Abstract

What is it for an agent to follow a rule, rather than merely act in accordance with it? An intuitive and plausible answer to this question is that to follow a rule is to perform an intentional act such that S follows a rule R iff S intends to act in accordance with R and subsequently acts on that intention. This intentional account of rule-following faces essentially two problems, widely seen as fatal: (1) Kripkean sceptical arguments, originally derived from Wittgenstein, suggesting that the requirements of a rule outstrip the possible content of our intention, and (2) a regress argument due to Boghossian, namely that the intentional view requires the agent to represent the conditional content of the rule in such a way that an inferential step is needed for the agent to move from the antecedent to the consequent, which, given some plausible assumptions about inference, leads to a regress.

In this paper, I will defend the intentional view of rule-following, using a game-theoretic account of semantic content outlined in Berg (2022) and Berg (forthcoming). I will argue that by placing the agent in the basic constitutive practice of using the terms that figure in an expression of the rule, we can give an answer to (1). Likewise, by positing that the agents have an in-built mental mechanism which is such that it responds to the constitutive structure of the content being represented (see Quilty-Dunn and Mandelbaum 2018) and by placing the resulting movement of thought in a basic constitutive practice of inferring, we can avoid Boghossian's regress.

What Is It to Follow a Rule?

Ásgeir Berg

January 19, 2026

Since the publication of Kripke's (1982) interpretation of Wittgenstein's rule-following considerations, there has been much philosophical interest in the notion of rule-following. Most of this work has concerned the notion of *semantic content*—more specifically, the question of how semantic content is determined, given that any purported fact about a given speaker seems to underdetermine what they mean. However, Kripke's sceptical arguments against the possibility of semantic content are only *superficially* about rule-following as such—the question of what it is to follow a rule—since we can, at least in theory, have an account of semantic content without thereby having answered the question of what it is to follow a rule (as opposed to merely acting in accord with it). And so, even if we could give an account of semantic content that avoids the problems laid out by Kripke's sceptic, it does not follow that we have thereby given an account of what it is to follow a rule as well.

Yet in the wake of Kripke's book, a very intuitive picture of what it is to follow a rule has been widely rejected, or at least found highly problematic, even by those who take it to be obviously true: namely, that to follow a rule is to perform an intentional act such that S follows a rule R iff S intends to act in accordance with R and subsequently acts on that intention. I will refer to this as the *intentional view of rule-following* and in this paper, I offer a defense of this view by developing an account that follows these broad contours.

The structure of the paper is as follows. In §1, I will explain why we require an account of rule-following in the first place and why the intentional view is both intuitive and attractive. In §2, I will lay out the problems for the intentional view, focusing on familiar Kripkean arguments and a well-know regress argument due to Boghossian, as well as some other desiderata we'd require for a successful account of what it is to follow a rule. In §3, I will lay out my own account, making use of a game-theoretic account of semantic content laid out in Berg (2022) and Berg (forthcoming). I will argue that this account can both preserve the intuitive elements of the intentional view and avoid the problems laid out in §2.

1 The intuitive analysis of rule-following

Consider the following rule, taken from Boghossian (2008, p. 485):

(ER) Answer every email as soon as it arrives.

Under what circumstances should we say that S is following this rule? It is clear that not everyone does so. Equally clearly we can imagine that S does in fact answer every email immediately without thereby also imagining that S is *following* the rule—they may just so act, perhaps coincidentally. A similar case would be if we supposed that S drew cards with numbers on them, writing down each number in a sequence. If by some coincidence, S were to then produce the Fibonacci sequence, we should not say that S was thereby following the rule "Write down the Fibonacci sequence!"—as what they did was completely coincidental, even if it was in accordance with the sequence. Such cases suggest that it can't be right to say that by performing certain actions, one must thereby be following a rule that accords with those actions, since conformity with a rule can be merely accidental.

It follows that there is a distinction between following a rule and acting in accordance with a rule. But what does this distinction consist in? What is the right relation between

the purported rule-follower and the rule they follow? The account I will defend for the rest of the paper is roughly captured by the following claim:

(RF) To follow a rule *R* is to (i) act on one's intention to act in accordance with *R*, and (ii) for one's action to actually accord with *R*.

As is common in the literature, this definition treats "following a rule" as a success term and according to it, one cannot follow a rule without acting in accordance with it. This is not compulsory, since one could give a different treatment by making changes elsewhere. Not much will hinge on this, however, since the focus of this paper will be on part (i) of the definition: explaining what distinguishes *S* merely acting in accord with a rule from *S* genuinely following it.

This distinction is often made by reference to *S*'s reasons for action: *S* follows a rule *R* only if *R* is (at least one of) *S*'s reasons for doing what accords with *R*. On this view, the rule rationalises *S*'s action and we can thereby say that *S* follows the rule, rather than merely acting in accord with it.[1] I will not follow this strategy exactly, rather defending the view that if *S* intends to follow *R* and acts in accordance with *R*, then *S eo ipso* had a reason to so act—that *S* has the intention to follow *R just is* a reason for *S* to act in accordance with *R*. After all, it is hard to see what it is for *R* to be the reason for *S*'s acting in accordance with *R* other than *S* having an intention to so act.

