner of 58th and Lexington Avenue, when you are standing on the corner of 57th and Lexington, etc. An adequate account of rule-following must allow finite creatures like ourselves to follow rules even though they have infinitely many distinct applications.

Secondly, someone can be following a rule without in every instance conforming to it. For instance, someone following the rule to take a pill every six hours might take a pill every five hours and thus fail to conform to the rule she is fol-

lowing because her watch is broken and goes a little faster than it should. An adequate account must allow for such failure to conform to a rule one is following and – though this is not explicitly stated by Kripke – also be capable of explaining certain systematicities in the kinds of failure that happen. A simple-minded dispositional view is bound to fail for this reason.

Thirdly, when a subject follows a rule, the subject is *justified* in conforming to the rule. Put in slogan form, rule-following is normative. Dispositions cannot possibly generate normative requirement so no dispositional theory of rule-following has any chance of being adequate.

Let me briefly rebut these challenges. The last point is rather mysterious. If we are thinking of social norms such as the rules of etiquette and other rules that govern social interaction, we might plausibly think that these norms generate normative requirements: one ought to – in some sense other than the moral sense of ‘ought to’ – follow rules constituting social norms. Now, a social norm is not simply a habit that is shared by a large number of people; e.g. just because everyone is in the habit of brushing teeth in the evening does not mean that there is a social norm requiring one to brush teeth in the evening. So a theory of social norms would need to tell us how social norms differ from generally shared habits and how social norms acquire their normative force.³ But a theory of rule-following as I am understanding it in this paper is *not* a theory of social norms. Presumably, an account of social norms will make use of a theory of rule-following but there is no good reason to suppose that the normative force of social norms is to be derived

³The subject of the normativity of social norms forms an important topic in the philosophy of social science. See, for instance, the recent discussions by, among others, Stueber (2005); Turner (2007); Rouse (2007)

from the nature of rule-following alone.

To see this, suppose Hans, an SS officer, follows the rule to kill any Jews in sight. Is there any good sense in which he is *justified* in conforming to this rule? It seems to me that there isn't. The mere fact that someone is following a rule cannot entail that he or she is justified in conforming to it. What is true is that given Hans is following this rule he would count as having broken the rule if he does not shoot a Jew in sight — but, this is important, having broken the rule need not be in any sense a bad thing. If there is any sense in which Hans is failing to do something he ought to, that must be due to something more than the mere fact that he is following the rule.

Even though it may not be true that Hans ought to conform to the rule stated above, if he fails to conform to it that is noteworthy in a way that it would not be noteworthy if a rabbi fails to shoot under the same circumstances. Someone who is in the habit of checking that his alarm clock is working before going to sleep might fail to check one evening and that would be noteworthy. But if I fail to check my alarm clock at night, it would not be noteworthy as I am not in any such habit. Of course, failure to do what one is in the habit of doing usually is not something to be criticized as wrong or mistaken. However, it is a noteworthy *failure* to conform to a rule and it would not be noteworthy unless one was following the rule in question. In other words, following a rule has success conditions. Might this, the fact that rule-following has success conditions, perhaps be what is meant by the slogan that rule-following is normative?⁴ If it is, it does not pose a problem.

S successfully follows the rule R only if S conforms to the rule. Any adequate

⁴The discussions by Blackburn (1984, 281–2) and Boghossian (1989, 517), for instance, take this route of understanding the problem of the normativity of content.

account of rule-following must be capable of specifying success conditions but it is not clear why this should be thought to be a particular problem for a dispositional account. In fact, I am not sure that it is a problem for any account of rule-following. After all, S conforms to the rule expressed by ‘if C, do A!’ just in case S does A if C. Once we can specify the rule S is following, specifying the success condition is trivial.

Maybe it will be complained that I am missing the point. Gampel writes “It is arguably essential to any rule to be capable of guiding and justifying, by defining a difference between correct and incorrect action.” (1997, 227) I take this to be in agreement with what I say here. However, Gampel argues that this entails difficulties for dispositional accounts of rule-following on the grounds that no naturalistic states could be *essentially* justifying. Wright argues in a similar fashion when he says “A solution to the problem would [...] be to show how suitably circumscribes facts about how one does, will, or would use a particular expression actually constitute facts about how one ought to use it.” (2002, 110). The reasoning here is not convincing. We must be careful in distinguishing the rule itself from an agent’s state of following the rule. An agent’s following the rule *if C, do A!* is constituted by facts about what he does, will or would do, among other things, but the fact that the rule *if C, do A!* requires doing A when C for conformity is decidedly *not* constituted by what the agent does, will or would do. The latter point about what it takes to conform to a rule is a simple platitude. Missing this is akin to confusing what it takes to *believe* that there is an odd number of stars with what it takes for it to be *true* that there is an odd number of stars. The facts constituting the mental state of believing that there is an odd number of stars do not constitute facts about when it is true that there is an odd number of stars. It is true that there is an odd

number of stars just in case there is an odd number of stars. There is no mystery here.

Now the second challenge. The challenge is that a dispositional account cannot make sense of the possibility of error. Why would one think this? The motivation for this claim appears to be the thought that a dispositional account must take the relevant dispositions to be dispositions to conform. So if the relevant dispositions are the ones that a subject has in the actual world, then it follows that the subject cannot – in the actual course of the world – fail to conform to the rule he is following. The typical response to this is to attempt to specify some ‘normal’ conditions and then to say that the rule a subject S follows are the ones that S is disposed to conform to in those normal conditions. Insofar as the actual conditions deviate from normal conditions, such a theory can account for the possibility of error.⁵ But why does anyone think that a dispositional theory has to take the relevant dispositions – whether they be dispositions in the actual world, or normal worlds or whatever – as dispositions to conform to the rule? Maybe the thought is that otherwise we have to classify some dispositions as dispositions to make mistakes but this requires already knowing which rule the subject is following. If this is the thought, it is misguided. Consider the following proposal:

S follows the rule *if C, do A!* if and only if: if S were to have an experience as of its being the case that C, he would do A.

Of course, this proposal is inadequate but notice that this is a dispositional proposal and yet the disposition in question is not the disposition to conform: S will

⁵For views of this kind see: Fodor (1990); Forbes (1984); Hindriks (2004); Pettit (1990); Wright (2002). Of course, there are great differences between the views advocated by these individuals.

fail to conform when C is the case but he fails to have an experience as of C (assuming there is nothing else that moves S to do A in the particular circumstances). There is no reason why a dispositional theory must treat relevant dispositions as dispositions to conform. The motivation for the challenge by Kripke's skeptic rests on a mistake about what dispositional theories can do.