My notion of a "rule" will be a rather weak one. I will not, for instance, be concerned with what it means for a rule to be in force for a person—such as in the case of games—or with what it means for a rule to be binding for a person, independently of their intention, as might or might not be the case with ethical principles. It follows that when I say that *S* should or must perform a certain action in a given circumstance, I will not mean thereby that *S* has some sort of duty to do so, but merely that only by performing that action will *S* count as having fulfilled their intention of following the rule. These issues, however, would not affect the overall shape of the account. After all, if *S* intends to follow a rule

1. See e.g. Hacker and Baker 2009, p. 136–7, Glock 1996, p. 325, Glüer and Wikforss 2009.

that is not in force, there is no reason to think that *S* is not following it, even if their action is in some sense silly (e.g., if they intend always to move their king two squares in a chess game or the like). And while ethical principles may plausibly bind everyone, there is still a question of explaining how they are followed, and there I believe we can give the same answer, with slight modifications.

It will also be a consequence of this account that semantic content or the representational content of intention cannot be explained in terms of our following rules with respect to that content. After all, as Boghossian notes, if rule-following requires intention and intention is itself a contentful state, as seems eminently plausible, then the constitution of that very content cannot be explained in terms of our following rules, on pain of a regress.[2] That is to say, if we think that our very thoughts, including intentions, token expressions in some way (perhaps of mentalese) and that those expressions themselves get their meaning in virtue of us following rules, it then follows that in order to have intentions (including intentions to follow rules) some expressions need to have meaning. That in turn requires that we follow rules with respect to those expressions. But if rule-following requires intentions, we would have to have prior intentions, namely in order to follow the rules that gave the intention its content, and so on. It can therefore not be, one might think, that rule-following involves tokenings of expressions by mental states, including intentions.

This tension will be resolved by an appeal to an externalist account of semantic content according to which the content of *S*'s intention is determined by the practice *S* is embedded in, not by *S*'s following any semantic rules *per se*. For how this is spelled out in case of *S* meaning one thing rather than another by their words, I refer to Berg (forthcoming).

Following a rule and acting in accordance with a rule Above, I introduced the distinction between rule-following and accord by means of simple examples of actions

2. Boghossian 2008, p. 487. See also Wright 2007.

that merely coincided with the prescription of some rule, suggesting that the rule wasn't really being followed in the requisite sense. This distinction is not merely intuitive, however. There are also more theoretical reasons to insist on it. For instance, any action can be fitted onto some rule, if we merely describe it in the right way. For example, if I've always answered every email right away until now, but tomorrow do not manage, we just need to identify some relevant feature of the occasion where I didn't do so and add that to the rule as an excepction, e.g.

(ER*) Answer every email as soon as it arrives, except at 10:31 a.m. on December 13, 2025.

It follows that if rule-following and acting in accordance with a rule amounted to the same thing, then one would always be following a rule, no matter how one acts.

Similar examples show that one can also always find *infinitely* many rules that fit any act. Just modify ER* with the following rule template, such that for any n, there is a rule, "After you first do not answer an email right away, do not do so for the next n emails". It is clear that if I do not answer my next email right away, my act would still be in accordance with any of those rules. And so, without the distinction between rule-following and accord, one would not only always be following a rule, but an infinity of rules.[3]

The problem of error Furthermore, without the distinction between rule-following and accord, there would be no such thing as making a mistake in the course of following a rule. This is because any purported mistake would then show that one was following a different rule after all—indeed an infinity of them. For example, suppose I am calculating a simple sum. If I were to write down, say, "67 + 58 = 5", we would want to say that I made a mistake. But if rule-following and acting in accordance with a rule is the same thing, we would have to admit that I was in fact following e.g. the rules of *quaddition* instead of the rules of *addition* and thus that I did not make a mistake. After all, what I

3. For a similar point, see Boghossian 2008, p. 480.

did accords with *quaddition* and not *addition*. I will call this, following Kripke's discussion of a similar problem for dispositionalist accounts of semantic content, the *problem of error*.

This problem hits at the core of what it is to follow a rule, since if I am genuinely following a rule R, what I do can be normatively evaluated with respect to R: Some actions accord with R and others do not, and if they do, what I did was right with respect to R and if they do not, what I did was wrong with respect to R. If rule-following and accord amounted to the same thing, I could therefore never make a mistake in following a rule, since the mere fact of accord would shows that I correctly followed whatever rule R^* it was that my act accorded with, and hence that my action should have been normatively evaluated with respect to R^* and not with respect to R after all.

Rule-following therefore seems to require the possibility of mistakes; if I follow a rule and do not act in accordance with it, that was a mistake. To follow a rule is therefore more than just according with some empirical regularity. This suggests that no merely causal process could, in itself, be said to follow a rule in the requisite sense (colloquial usage, of course, often conflates rule-following with mere accordance). A calculator, for example, encodes a set of rules in its circuits and hence, if any causal process could be said to follow rules, a mechanical calculation would be a prime candidate. Of course, a calculator can give the wrong output, and that is a mistake of sorts. Suppose for example that an inventor has made a complex calculating machine out of gears and levers. In the course of a calculation, friction causes some of the gears to heat and expand, leading to a false result. In the relevant sense, this would be like a mistake, and so, again in the relevant sense, we might be tempted to say that the machine is following a rule (and relative to that rule, malfunctioned).

But now suppose that the inventor designed the machine in such a way that the gears are supposed to overheat and expand when calculating this sum, and thus that the machine was working as they intended, and that the answer it gave was what its inventor

expected and wanted. Here we would have two identical causal processes leading to the same result. If the first was in error, we should therefore say, by parity of reasoning, that both were the result of a malfunction, but that seems hard to sustain in light of the fact that in one case, the inventor meant for the gears to overheat and meant for the machine to give the reply it gave. This suggests that what the wrong result of a causal process is can only be evaluated relative to some rule and which rule that is seems to be underdetermined by the process itself—since here we have the same process evaluated relative to two different rules. It follows that naturally occuring causal processes cannot be said to follow rules; only accord with them.

It is possible, however, to solve the problem of error in a different way and say that there is no distinction between rule-following and accord, and that I thus always follow an infinity of rules, but what we need is to explain why my action should only be normatively evaluated with respect to at most one rule R. In what follows, I will nevertheless prefer to speak of the distinction between rule-following and accord, even if the distinction can be given in a slightly different form. In either case, we're looking for some relation between the purported rule-follower and the rule that can mark out the relevant distinction. The rule-follower's intention would be a natural candidate in both cases, and so it does not matter in which form we present the necessary distinction, whether between rule-following and mere accord or between normative evaluation with respect to at most one rule or no such evaluation.

Intention as the missing link In all of these cases, what we are missing is some relation between the rule-follower and the rule. But what? Boghossian plausibly (but tentatively) suggests that what makes the required distinction between rule-following and rule-accord is S's *internalisation* or *acceptance* of the rule (Boghossian 2008, p. 482). But how should we understand these notions? Boghossian thinks of these as states that can explain or provide a reason for S's actions in terms of the rule, which is why he takes

them to be equivalent to *S* having a commitment "to uphold a certain pattern in his thought or behavior" (Boghossian 2008, p. 482). That seems right—that internalising or accepting a rule is to take it as in force or binding for oneself, to accept that one ought to follow it.

However, it seems clear that one can accept a rule, or acknowledge that it is in force, without thereby following it. Consider, for example, the castling rule in chess. Presumably, every chess player accepts this rule. Yet it does not follow that they will castle whenever the position on the board allows it; they may simply judge that castling is not advantageous in a given situation and choose not to do so. This is of course because the castling rule *allows* castling, but does not mandate it. Similarly, every player on the pitch in a Premier League game accepts that certain rules are in force—or so we may assume. Yet players still break the rules on occasion. This is not because they reject the rules or believe themselves to be playing a different game; rather, they knowingly violate the rules in the hope of getting away with it. It is thus possible to knowingly abstain from following a rule or to follow a different rule, even if one accepts that certain rules are in force.

Likewise, commitment to a rule does not appear to be a sufficient condition either. Suppose, for example, that a compulsive and religious smoker commits himself to stop smoking by swearing an oath on the Bible. Every so often, he feels the desire to smoke and, being addicted as he is, forms the intention to smoke on each such occasion. It is possible that everytime this happens he is prevented by circumstance from actually smoking: the shop is closed, he has no cigarettes to hand, or he is interrupted before he can light up. In such a case, the smoker is committed (by their oath) to not smoke, but his commitment plays no role in his abstinence. Hence, he acts in accordance with the rule, but since his commitment is not the cause of his actions, he cannot be said to follow the rule.[4]

One plausible starting point in trying to find the missing link between rule-following

4. Similar examples could of course be constructed with acceptance or acknowledgement in place of committment.

and accord is to reflect on the fact that rules seem, by their very nature, to have conditional content: a rule, it seems intuitive to say, tells us what is mandatory or permissible when some situation obtains. Boghossian's email rule, again, seems typical in this regard, as it tells us that if we get an email, we should answer it immediately (or something to that effect). Here, the antecedent specifies a circumstance ("getting an email") and the consequent what one should do if it obtains ("answer the email immediately").[5] Given this conditional content of rules, I will say that to follow the email rule is to have formed the *intention* that *if* one gets an email, *then* one will answer it right away and then act on that intention in the right circumstances.[6] At this point, this picture of rule-following seems almost truistic—and perhaps obvious. After all, what kind of rule would that be that does not have conditional content? Isn't that what a rule is, something that tells us what to do in specific circumstances?