Finally, the first challenge: rules have infinitely many instances of application but we are finite creatures. At times, Kripke makes it sound as if the worry is that a finite creature like us cannot have infinitely many dispositions and that this is inconsistent with rule-following requiring the ability to apply the rule in indefinitely many circumstances.⁶ If this really is the worry, it is hard to take it seriously. A sugar cube has the disposition to dissolve in a glass of water in L.A, has the disposition to dissolve in a glass of water in N.Y., in Bucharest, Shanghai, Sidney, Johannesburg, Buenos Aires, etc. So a sugar cube has all the infinitely many dispositions needed to unfailingly conform to the rule *if you are thrown into a glass of water under normal conditions, dissolve!*⁷ Of course, a sugar cube only conforms to this rule without following it but the point is that there is no difficulty in thinking that a finite object can have infinitely many dispositions needed to conform to a given rule under indefinitely many circumstances.

There is, however, a related point that deserves more attention and I suspect is what Kripke is really driving at. As creatures of a fairly specific type with fairly specific physical features, there are limitations on the kinds of dispositions we can have. For instance, there is no useful sense in which we could acquire the

⁶E.g. "[A dispositional theory] ignores an obvious fact: not only my actual performance, but also the totality of my dispositions, is finite."(1982, 26).

⁷Normal conditions are picked in terms of conditions that prevail in moderate zones on the surface of the earth.

dispositions needed to conform systematically to the rule *if you are exposed to sound of n Hz, emit electro-magnetic waves of n times 10^{12} Hz from your fingertips!* It seems to me that for this rule the right thing to say is that *we cannot follow it at all*⁸ and not just that it is very easy to fail to conform to it. An account of rule-following which entails that we can follow any and all rules must be false.⁹ This might be far more threatening than it appears. For consider the rule to respond with the sum whenever you are presented with two numbers in a certain way. Can we not follow this rule? But if there really are restrictions on what kinds of rules we can follow that arise from the way we are built, how can we be so confident? In fact, there seem to be decisive considerations against the possibility of following this rule. Let us assume that the rule, when precisified, is something like: *when you are presented with standard Arabic numerals for numbers n and m written on a piece of paper where the two numerals flank a sign that looks like '+', write down the standard Arabic numeral for the sum of n and m on the same piece of paper!* So when you see '1+1' on a piece of paper, the rule tells you to write down '2', when you see '154+231' on a piece of paper, the rule tells you to write down '385', etc. In fact,

⁸'We' needs emphasizing since there might be creatures — like the X-men — who can follow this rule. Whether or not one can follow a rule depends on which dispositions one has and can acquire.

⁹If concept possession is a matter of following certain kinds of rules, this means that there will be limitations on what we can think. We are already familiar with certain limitation that arise from considerations motivating content externalism (e.g. Putnam (1975); Kripke (1972); Burge (1979)) and what I am suggesting is another. Just as an aside: Aristotle (*De Anima* III.4) argues that the intellectual part of the soul cannot have matter (and hence is pure form) for otherwise there would be limitations on what we can think. We, as good materialists, must run *modus tollens* and conclude that since the mind is enmattered, there are limitations on what we can think.

the rule can be thought of as encapsulating infinitely many distinct rules which can be obtained by substituting for n and m into the following schema:

if you are presented with standard Arabic numerals for n and m where these flank the sign '+', write down the standard Arabic numeral for the sum of n and m !

It is clear that once the numbers n and m get very large, we will fail to have dispositions to conform to the corresponding instance rule simply because the numerals get too big for us to read them: imagine the numerals printed on the Great Wall of China from end to end — 4000 miles long — in a 10pt. font. If we want to say that we can follow every instance of the above schema, there is a serious problem for a dispositional account. The fact is, there are limitations on what kinds of dispositions we have and the dispositions needed to follow the rule about writing down the numeral for the sum of two numbers might simply outrun the kinds of dispositions we have. Even if we look for idealized circumstances in which agents are to have the required dispositions, it is hard to see that an idealization that allows for radical changes in our constitution are legitimate: would it be legitimate to allow a situation in which I am a mutant of the X-men variety as the 'ideal' or 'normal' circumstances?¹⁰ What should we say?

The right thing to do here is to agree that we cannot follow all instances of the above schema — we can, however, follow some instances; viz. those where the numbers involved are relatively small. But would this not mean that we do not know how to add, or that there would be an indeterminacy as to whether we are engaged in addition, quaddition or whatever? This is the kind of worry that Kripke

¹⁰For a more detailed discussion of the difficulties involved, see, for instance, Kusch (2005).

and Wittgenstein press in their respective discussions of rule-following.

Here is one way to raise the worry that we must resist. One might think that whether or not an agent comprehends a rule is determined by whether she can follow that rule. One might further think that to comprehend the above schema is to be able to follow every single instance of the schema. If there are instances of it an agent cannot follow, then that has to mean that an agent does not comprehend the schema. But it is evident that we do comprehend the schema so it cannot be that we can follow only some instances of it. This is a faulty piece of reasoning. Consider again the rule *if you are exposed to sound of n Hz, emit electro-magnetic waves of n times 10^{12} Hz from your fingertips!* The fact that we cannot follow this rule does not show that we do not understand it as it is part of our ordinary understanding of rule following that there are rules that we can understand without being able to follow them. We understand full well what the rule demands and see that we cannot do it. So the mere fact that we cannot follow all instances of the above schema does not threaten our comprehension of it.

But can we be said to have mastered addition if we cannot follow every instance of the above schema? After all, if we cannot follow all the instance rules, then there will be pairs of numbers for which we cannot produce the sum and for pairs of small numbers, the failure to produce the sum of the two numbers does serve as evidence that addition has not been mastered yet. At the same time, we must also note that failure to produce the sum for large numbers is not immediately taken as evidence that addition has not been mastered yet. Why is this?

The genius of the standard Arabic notation we use is that we can generate the numeral corresponding to the sum of two numbers n and m in a recursive fashion (provided n and m are themselves given to us in the Arabic notation). We start

from the right-most digit and move leftwards one digit at a time until we run out of digits. My readers know how to do this. To know how to add using the Arabic notation is to have mastered the finitely many rules needed to do the digit-by-digit addition plus mastery of the rule to stop when the digits run out. We might not have the memory to keep track of the summands or the whole sum but this does not prevent us from correctly calculating the sum of two numbers.¹¹ This is why we can start calculating the sum of two numbers written using a tiny font in the Arabic notation from end to end on the Great Wall of China without having to first read the whole numerals — we could do it while being forced to peek through a straw seeing only one digit at a time so long as we have the memory for carrying and the ability to shift our focus by one position at a time. The fact that once the numerals get too long we cannot finish the procedure is no evidence that we have not mastered the procedure for generating the Arabic numeral for the sum of two numbers where these are themselves given in the Arabic notation.¹² Another thing to note is that our grasp of the concept of a sum cannot possibly depend on our mastery for the procedure for generating the Arabic numeral for the sum of two numbers. If it did, we would have to say that the Romans did not possess the concept of a sum because they did not even use the Arabic notation

¹¹Let us not deceive ourselves into thinking that mastery of the procedure requires the ability to do it in our heads for arbitrary numbers. No teacher in her right mind would insist that her math students learn to do addition in their heads for arbitrarily large numbers. It is enough that we can carry it out on paper.

¹²There is still the epistemological problem of how much evidence we need to be satisfied that someone has mastered all the finitely many rules needed for mastery of the procedure of addition using Arabic numerals. We could never *prove* that someone is following the *right* rules as opposed to some ‘bent’ ones but this should be no more threatening than the fact that we cannot prove that the sun will rise tomorrow.

and their system of numerals was not capable of expressing infinitely many natural numbers¹³ let alone irrational numbers, or zero for that matter.¹⁴ Since this is so, our inability to add, i.e. the inability to produce the numeral for the sum, when numbers get too big is no threat to our being able to mean the sum by the sign '+'. The meaning of '+' can be given by something like: if 'n' refers to the number x and 'm' refers to the number y, 'n+m' refers to the sum of x and y.¹⁵

The response to the skeptic that I have sketched here is partly concessive. It is indeed true that given a dispositional account of rule-following, we must say that there are many rules that we cannot follow. But this does not mean that we

¹³So we cannot even say that the Romans had procedures for addition operating on Roman numerals which are isomorphic to the procedures we use operating on Arabic numerals.

¹⁴Of course, the Roman system of numerals can be extended on an ad hoc basis. The distinctive feature of the Arabic notation is that the rules for generating the sequence of numerals large than 9 (i.e. 10 onwards) determine a unique numeral for every natural number. No such rules exist for the Roman system. We can also invent a system of numerals which is indistinguishable from the Roman system for relatively small numbers but different for large numbers so that it is in fact capable of generating numerals for all integers. But such a system is not the Roman system of numerals.

¹⁵'n' and 'm' need not be Arabic numerals. ' $\pi+e$ ' refers to the sum of π and e even though ' π ' and ' e ' are not Arabic numerals. Also notice that the talk of numbers being too big for us to grasp (e.g. (Kripke 1982, 26–7) that one often hears in discussion seems to confuse numbers with numerals. The decimal expansion of π is too long for us to grasp but there is no difficulty for us to think about the number π . π is the ratio between the area of a circle and the area of a square standing on the radius of the circle. Similarly, a very large number when written out in the standard way might require a numeral too big to take in but it can easily be referred to by a single letter 'm', like this: let m be the largest number that can be written on the Great Wall from end to end using the Courier 10pt. font in the standard Arabic notation. Here is at least one thing I know about m: m is not a prime number (since all its digits must be '9' so that the whole thing is divisible by the number 9)

cannot follow any rules at all. Thus, the reductio argument that Kripke's skeptic is attempting to run fails because it is *not* true that the supposition that we follow the rule R leads to insurmountable problems for *every* R. Moreover, we have reason to believe that the concession that we cannot follow every rule does not entail that we must also concede that we cannot follow the rules for addition or that we cannot grasp the concept of an arithmetical sum.

In any case, the important thing to keep in mind is that a dispositional account of rule-following will have the consequence that there are limitations on what kinds of rules a given subject can follow. Sometimes this will mean that seemingly unproblematic rules will turn out to be impossible to follow. However this should not be taken to encourage skepticism about rule-following in general. What I am suggesting here is that rule-following might turn out to be a rarer phenomenon than one might have thought, not that it does not exist at all.

I hope I have done enough to diffuse the strength of the skeptical challenges raised by Kripke. Of course, my points do not amount to a proof that a dispositional account of rule-following will work. Whether a workable dispositional account exists can only be found out by trying to build one. In the following I will attempt to start a sketch of a dispositional account which I believe gives us reason to be hopeful.

3. Clarifications

Let me begin with some clarifications that are needed to avoid confusion. I shall use 'if C, do A!' as the canonical expression of a rule. I shall be using underlined italics to refer to the rule expressed; i.e. *If C, do A!* is the rule expressed by the phrase 'if C, do A!'. 'C' refers to a condition that I will call the input condition, 'A'

refers to an action type that I will call output and the rule *If C, do A!* is such that conformity to it requires doing A if C obtains.¹⁶

In English, expressions of the form ‘If C, do A!’ are variously called rules, directions, instructions and the like. These terms are not interchangeable. For example, when I am given directions for going from Los Angeles to San Diego, I am hardly given a rule. A rule, in ordinary usage of the term ‘rule’ seems to be something that is to be interpreted as requiring to do A *whenever* C, not just today or some other time and circumstances given by the context. Moreover, calling something a rule often carries the connotation that one *ought* to conform to it. In this paper, I will be using the term ‘rule’ in such a way that it does *not* carry the connotation that one ought to conform to it. Also, the existence of a rule does *not* in any sense imply that someone is actually following it. Rules simply are the referents of expressions of the form ‘If C, do A!’ As for expressions of the form ‘if C, do A!’ that we would ordinarily do not call ‘rules’ because they are not meant as ‘whenever C, do A!’, I will still call them rules for we can specify C in such a way that in the actual world it can be satisfied only once.¹⁷

Ordinary language locutions like ‘S follows the rule R’ are ambiguous in an

¹⁶I will not bother with the ontological question whether or not there are entities of the kind rule that are referred to by expressions of the form ‘If C, do A!’ I am not sure what a doubt here really amounts to. The notation using underlined italics can be thought of as a device to clearly mark when an expression for a rule is used as opposed to mentioned (a remark like ‘the rule R needs interpretation’ mentions the expression for the rule R since it is linguistic items that need interpretation).

¹⁷On March 21, 2006, at 3pm, Helen says to John: ‘if George is around, bring him here!’ and she does not mean that whenever George is around, he is to be brought to her. We can understand her as saying: whenever it is March 21, 2006, 3pm, and George is around, bring him here! (or some such).

important way. In one sense, following a rule implies that the subject actually conforms to it. Hence, the puzzling nature of a statement like this: 'John follows the rule to brush his teeth at 11pm. As you can see, he always brushes his teeth at 10.30pm.' If John brushes his teeth at 10.30pm, how can he be said to be following the rule to brush his teeth at 11pm? On the other hand, we can also say: 'John is brushing his teeth at 10pm because DST just ended but he has forgotten to set his clock and he follows the rule to brush his teeth at 11pm.' Clearly, in this latter statement, saying that John follows the rule to brush his teeth at 11pm does not imply that he conforms to it. If it did, the statement should strike us as contradictory which it does not. In this case, John's following the rule to brush his teeth at 11pm is not understood as entailing his conforming to it, but is understood as being the kind of state that, when combined with other facts, is capable of explaining why he brushes his teeth at 10pm. From here on, I will be using the locution 'S *observes* the rule R' for this sense of 'S follows the rule R' that does not imply conformity to it. The locution 'S *follows* the rule R' is reserved for the conformity implying sense.