And how can one follow a rule—in the requisite sense—if not by having an intention to do what the consequent specifies, in case the antecedent obtains? By positing *S*'s intention in following the rule, we seem to get the right kind of connection between *S* and the rule: We can explain why *S* followed the rule, and not merely acted in accord with it and we can solve the problem of error. *S*'s intention of following the rule would likewise explain what it is for *S* to have internalised or accepted the rule—after all, if *S* hasn't somehow grasped the rule nor accepted it, in what sense can *S* be said to have the intention to follow it?

There are, of course, plausible examples of rules that do not appear to have conditional content. Two such cases might be: "Always be happy" and "Do not steal".[7] I only have a space for a small remark, which is that such rules still have a distinction between following and accord; someone sitting alone in their living room reading a book presumably isn't thereby following the rule that one ought not to steal, just because they happen

5. Again, for reasons of space I am not taking a stand on whether rules are imperative or propositional. See Boghossian 2008 for discussion.

6. See Boghossian 2008 for a similar point.

7. These cases were suggested to me by Elmar Unnsteinsson and an anonymous reviewer, respectively.

not to be stealing at that time, and a person of a sunny disposition isn't thereby following the otherwise questionable advice to always be happy. I would suggest that we can fit such rules into the intentional account by seeing them as being followed in so far as the rule-follower intentionally refrains from doing what the rules prohibit when a situation arises where they find themselves tempted or otherwise prone to do so. For example, one might follow the rule "Do not steal" by reminding oneself of it when tempted to steal, and refraining from stealing for that reason. Similarly, one might follow the rule "Always be happy" by trying to cheer oneself up when feeling sad. Such cases, however, will not play a big role in what follows. On such a view, it might be said, their conditional content is the contrapositive.

2 Problems for the intentional view of rule-following

So far, I've argued that any account of rule-following must be able to explain the difference between rule-following and rule-accord and the best candidate for this role is the purported rule-follower's intention. This intuitive picture of rule-following faces a number of very difficult problems, however, which have led to its almost universal rejection, even among purported defenders of the view, like Boghossian and Wright. In this section, I will give an overview of these problems before moving on to my own account and how it can avoid them.

Kripkean sceptical problems The first set of problems I want to discuss relate to sceptical arguments of the sort raised by Kripke in his (in)famous discussion of Wittgenstein's rule-following considerations, which have to do with the infinitary character of rules. How can an intentional state carry with it a commitment to an infinite pattern—determining for infinitely many possible cases what is correct to do? If I intend to add, for example, it will be correct for me to write down "57 + 68 = 125" but wrong to write down "57 + 68 = 5". But how can my intention carry with it this implication, for any

given case? It is not as if my intention somehow represents or encodes every possible case it might apply to specifically; the content of the intention is general, not specific. Or more simply: I didn't think of this case when I formed my intention to add, and yet it applies to it. One could say, as Wright has done, that this is simply what an intention is, that it has content with such infinite character, and hence that there is no mystery as to how we can intend to follow rules that settle indefinitely many cases.

However, it seems plausible to think, given that S has a certain intention, that there would be some fact about S that constitutes the fact that S has that intention—for example by some configuration of S's brain or their environment. This is not a question of S being in the position to *know* if they have a given intention, but the question of what constitutes the fact that S has *this* intention, rather than another. So even if Wright is surely right that this is simply what intention is, that it has content with such infinite character, there nevertheless remains the question of what makes it the case that S intends one thing, rather than another. An adequate account of what it is to follow a rule must thus, given Kripke's underdetermination arguments, explain why S has one intention rather than another.[8]

Boghossian's regress argument There is an unrelated problem, however, potentially even more devastating: What I will call *Boghossian's regress argument*.[9] The starting point of this argument is the plausible assumption that to infer is to follow an inference rule. Boghossian himself is led to this view for much the same reasons as I gave above for the view that to follow a rule requires an intention: S might think that p and because of that think that q, but it does not follow that S has thereby inferred that q from p—or to put in more familiar terms, to act in accordance with an inference rule is not necessarily to have inferred. And so, just as in the case of rules in general, we need to be able to make

8. Primitivism about intention is of course a possible view. We could also say, as Wright suggests, that intention is judgement-dependent, that if S's best judgement is that they intend to φ, then they intend to φ. These are valid options, but there is not space to examine them further here.

9. See e.g. Boghossian 2008, 2012, 2014.

this distinction. If we assume that inference consists in applying inference rules, we have our solution: *S*'s 'movement of thought' from *p* to *q* was an inference because *S* applied an inference rule.