The fact that S *observes* the rule R can be used to explain certain facts about his behavior. For instance, the fact that S observes the rule to brush his teeth at 11pm can, in conjunction with other facts, explain why S brushes his teeth at 11pm. But we can also appeal to S's observing a rule, in conjunction with other facts, to explain why S fails to conform to the rule as we have seen in the previous paragraph. It is the notion of observing a rule that is of interest in this paper. It is the notion that we use in explanations of behavior and an analysis of rule *following* presumably will have observing of a rule as a component.

The explanatory role of observing a rule shows that it is different from the in-

tention to conform to a rule. To see this, suppose Jimmy intends to conform to the rule *if the President is in the White House, jump three times!* Jimmy has no access to information about the president's whereabouts (say he is the sole survivor of an airplane crash and is currently stranded on a desert island trying to pass his time) so flips a coin and decides that the President is in the White House and jumps three times. Even if the President is in fact in the White House and Jimmy has therefore managed to conform to the rule, it is still an accident that he managed to conform to it. The intention and the result of the coin toss can only explain why he jumped three times, but not why he jumped three times when the President was in the White House. When the explanandum is conformity to the rule and not simply the performance of the output action of the rule, the mere intention to conform to the rule cannot explain it.

So an intention to conform to R is not sufficient for observing R. Is it necessary? It also appears *unnecessary*. When we think of rule-observing, it is tempting to think of cases in which we are given explicit instructions on what to do and we try to conform to the instructions in a more or less conscientious fashion.¹⁸ In such a case, we must intend to conform to a given rule. When we let such cases guide our thoughts on rule-observing, the paradigmatic rule-observer is a well-meaning, culturally sensitive tourist in a foreign country who tries to conform to the local customs by constantly consulting a guide book. But a little reflection should make clear that it is a bad idea to treat the tourist as the paradigmatic rule-observer. After all, the tourist is trying to be like the locals and one thing the

¹⁸Apparently, this is what Boghossian has in mind when he writes "... the ordinary concept of following a rule ... is the concept of an *intentional* act: it involves the intentional attempt to bring one's behavior in line with the dictates of some grasped rule." (1989, 516–7, his emphasis)

locals are certainly not doing is consulting guide books. The locals are observing the rules of custom that are spelled out in the guide book (assuming the guide book is a decent one) but the way they do it is rather different from the tourist's way of doing it. You might, of course, think that what is going on is that the locals consult rule-books that are in their heads: locals are like the enthusiastic tourist who has memorized his guide book and is capable of recalling its contents in a flash — so fast he does not recognize that he did consult the memorized book. But what could motivate treating locals as a subspecies of tourists other than the preconceived notion that the tourist is the paradigmatic rule observer? If the locals are taken as the paradigmatic rule observers, it is not at all clear that rule-observing requires any intention to conform to the rules. In fact, it is not even clear that a mental representation of the rule is required — witness the enormous difficulties we typically have in spelling out the finer details of social norms even though we have no difficulties observing them.¹⁹

Suffice this for initial clarification of terminology and the target concept of this paper. Let me quickly summarize what I take to be adequacy constraints on a decent account of observing rules:

1. Observing a rule must allow for failure to conform to a rule.
2. Observing a rule must be capable of explaining at least some instances of conformity to the rule.

¹⁹I am not raising the question whether to treat much of human activity as cases of rule-observing is feasible as some philosophers have (e.g. Dreyfus (1972); Searle (2001)). How much of human activity can be construed as cases of rule-observing will depend on what it takes to be observing a rule. All I am claiming here is that rule-observing does not appear to involve looking up a rule-book.

3. Intention to conform to a rule is not sufficient for observing a rule.
4. Intention to conform to a rule is not necessary for observing a rule.

4. Sketch

Phil is baking bread for the first time in his life. He constantly consults a cookbook to ensure a decent result. One of the things the cookbook says is to put the dough in a warmer place if it does not rise properly. As it seems to Phil that the dough is not rising properly, he puts it in a warmer place. Other things being equal, it is true of him that he would not have put the dough in a warmer place unless he *believed* that the dough is not rising properly. And this is so because he is observing the rule *if the dough is not rising properly, put it in a warmer place!* We would not accept the explanation that Phil put the dough in a warmer place because he was observing this rule if his action had nothing to do with his belief that the dough is not rising properly: perhaps Phil is a fidgety person and it is this that made him move the dough; if so, he did not move the dough because he is observing the rule in the cookbook.

When is an appeal to Phil's observing the rule even a candidate explanation for his putting the dough in a warmer place? The considerations so far intimate that Phil has to be such that if he were to believe that the dough is not rising properly, the tokening of this belief would dispose him to put the dough in a warmer place. The belief would only dispose him to perform the relevant action because many kinds of things can intervene. Perhaps, he forms the beliefs and at just that moment the phone rings and Phil fails to put the dough in a warmer place. But this kind of failure to do what the rule says must be a case of an intervention. The requirement

that the belief that the dough is not rising properly disposes Phil to put the dough in a warmer place captures this: the belief disposes him to perform the action but something prevents the actualization of this disposition. If even this much were not true of Phil, we would not consider him as observing the rule to put the dough in a warmer place if it is not rising properly: how could he count as observing the rule if his realization that the input condition is met makes no difference as to whether or not he is going to move the dough? Let us generalize this to the following necessary condition:

S observes the rule *if C, do A!* only if: if S tokens the belief that C, this belief would dispose S to do A.

This is only a necessary condition. Given the requirement that S's observing the rule R must be capable of explaining S's conformity to the rule, we would need something like the reliability of the belief that C. For instance, if Phil above is not capable of telling when the dough is or is not rising properly, his managing to do what the cookbook tells him to do could hardly be explained by the fact that the tokening of the belief that the dough is not rising properly disposes him to put it in a warmer place: it would still strike us as pure luck. Should we add the reliability of S's belief that C as another necessary condition? It is tempting to do so but I want to propose a more generalized condition.

In order to be observing the rule *if C, do A!*, the agent must be such that her performance of A is a response to the obtaining of the input condition C. She need not be such that she unfailingly does A whenever C obtains — errors are possible — but there must be some kind of systematic connection between the obtaining of C and the performance of the output A. One way in which such a connection can exist is when the agent can reliably form beliefs about whether or not C obtains

and she performs A if she believes C obtains. But such a connection need not be mediated via a belief about C. Other states can make an agent perform an action and be such that it is correlated with the obtaining of C.

For example, consider the rule to raise one's hand if one wants to say something. I observe some such rule. In general, when I raise my hand because I want to say something, I do not first form the belief with the content that I want to say something. Rather, what usually happens is that my wanting to say something makes me raise my hand. Of course, sometimes I might form such a belief about my own mental states — e.g, I am not so sure if I want to say something so I scrutinize myself and come to the conclusion that I do want to say something and this belief then makes me raise my hand — but normally raising my hand because I observe the rule to raise my hand if I want to say something does not involve forming beliefs about my own mental states. Nevertheless, there is a systematic connection between my wanting to say something and my raising my hand.