This gets us into trouble, however. Assuming that *S*'s intention to act in accordance with the rule is what makes the distinction between rule-following and accord, it is clear that in order to follow a given rule, a kind of "double success" is required (Wright 2001, p. 207): that *S* must be able to (a) correctly identify that the relevant circumstances, described by the antecedent of the rule, obtain and (b) be able to identify what action will be required or permitted, in light of that situation. Taking the email rule as an example, this suggests the following model for rule-following:

(Rule) If I receive an email, I should answer it immediately.

(Minor premise) I've received an email.

(Conclusion) Therefore, I should answer it immediately.

But how then does *S* move from grasping the rule and recognising that the minor premise obtains to the conclusion? Since *S* is required to grasp that the antecedent obtains and for that *reason* move to the consequent, the most obvious candidate is that *S* obtains the conclusion by *inferring* it from the premises—and if we think that to infer is to follow an inference rule, that would be by applying modus ponens to two premises, or something sufficiently analogous.

However natural and plausible this seems, the two assumptions that to follow a rule is an intentional act and that inference involves rule-following now lands us in a difficulty, since for *S* to act on this intention, *S* needs to recognise that the antecedent of the condition obtains, i.e. that the condition of the rule is satisfied, and then because of that, fulfil their intention. A regress is immediate: if following a rule is a question of recognising that the antecedent of a conditional is satisfied and then drawing a conclusion, one needs to apply an inference rule (e.g. modus ponens), and if inference is a matter of

following an inference rule, one needs to do so again, etc. ad. inf.

Options for an account of inference We might think therefore that the matter is simple: We just adopt some other theory of inference—one that does not involve the application of an inference rule or simply deny that this movement of thought is an inference after all. The latter course of action seems blocked from the outset as it seems hard to imagine, as Crispin Wright points out, that this step is not inference of some kind—after all, what kind of movement of thought is it when *S* goes from their intention to follow the rule and the recognition that the antecedent of the rule obtains to the action that the consequent mandates or allows? Isn't that just... inference?[10]

That leaves the former option, but the difficulty here, however, is that the problem we are trying to solve is to explain why *S* can properly be said to follow a rule and not merely act in accordance with it. So it doesn't really matter how we construe the step from recognising that the trigger condition of an accepted rule obtains to the taking of an action that the rule mandates or permits—we can call it inference or not—if that step is itself not normatively evaluable, we might worry that we haven't solved the problem, but merely hidden it deeper in the account. That is to say, how can we explain what the difference between rule-following and accordance with a rule is, if the very explanation relies on a step where a comparable distinction is papered over?

It might be plausibly suggested the regress shows that the movement from the premises to the conclusion is somehow, at least at some step in the regress, *subpersonal*—that the rule-follower cannot explicitly accept the rule and recognise that the minor premise obtains, but rather that some subpersonal state plays that role. However, as Boghossian points out, we seem to have two options here: either (1) that subpersonal state is one where the requirement of the rule are explicitly encoded or represented and that some subpersonal, inferential-like process acts on what the rule calls for under the circumstances or (2) it is not. In the first case, we don't seem to be much better off, since if the

10. Wright 2012, p. 386

requirements of the rule must be explicitly represented and acted on by a subpersonal process, then presumably the requirements of the inference must be likewise explicitly represented and acted on, and so on. This suggests, it might be noted, that the regress isn't really about the application of inference rules, but rather about the explicit representation of the requirements of the rule, and that of what is required for the inference. This puts a constraint on what kind of account of inference we can adopt.

What about (2) then, the option that there is some subpersonal state such that the requirements of the inference rule are not explicitly represented? Ultimately, I believe that some such account must be right, for the simple reason that there is no other option left over; it just cannot be, barring a regress, that at each step, the requirements of the rule have to be explicitly represented and acted upon.

But can such an account be construed as rule-following proper? An important benefit that the rule-following model of inference can bring us is that rules can explain how inference is normatively constrained: When we follow a rule, some things accord with the rule and other things do not. It follows that some things are correct—relative to the rule—and others are incorrect. And since we want to say that some inferences are correct and others incorrect, it is natural to say that this is because the inference accords with some inference rule, and how better to explain this than by positing that we actually do follow these rules when inferring?[11] If we suppose that a subpersonal process or state takes care of this for us, we might therefore worry that it would not be subject to normative evaluation.

This aspect of Boghossian's (and Wright's) discussion of inference is often side-stepped, wrongly in my view, as this feature of an account of inference is absolutely crucial, because if one does *not* think that inference is rule-governed, then one needs to somehow explain how inference is nevertheless normatively constrained—and that explanation cannot itself rely on rules. It should be noted, however, that Boghossian himself warns

11. Blackburn 1984 openly assumes, for example, that one can move from one to the other in all contexts.

against conflating rule-following and normative constraint (2008, p. 480). He points out that there are intuitive cases where we would not say that a person is following a rule, even if we can normatively assess their behaviour or evaluate it in terms of a rule and so, at least in some cases, an act can be assessed normatively without the rule playing a part in the agent's actions.