More generally, what examples like this suggest is that one observes the rule *If C, do A!* only if there is a state which is such that if one is in that state, C obtains and if one is in that state, that state also disposes one to do A. Here, the conditionals are not the material conditional but are meant to capture a positive correlation. Roughly, the proposition expressed by 'if P, Q' is true just in case in most nearby possible worlds in which P is true, Q is also true. Now, in the case of the rule *if you want to say something, raise your hand!* the state in question might be the input condition itself — my wanting to say something is trivially such that if I am in this state I want to say something, and it is this state of wanting to say something which disposes me to raise my hand. In other cases, it might be the belief that the input

condition is met — as in the case of Phil above following the cooking instructions.²⁰ In yet other cases, it might be yet another kind of state — for instance, in the case of instructions given to move our bodies in certain ways (like dancing lessons), the state that is appropriately correlated with the input condition is probably some proprioceptive state and it is this state that disposes one to move in the right way.

So far we have two suggested necessary conditions:

S observes the rule *if C, do A!* only if

i) there is a state σ such that if S is in σ , C obtains²¹ and the tokening of σ disposes S to do A;

and

ii) if S tokens the belief that C, this belief would dispose S to do A.

While it is not my ambition here to provide necessary and sufficient conditions for rule-observing, it is clear that these conditions need more refining.

Consider the following case. Bassanio is asking to marry Portia but in order to win her over, he must choose correctly from a group of three boxes — one is gold, one silver and the third is leaden. The right one contains Portia's portrait. Bassanio is attempting to conform to the following rule: *if a box contains Portia's portrait, open it!* He comes to believe, based on a whim, that the leaden one contains the portrait, opens it and it turns out that that is indeed the right box to open. Bassanio managed to conform to the rule *if a box contains Portia's portrait, open it!* But can

²⁰The kind of reliability needed here is similar to the one that Sosa (1999) calls safety. A belief is safe only if: in nearby possible worlds in which S believes that p, p is true.

²¹This is not the material implication. It is an ordinary language conditional and its semantics is roughly: in most nearby possible worlds in which S is in σ , C is also true.

he be said to have observed this rule? Hardly. If he is observing this rule, then this fact must play certain explanatory roles. In particular, the fact that he is observing the rule *if a box contains Portia's portrait, open it!* must, other things being equal, explain why he managed to conform to it. But we would not explain the fact that he managed to open the right box by appealing to his 'following' the rule to open a box if it contains Portia's portrait. Rather, we would say that Bassanio got lucky.

Notice, though, that the conditions i) and ii) above for observing a rule are satisfied: If he were to believe that a box contains Portia's portrait, this belief disposes him to open it; and there is a state, e.g. that of believing he is in front of the leaden box, which is such that if Bassanio were in that state, the box in question contains Portia's portrait and this state disposes him to open the box. We need to tighten these conditions.

The reason why Bassanio is not observing the rule to open the box containing Portia's portrait is that his belief that the leaden one contains the portrait is true merely by accident. Because the belief is true by accident, Bassanio's disposition to open the leaden box upon forming this belief cannot be explained by appeal to the fact that his belief is true: he does not open the leaden box *because* it is the right one to open. Compare this to what we would say if Bassanio *knew* that the leaden box is the one which contains the portrait. In such a case, we would not hesitate to say that Bassanio opened the right one because he is observing ('following') the rule to open a box if it contains Portia's portrait. And this is so because it would be correct to say that Bassanio opens the leaden box because it is the one containing Portia's portrait. This suggests that the state σ in i) must be such that it disposes the agent to do *A* because σ is such that if the agent is in σ , *C* obtains.

Let me suggest one more refinement. Suppose Bassanio starts off by knowing

that the portrait is in the leaden box but that Portia finds out that this all-important information was somehow leaked to him and that she changes the location of the portrait to the silver box. As a result, Bassanio fails to conform to the rule *if a box contains Portia's portrait, open it!* But surely Bassanio in this case is 'following' the rule just in the same sense as in the case in which Portia does not change the location of the portrait; i.e. in this case, Bassanio is observing the rule unlike the initial case in which he came to believe that the portrait is in the leaden box on a mere whim. Bassanio in this case is observing the rule because the disposition to open the leaden box upon coming to believe that it is the leaden box was set up because at the time when the disposition came into being it was true that if Bassanio has that belief, the box is the right one.

So here is my refined and, for the current paper, final suggestion of necessary conditions for rule-observing:

S observes the rule *if C, do A!* only if:

i) there is a state σ such that

a) S's tokening of σ disposes S to do A

b) a) is true because when the disposition came into place it was the case that if S is in σ , C obtains.

and

ii) if S tokens the belief that C, this belief would dispose S to do A.

Let me note a few advantages of this proposal. First, what i) accomplishes is that it is not necessary to *believe* that the input condition C obtains in order to succeed

in doing what the rule requires one to do.²²

Secondly, it accommodates the point that if an agent succeeds in doing as the rule requires, the success should in many cases be explainable by the fact that he observes the rule (the explanation would go along the lines of ‘i is true of S and S was in σ because C; and that is why he succeeded in doing as the rule requires’).²³

Thirdly, the conditions do not require any explicit, implicit, tacit or whatever representation of the rule by the agent. The relevant dispositions can be had because of wiring in the brain, as it were.

Fourthly, the necessary conditions that I have stated point to a dispositional account of observing a rule but the conditions do not require any actual success in doing as the rule requires. Imagine Vaclav living in the more terrible days of Czechoslovakia. He observes the rule *if there is a government undercover agent nearby, beat the hell out of him!* If he were to believe that there is an undercover agent nearby, he would beat the hell out of him. So condition ii) is satisfied. Condition i) is also satisfied since the relevant state could just be the belief that there is an undercover agent nearby (Vaclav, like most others, forms beliefs based on evidence so that if he believes that there is an undercover agent nearby, chances are he is right; and we are so wired that our beliefs guide our actions because beliefs are reliable). But Vaclav might never actually beat up anyone even though he is in fact surrounded by agents all the time because they are undercover and Vaclav

²²Let me reiterate that σ can be a belief. It is probably possible to do as the rule *if three circle appear on the screen, hit the button!* requires without first forming the belief that there are three circles on the screen. The same is almost certainly not true for the rule *if seventeen circles appear on the screen, hit the button!*

²³Naturally, there can be accidental success even if one observes the rule. The point is that it must be possible for the agent to succeed in doing as the rule requires *because* he observes the rule.

never suspects that even many of his closest friends are spies.

My proposal does a fairly good job in meeting the demands I placed on an account of observing rules. Let me discuss a few more examples to solidify this point. Consider our use of language. There is a state that is correlated with an uttered sentence's having a plural subject which causes the agent to use a plural verb in the same sentence. This state causes the agent to use a plural verb because that state is correlated with the sentence's having a plural subject (the establishment of such connections is one of the things that happen in learning a language). Moreover, if the agent were to believe that the sentence has a plural subject, this will dispose her to use a plural verb (and this disposition may override the disposition generated by the state that is fairly well correlated with the grammatical subject's being singular; for instance, when one realizes that a certain noun like 'police' is plural despite appearances and sets out to correct one's own linguistic dispositions). So one observes the rules of grammar — even though one would have difficulties spelling out the rules.

One might, however, wonder how many of what we ordinarily call rules can be formulated in the form 'if C, do A!' That is, how wide is the scope of my discussion of observing rules? Is it perhaps too narrow to be able to handle people's obeying rules of etiquette as a type of rule-observing? For instance, an injunction like 'Be polite' is not naturally put in the form 'if C, be polite!' for some C.²⁴ Does this mean that my discussion will not be able to handle rules like 'be polite' or 'dress appropriately'? Not quite. To be polite is to have a highly complex set of dispositions. For instance, a polite person is disposed to say 'thank you!' if she notices that someone held the door open for her; a polite person is disposed to

²⁴I am grateful to an anonymous reviewer for raising this issue.

wait and listen when she is aware that someone is talking to her; a polite person is disposed to pay attention to other people's movements when walking down a crowded street; etc. We can explain these dispositions by thinking of these as due to the person's observing rules such as *if someone holds the door open for you, say 'thank you'!*, *if someone is talking to you, wait and listen until s/he has finished!*, *if you are on a crowded street, pay attention to others' movements!*, etc. If one observes such rules one will be disposed, among other things, to say thank you when one notices that someone held the door open, to wait and listen when one is aware that one is being talked to, etc. That is, the dispositions constitutive of politeness can be nicely captured as resulting from the fact that a polite person observes rules along the lines just mentioned. In telling someone to be polite, we are telling her to follow a very large number of rules of the form 'if C, do A!' Which rules these are may be indeterminate on the fringes just as it is probably indeterminate just which dispositions constitute politeness. But this should not detract from the fact that we can usefully tell someone to be polite just as we can meaningfully describe someone as being polite. *Mutatis mutandis* for similar cases like the injunction to dress appropriately or to be generous.

Now consider the rule *if the temperature is above 90F, sweat!* There is a state which is correlated with the temperature's being above 90F and which also makes one sweat because that state is correlated with the temperature (after all, sweating's function is to regulate temperature so whatever state causes one to sweat does so because such a state tracks facts relevant to temperature regulation). So condition i) is met. But we certainly do not observe the rule *if the temperature is above 90F, sweat!* The reason is that the following is *not* true: the belief that the temperature is above 90F disposes one to sweat. Condition ii) is not met and this

explains why we are not observing the rule to sweat when hot.

Lastly, I do not observe the rule *if someone is not an Aryan, then do not befriend that person!* since my belief that someone is not an Aryan has no tendency to make me refrain from befriending that person.

The necessary conditions stated above do a good job correctly indicating whether or not a given subject is observing a certain rule. Since they are only necessary conditions, there are bound to be cases for which they do not suffice to correctly classify whether we are dealing with a case of rule-observing or not. Those cases will provide the bases for further elaborations on the conditions. Let me briefly discuss one particular case that is of interest. Our emotional responses seem to satisfy the conditions I have provided for rule-observing. For instance, consider someone who has a phobia of spiders. Someone who has such a phobia seems to satisfy the necessary conditions for observing the rule *if there is a spider nearby, feel fear!*: he is such that a belief that there is a spider will dispose him to feel fear; and there is a state (e.g. the perceptual state as of seeing a spider) which is such that were he in that state, it would dispose him to feel fear and this disposition came into place because the state is well correlated with the actual presence of spiders. Should we say that arachnophobia is a case of rule-observing? On the one hand, one might point to the fact that it is very natural to treat the formation of certain mental states as cases of rule-observing. For instance, one might think of concept possession as a case of rule-observing where this means that possessing concepts is a matter of observing certain rules of judgment; e.g., to possess the concept of a conjunction is to observe the relevant introduction and elimination rules in making judgments involving the concept of conjunction. It is also fairly natural to think of reasoning, which is one way of forming beliefs, as a case of

rule-observing. One might then think that the formation of emotions should be assimilated to these cases and be treated as cases of rule-observing. On the other hand, we might think that it is not a case of rule-observing since we cannot form emotion at will any more than we can sweat at will and try to tighten the conditions so that the formation of emotions are not cases of rule-observing. But this route will put pressure on the idea that we can treat the formation of at least some mental states as cases of rule-observing since it is not clear that we can form any mental states at will. We certainly cannot form beliefs or desires at will. It seems to me that the issue rests on the further theoretical advantages or disadvantages of treating the formation of emotions as cases of rule-observing. For instance, we think of emotions as rationally evaluable. A phobia is a case of an irrational emotional reaction and perhaps the best account of how emotions can be considered rational treats the formation of emotions as cases of rule-observing. But without further insight into the nature of emotions and how an account of emotions will interact with accounts of other mental states and rule-observing, I do not think it is possible to settle the question whether the formation of emotions is a case of observing rules.

Let me finish the present discussion with some remarks on further applications of this approach.

5. Observing a Rule Directly

Rule-observing is a type of action. Actions can stand in constitutive relations that a 'by' locution often signals; e.g. Oswald killed Kennedy *by* shooting him. Oswald's shooting is not identical with his killing Kennedy since he could have shot Kennedy without also killing him. However, it is part of the sequence of events that

together constitute the assassination of Kennedy. Two cases of rule-observing can stand in this relation, too. For instance, in order to comply with the rule to change my email password every three months, I might set up a system of periodic alerts sent to me and observe the rule to change my email password whenever I receive an alert to do so. In this case, I observe the rule to change my password every three months *by* observing the rule to change my password when I receive an alert to do so. Now suppose that whenever I receive an alert, a log entry is created in the computer system for administrative purposes. So in observing the rule to change my password whenever I receive an alert, I also conform to the rule to change my password whenever a certain entry is created in the log. However, I clearly am not observing the rule to change my password whenever an entry is created in the log. After all, I might not have any incling that logs are taken.

Here we have a case in which someone observes a rule R_1 such that conforming to R_1 will also lead to conforming to R_2 but the agent only observes R_1 . So it is not necessary that given S observes R_1 , there is another rule R_2 such that S observes R_2 by observing R_1 . This is a virtue since it blocks a regress. We also want an account of rule-observing to block a regress in the other direction. It better not be the case that, necessarily, given S observes R_1 , there is another rule R_2 such that S observes R_1 by observing R_2 .²⁵ The necessary conditions of rule-observing strongly suggest

²⁵This kind of regress worry is discussed, among others, by Millikan (1990) and Wright (1989).