In the next section, I will go on to develop my own account of what it is to follow a rule by means of my own account of semantic content (see Berg 2022, Berg, forthcoming). According to this account, the meaning of a word is given by the practice of using the word (where 'practice' is given a precise definition). The idea I will develop for the rest of the paper is that by using this account of semantic content, we can do both at the same time, (i) give an account of inference that does not require rule-following, while nevertheless construing inference as being normatively assessable, and (ii) explain how an intentional state can carry with it indefinite requirements—clearing the way for the intentional view of rule-following.

3 Basic constitutive practices

What I've promised might seem a tall order: to provide an intentional account of rule-following, while at the same time avoid sceptical arguments of the Kripkean sort, aiming to show that intention cannot provide correctness conditions for the infinitely many steps of a rule, *and* avoid Boghossian's regress argument. Fortunately, I am not starting from scratch. In previous papers, I've argued that the semantic content of a given symbol 'F' is determined by the game-theoretic equilibrium of the basic constitutive practice of using 'F' and via this notion it is possible to give a community solution to the rule-following paradox that is both objective and settles indefinitely many cases of the use of a 'F', thereby defining its semantic content.[12]

12. I will try keep this section to a minimum, and instead refer the reader to Berg 2022 and Berg, forthcoming for a more detailed discussion and answers to objections, in particular concerning the objectivity of such a community account of semantic content.

The basic idea in this section is that if we assume that the meaning of the term 'F' is indeed given by the basic constitutive practice of using 'F' and that *S*'s mental states token 'F', in some way or another, then the practice of using 'F' will provide correctness conditions for the content of those mental states—i.e. determine what they are about. This will apply, in particular, to *S*'s *intentions* and will allow us to stipulate that in order to follow a rule, *S* must have an intention to do so that can settle any case. The account is thus an externalist one—whereby the content of *S*'s intention is determined by the context in which *S* is embedded.

I will then propose to solve the regress problem by positing that there is some kind of subpersonal mechanism in *S*'s mind such that whenever the antecedent of an inference rule that takes us from the antecedent of the rule being followed to its consequent (i.e. modus ponens) is represented by the system, the consequent will also be represented.[13] By positing that such a movement of thought can also be placed in a basic constitutive practice of inferring, a purported inference can be normatively evaluated. There can thus be an external criterion by which we can normatively assess inference, without the agent doing the inferring needing to represent the rule and its requirements to themselves—thereby avoiding Boghossian's regress. The resulting account is a deflationary account of inference that does not conceive of inference as involving the following of inference rules—rather, positing that to infer just is to have a movement of thought which is embedded in such a practice.

Putting it all together, the result is an account of rule-following where it is indeed possible that to follow a rule is to act on one's intention to follow the rule, involving an inference from the antecedent of the rule to the consequent—without regress. I must stress, however, that this is not a paper responding to underdetermination arguments about semantic content as such and that I am taking my own solution to the rule-following paradox for granted. The point of this paper is to use this solution to give an

13. And here, I will make use of work by Quilty-Dunn and Mandelbaum (2018).

account of what it is to follow a rule. There is therefore not space to consider every aspect of or objection to the account I will give in the next section. The underlying claim of the paper is therefore conditional, namely, that *if* my purported account of semantic content works, then we can *also* offer a solution to the problems that have plagued the intentional account of rule-following, and possibly rehabilitate it as a candidate answer to the question of what it is to follow a rule.

Defining basic constitutive practices

Searle's distinction between regulative rules and constitutive rules is well-known (Searle 1969, p. 33). According to this distinction, certain rules, like those of chess, constitute what it is to perform certain actions—there would not be, for example, such a thing as 'moving a pawn' if not for the rules of chess. In Rawls' terminology, the rules of chess are the 'stagesetting' that constitutes chess, and thus makes the moves in a particular game of chess possible.However, since I want to account for what it is to follow a rule, and this is the crucial feature of the account I will give, not all stagesetting must come in the form of rules, *contra* the orthodox view. Instead, I will appeal to the *game-theoretic structure* of what I call *basic constitutive practices* to provide the necessary stagesetting. Thus, it will be the structure of some basic constitutive practice P that defines what it is to take part in P. The core of the account is that semantic content is given by such basic constitutive practices. For example, to mean *addition* by the symbol '+' is to take part in the basic constitutive practice of using the symbol '+'. Correctness is then determined by the structure of the relevant practices.

The structure of basic constitutive practices is analysed in terms of iterated coordination games, using a game-theoretic framework due to Peter Vanderschraaf (Vanderschraaf 2018). We assume that agents taking part in the practice are engaged in a sequence of interactions (those are the coordination games) where each agent receives a positive pay-off iff coordination was successful. The exact nature of the pay-off is left open, but

for now I conceive of it as merely a formal device used to mark off an occasion of successful communication. We say that an interaction is in equilibrium if no agent would unilaterally change their move and consequently receive the same or higher pay-off. The equilibrium is thus a state in which nobody would (or should) do anything other than what they did, given what everyone else has done.