Wittgenstein also insists on the need to stop the regress: “It can be seen that there is a misunderstanding here from the mere fact that in the course of our argument we give one interpretation after another; as if each one contented us at least for a moment, until we thought of yet another standing behind it. What this shews is that there is a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call “obeying the rule” and “going against it” in actual cases.” (1958, §201) The problem Wittgenstein points out arises from

that such a regress can indeed be avoided. Given the rule *if C, do A!* and a condition D such that when C is true D is true, an agent observing *and* conforming to the rule will also be conforming to the rule *if D, do A!* However, it need not be the case that the belief that D obtains disposes the agent to do A. Nor need there be any state of the agent which is positively correlated with D and disposes the agent to do A because of this correlation.

Let us say that S observes a rule R *directly* just in case there is no other rule R₂ such that S observes R by observing R₂. Wedgwood (2002) argues that rules that we can observe directly must be of a specific kind. More specifically, he holds that rules we can observe directly must have ‘internal’ input conditions.²⁶ ‘Internal’ input conditions are things like non-factive mental states. Two things stand out about this proposal — one of which I believe is correct, the other false. The first is the commitment to the idea that belief formation and revision is a case of rule-observing. If one held that rule-observing required the intention to conform to the rule in question, this will probably strike one as a non-starter. It is highly dubious that we have any intentions to conform to all sorts of inference rules. However, the necessary conditions laid out so far do not require any representation of the rules in question and since, as pointed out earlier, we should avoid taking the culturally sensitive tourist to be the paradigmatic rule-observer, there is little reason to believe that a representation of the rule will turn up as a necessary condition for observing a rule. Moreover, it is clear that we do satisfy the necessary condi-

treating rule-following as being essentially a matter of interpreting a representation of the rule which is itself a type of rule-following.

²⁶His discussion is in terms of rule ‘following.’ It is not entirely clear whether he means the conformity entailing sense or the sense that I am capturing with the term ‘observing.’ Pollock (1986, chapter 5) comes to a similar conclusion.

tions above to count as observing all kinds of rules governing what to believe. For instance, take the rule *if there are three objects in front of you, believe that there is an odd number of objects in front of you!* My tokening the belief that there are three objects in front of me disposes me to believe that there is an odd number of objects there. There is a state which disposes me to believe that there is an odd number of objects and which is also such that if I am in that state, there is an odd number of objects in front of me; e.g. the belief that there are three, or perhaps even the perceptual state as of three objects. And the reason that such a state disposes me to believe that there are three objects in front of me is that this state is reliable (presumably, the systematic nature of our belief formation and revision processes has a lot to do with the reliability of many of the crucial states involved). There does not seem to be an in principle difficulty for maintaining that belief formation/revision is to be understood as a species of rule-observing.²⁷

This brings me to the second point. According to Wedgwood, any rule we can observe directly must have internal states as input conditions. So while he can agree that we observe the rule *if there are three objects in front of you, believe that there is an odd number of objects in front of you!*, he will have to say that we do not observe this rule *directly* since the input condition is not an internal state — it concerns how it is with the external world. Why does he think this? Wedgwood's strategy is to look at the folk psychological explanation of what is taking place. He points out that on the intuitive, folk-level of explanation, the causal sequence of events in the case of someone who conforms to the above rule is something like:

1. three objects in front

²⁷Of course, further necessary conditions might rule out belief formation as a case of rule-observing. I am open to this possibility.

causes

2. visual experience as of three objects

causes

3. belief that there are three objects

causes

4. belief that there is an odd number of objects

While 1. does cause 3., it is not the proximate cause so that, according to Wedgwood, it cannot be said that the agent observes directly the rule *if there are three objects in front of you, believe that there are three objects!* But 2. is, at the folk-level of causal explanation, the proximate cause of 3. so that, according to Wedgwood, the agent observes *directly* the rule *if you have a visual experience as of three objects, believe that there are three objects!*

I do not wish to contest his account of the correct causal explanation at the folk-psychological level. But I do want to point out that the fact that an agent's being in state A is a cause, proximate or otherwise, of the agent's being in state B does not show that the agent observes, directly or otherwise, the rule to get into state B when in state A. Otherwise, we would have to say that we follow rules like *if you are nervous, become error-prone!*, *if you are hungry, become grumpy!* and, perhaps, *if you are in love, become blind!*²⁸ Wedgwood's points about the causal explanations at the folk-psychological level are plainly inadequate to establish the thesis that the

²⁸Dion Scott-Kakures has pointed out to me that none of these have what we would normally call actions as output. However, we can easily think of rules like *if you are nervous during a job interview, say something stupid!* And the like. That is, we can simply replace 'become error-prone' or 'become blind' with actions typical of being in such states.

only rules we can follow directly are rules with internal input conditions. This by itself does not mean that Wedgwood's thesis is false — only that it is ill-supported by his arguments. However, there are reasons to think that the thesis is indeed false. Consider the rule *if you have a visual experience as of three objects, believe that there are three objects!* According to Wedgwood, we observe this rule. If he is right, it must be the case that the tokening of the belief that I have a visual experience as of three objects disposes me to believe that there are three objects. This strikes me as false. I simply would not reason 'I have a visual experience as of three objects, so there are three objects.' It does not take long to agree that this reasoning would be a *non sequitur* and part of the reason why it is so easy to agree on this seems to be due to the fact that we plainly do not observe the rule to believe that there are three objects when we have a visual experience as of three objects.²⁹

The right thing to say here seems to be that we observe directly the rule to believe that there are three objects if there are three objects in front of one. The visual experience as of three objects disposes us to believe that there are three objects and this is so because vision is reliable; this satisfies condition i) and explains why normally the visual experience as of three objects also produces the belief that there are three objects. Condition ii) is also met since it is trivially true that the belief that there are three objects disposes one to believe the same.

Notice that to say that someone observes the rule *if P, believe that P!* is not to

²⁹The realization that I am not following the rule *if you have a visual experience as of three objects, believe that there are three objects!* requires a second order awareness of how my beliefs are formed. However, this does not mean that the second order awareness is part of what it is to observe a rule. Rather, the second order awareness helps us in figuring out which rule we observe at the first order level and hence helps us figure out what it takes to conform to the rule we observe.

say that she is disposed to believe P when P. After all, the conditions i) and ii) do not entail that one is disposed to believe that P when P is true. All that is needed is that there be a state σ which disposes her to believe that P and which is also such that if the agent is in σ , P is true. It is not required that the agent is even likely to be in σ when P is true. This is why I can be observing the rule *if the Dow Jones Index is up, believe that the Dow Jones Index is up!* even though most of the time I pay no attention to the financial news and so fail to conform to the rule. I manage to observe the rule because there is a state, e.g. the state of reading in the New York Times that the Dow Jones Index is up, such that if I am in that state the index is up, and the state also disposes me to believe that the index is up.