Our goal now is to define the correctness conditions for the use of a symbol 'F' by building up the basic constitutive practice of using it in terms of coordination games and the underlying idea is that we can use the structure defined by the sequence of coordination games as the stagesetting by which we can evaluate an agent's actions against, stipulating that if they lie on the equilibrium, then they count as an instance of the action that the basic constitutive practice is a practice of. In this way, basic constitutive practices can provide normative constraint for themselves without the need for further stagesetting.

The basic constitutive practice of adding I will use the basic constitutive practice of using the symbol '+' as an example, but the idea is meant to generalise. As a first step, consider what I will call the "simplified '+' game". The game is played by two players who are asked a question of the form 'What is 57 + 68?'. The players can give one of two answers, '5' and '125'. We will assume that the game is repeated indefinitely and that the pay-offs for the possible moves are described by the matrix in Figure 1. If the agents keep

	'5'	'125'
'5'	$1, 1$	$0, 0$
'125'	$0, 0$	$1, 1$

Figure 1: Simplified '+' game for two agents

playing this game, we can define a sequence of such games

$$(\Gamma_t) = \Gamma_1, \Gamma_2, \ldots$$

where each Γ_i is of the same form as the game described by Figure 1. I will call such a sequence of games a *supergame* and the index t a period. I will also say that the actions of each agent at a given period t is an *act profile* $\boldsymbol{s}_t$ of t. For example, the act profile $\boldsymbol{s}_1 = (\text{'5'}, \text{'125'})$ would describe a game where the first agent gave the answer '5' in the first period, while the second agent gave the answer '125'.

The eventual point of this definition is that each period will stand for different occasions of use for the symbol '+' when it occurs in this form. We've therefore defined a coordination game for each instance of the question $57 + 68$ being posed—when I calculate the sum on Christmas Day, 2050 at 12:00:01, for example, we will have defined a game for that occasion, and so on for every possible instance. Notice that for any period t in (Γ_t) , the act profiles $\boldsymbol{s}_t = (\text{'125'}, \text{'125'})$ or $\boldsymbol{s}_t = (\text{'5'}, \text{'5'})$ pick out an equilibrium for the game Γ_t. If we then select an act profile that is in equilibrium for each Γ_t in the supergame, such a selection is called an *equilibrium path* of (Γ_t). An equilibrium path is thus a set of act profiles, where one and only act profile is selected from each period.

For example, we can select the act profile where the agents give the answer '5' in the first game and '125' in every game after that, which would be given by the set $\{\boldsymbol{s}_1 = (\text{'5'}, \text{'5'}), \boldsymbol{s}_2 = (\text{'125'}, \text{'125'}), \boldsymbol{s}_3 = (\text{'125'}, \text{'125'}) \ldots\}$, or if the agents always give the answer '5', we have a selection given by the set $\{\boldsymbol{s}_1 = (\text{'5'}, \text{'5'}), \boldsymbol{s}_2 = (\text{'5'}, \text{'5'}), \ldots\}$. Another possible equilibrium path would be where the agents start by giving the answer '5' in the first game, and then alternate, such that $\{\boldsymbol{s}_1 = (\text{'5'}, \text{'5'}), \boldsymbol{s}_2 = (\text{'125'}, \text{'125'}), \boldsymbol{s}_3 = (\text{'5'}, \text{'5'}) \ldots\}$. Since we can always make a different selection at any t, there are in fact infinitely many equilibrium paths through (Γ_t).

The idea behind this definition is that we can now refer to all the different ways a strategy can be manifested in practice. For example, if both players follow the strategy of throwing a coin and picking their answer based on the outcome, there is no single equilibrium path that corresponds to this strategy system—which equilibrium is selected depends on how the coin lands each time. In fact, the set of equilibrium paths that

this strategy system might realise corresponds to the set of all equilibrium paths through (Γ_t). In this particular game, since there are only two choices facing the agents, a given equilibrium path corresponds to a binary tree. In general, it is helpful to imagine an equilibrium path as a tree structure.

Now suppose we generalise the simplified '+' game and allow indefinitely many agents that can give any reply to take part. We can then, by replacing the numbers in the question that the agents are asked by variables, define a *generalised* '+' game for any *n* and *m* in the question schema 'What is $n + m$?' and enumerate them as follows:

$$(\Gamma)_+ = (\Gamma_t)_1, (\Gamma_t)_2, \ldots$$

where each $(\Gamma_t)_n$ is itself a supergame as defined above and thus repeated indefinitely many times. $(\Gamma)_+$ is thus an infinite sequence of supergames. This all sounds very technical, but the main thrust is quite simple: We've defined a supergame for any sum of the form "n + m", and for any such sum, we've defined indefinitely many coordination games—one for any possible occasion of use of that sum. There is therefore a coordination game for any possible use of the symbol '+'.