While a claim about which rule we in fact observe is not a claim about what rule we ought to observe, it is plausible given the assumption that we are by and large rational in our beliefs that the rules of belief formation we in fact observe are by and large also the rules that we ought to observe. Now, one might worry that if we observe the rule *if P, believe that P!* then it must be the case that we ought to believe that P when P which is clearly too demanding. But it is not at all clear why the fact that we observe the rule *if P, believe that P!* should have such a consequence. The most that follows is that we ought to observe the rule *if P, believe that P!*. Since observing the rule *if P, believe that P!* does not entail being such that one believes that P when P, the normative requirement to observe this rule does not generate a requirement to believe that P when P. Rather, it generates the requirement to be such that one counts as observing the rule *if P, believe that P!*

We must also not get confused by the apparent triviality of a rule of the form *if P, believe that P!* For many instances of P, it will not be easy, or even possible to observe such rules. For instance, the rule *if the galaxy you are looking at contains*

an even number of stars, believe that the galaxy contains an even number of stars! is not something we can observe with the current level of technology. Which rules we can observe depends heavily on our make-up and the way in which we are hooked up to the world. We can think of our cognitive machinery as allowing us to observe directly rules of the form *if P, believe that P!* for certain ranges of P.

6. Conclusion

I want to conclude this paper by noting two features of the way the discussion was conducted. One of the difficulties of the literature on rule-following is that so much of it is geared towards issues in the philosophy of language and mind. However, the concept of rule-following is most at home in action theory since rule-following is a concept of a type of action. This paper has been an attempt at understanding rule-following in action-theoretic terms and for this reason I have stayed away from issues in the philosophy of language and mind as much as possible. This does not mean that I think that rule-following discussions have no place in the philosophy of language and mind. As my discussion should have indicated, I am sympathetic to the idea that belief formation is a matter of observing rules of judgment and I am also sympathetic to the idea that concept possession is to be understood in terms of observing rules. But in constructing a theory of rule-following that is useful to the philosophy of language, it seems to me advisable to avoid begging the question against those who doubt that belief formation can be treated as a form of belief formation. Hence my developing of the necessary conditions for observing rules by looking at cases like following cooking instructions.

The necessary conditions I spelled out are in overtly intentional terms. If a reduction to non-intentional terms is thought to be an adequacy constraint on an

acceptable account of rule-following, my attempts here fail miserably. However, it is not obvious why such an account must be in non-intentional terms. Proposed analyses in action theory of certain types of actions like *intentional* action are typically not in non-intentional terms – why should we insist on an analysis in non-intentional terms when it comes to rule-following? If our ultimate aim is to give an account of concept-possession in terms of rule-following, one might worry that a non-reductive account leads to a vicious circularity in our theory. But, surely, if an attempt to utilize an independently motivated account of rule-following leads to serious problems in the theory of concept-possession, the right thing to do is to abandon the attempt to understand concept-possession in terms of rule-following. It is no objection to an account of rule-following that it cannot be utilized by theorists of meaning. Moreover, it is not obvious that the ‘circle’ would be vicious. Instead, we could think of the concept of rule-following as one of many interdependent concepts including beliefs, judgment and actions.³⁰ Whether this is so will depend on what kinds of connections we can draw out between these concepts and this is what I have attempted to do in this paper.³¹

References

Blackburn, Simon. 1984. The individual strikes back. *Synthese* 58 (3): 281–301.

³⁰For other criticism of the demand for reduction that is implicit in Kripke, see McGinn (2002); McDowell (1984); Boghossian (1989).

³¹I owe special gratitude to Paul Boghossian, Peter A. Graham, Peter Kung, Anna-Sara Malmgren, Joshua Schechter and Declan Smithies for in-depth discussions of and feedback on several earlier versions of this paper. I have also benefited from help by my colleagues at the Claremont Colleges and an audience at the University of Michigan.

- Boghossian, Paul A. 1989. The rule-following considerations. *Mind* 98 (392): 507–549.
- Burge, Tyler. 1979. Individualism and the mental. *Midwest Studies in Philosophy* 4:73–121.
- Dreyfus, Hubert L. 1972. *What computers can't do; a critique of artificial reason*. New York: Harper & Row.
- Fodor, Jerry A. 1990. *A theory of content and other essays*. Cambridge, Mass: MIT Press.
- Forbes, Graeme. 1984. Scepticism and semantic knowledge. *Proceedings of the Aristotelian Society* 84:223–237.
- Gampel, E. H. 1997. The normativity of meaning. *Philosophical Studies* 86 (3): 221–242.
- Hindriks, Frank A. 2004. A modest solution to the problem of rule-following. *Philosophical Studies* 121 (1): 65–98.
- Kripke, Saul. 1972. *Naming and necessity*. Cambridge, MA: Cambridge University Press.
- Kripke, Saul A. 1982. *Wittgenstein on rules and private language: An elementary exposition*. Cambridge, Mass: Harvard University Press.
- Kusch, Martin. 2005. Fodor v. Kripke: Semantic dispositionalism, idealization and ceteris paribus clauses. *Analysis* 65 (286): 156–163.
- McDowell, John. 1984. Wittgenstein on following a rule. *Synthese* 58 (3): 325–363.

- McGinn, C. 2002. Wittgenstein, Kripke and Non-Reductionism about meaning. In *Rule-following and meaning*, eds. A. Miller and C. Wright. Montreal and Kingston: McGill Queen's University Press.
- Millikan, Ruth Garrett. 1990. Truth rules, hoverflies, and the Kripke-Wittgenstein paradox. *The Philosophical Review* 99 (3): 323–353.
- Pettit, Philip. 1990. The reality of rule-following. *Mind* 99 (393): 1–21.
- Pollock, John L. 1986. *Contemporary theories of knowledge*. Rowman & Littlefield texts in philosophy, Totowa, N.J: Rowman & Littlefield.
- Putnam, Hilary. 1975. The meaning of 'Meaning'. In *Mind, language and reality*. Cambridge: Cambridge University Press.
- Rouse, Joseph. 2007. Social practices and normativity. *Philosophy of the Social Sciences* 37 (1): 46–56.
- Searle, John. 2001. *Rationality in action*. Cambridge, MA: The MIT Press.
- Sosa, Ernest. 1999. How to defeat opposition to Moore. *Philosophical Perspectives* 13:141.
- Stueber, Karsten R. 2005. How to think about rules and rule following. *Philosophy of the Social Sciences* 35 (3): 307–323.
- Turner, Stephen P. 2007. Explaining normativity. *Philosophy of the Social Sciences* 37 (1): 57–73.
- Wedgwood, Ralph. 2002. Internalism explained. *Philosophy and Phenomenological Research* 65 (2): 349–369.

Wittgenstein, Ludwig. 1958. *Philosophical investigations*. Oxford: Blackwell.

Wright, Crispin. 1989. Critical notice. *Mind* 98 (390): 289–305.

———. 2002. (Anti-)sceptics simple and subtle: G.E. Moore and John McDowell*.
Philosophy and Phenomenological Research 65 (2): 330–348.