This allows us to make a very important move. If we were to choose one possible equilibrium path from each supergame in $(\Gamma)_+$, we'd have a set of equilibrium paths corresponding to one possible interpretation of '+'. Suppose for example that for any game in $(\Gamma)_+$ such that $m < 67$ we would select an equilibrium path where every act profile is such that the equilibrium is given according to addition, but defined by $s_t = \{(\text{'5'}, \text{'5'}, \ldots)\}$ afterwards, our selection would correspond to Kripke's *quaddition* function. More simply, if we select an equilibrium path where every act profile's equilibrium corresponds to *addition*, our selection would correspond to *addition*. I will call such a selection a *second-order equilibrium path* through $(\Gamma)_+$. *A given second-order equilibrium path* thus corresponds to one possible interpretation of a term.

In any game-theoretic analysis, the agents taking part in the game follow some strategy. This presents a *prima facie* problem for any game-theoretic solution to the rule-following paradox that the above definitions are meant to solve: When the agents choose a strategy, the strategy cannot be such that the agents need to grasp its semantic content before making their selection, since otherwise, the account would be circular. For that reason, we assume that the agents who take part in $(\Gamma)_+$ are (a) such that they form dispositions to give some answer in the presence of the right stimulus through an ongoing process of linguistic training—by which I mean any stimulus from the environment that creates or shapes their linguistic dispositions—and that they are such that they respond to this training in similar ways, and (b) that they simply follow the strategy of giving the reply they are disposed to give in a given case.[14]

Given these assumptions, the agents need not grasp any rule or regularity before making their choice of action in a given game, and so, no circularity will be forthcoming. Furthermore, since we assume that the agents form a similar set of dispositions regarding the use of the symbol '+' after their training, the act profiles selected after a given interaction will be in equilibrium, and hence for any given supergame, a unique equilibrium will be selected, which in turn means that only one second-order equilibrium path through $(\Gamma)_+$ will be selected.

We can now appeal to the second-order equilibrium of $(\Gamma)_+$ as the required stage-setting, since a given use of the symbol '+' can be said to be correct for S iff it lies on the second-order equilibrium path of the basic constitutive practice of using the symbol '+'—and further, that practice can be said to be an addition practice, if the second-order equilibrium corresponds to what we, in the metalanguage, call *adding*. That is to say, what it is for an agent to add is for that agent to take part in the basic constitutive practice of adding and a correct addition is one that lies on the second-order equilibrium path of the practice. This rules out sceptical interpretations, because given the dispositions of all

14. See Berg, forthcoming for a further discussion of these points, especially regarding the agent's linguistic training.

the agents, only one practice is actual. My claim here is that by appealing to basic constitutive practices, we can explain how a particular action can fall under one concept, and not another: By being placed in the context of a practice whose structure provides the stagesetting for us to evaluate it against.

Consequently, the semantic content of 'F' as uttered by an agent in a basic constitutive practice of using 'F' depends on the dispositions of all of the agents in the practice, not merely that agent: *S*'s utterance of 'F' is correct if and only if it falls on the second-order equilibrium path of using 'F' in which *S* is embedded.[15] The account is thus an externalist one, not depending merely on facts about *S* alone—and so, even if the dispositions of the agents play a role, dispositions are not what determines meaning directly. Since the account is also a community solution, there would be no fact of the matter as to which of the two concepts, *addition* or *quaddition*, a community ought to use, other than practical considerations. It might be better to add than to quadd, but there is nothing preventing a community of quadders to exist—its members would simply mean something else by '+' than we do.[16]

4 The intentional account of rule-following

Now we are finally in shape to give an account of rule-following whereby *S*'s intention to follow the rule is what makes the distinction between *S* following a rule and *S* acting in accordance with it, as well as avoiding the problems laid out earlier in the paper, the (1) Kripkean sceptical problems about the content of an intention and (2) Boghossian's regress.

I will start with the first set of problems, which essentially divide into two: How can a finite state of *S* be such that it encodes an intention which settles infinitely many

15. The further step of explaining how *S* thereby *means F* by their utterence is discussed in Berg, forthcoming. For the rest of this paper, I only require that the correctness conditions for the use of 'F' are established.

16. For more discussion on this and similar points, see Berg, forthcoming.

cases of the rule ahead of time? And further, how can such a state itself have semantic content if we must assume that the state gets this content in virtue of *S* following rules with respect to it?

An externalist account of intentional content My proposal is that the content of *S*'s intentional state isn't merely determined by its own form or structure, but that its full content is determined by the basic constitutive practice in which *S* is embedded. By appealing to the externalist nature of the account, we can thus do for mental content what we did for semantic content by stipulating that the content of *S*'s mental states are not fully specified by what goes on in *S*'s mind, but by reference to the basic constitutive practice *S* is embedded in.

The first pass at such an account would thus be the thought that there is some fact about *S*, perhaps a state of their brain or mind, which tokens an expression of intention, but that the full content of that intentional state is determined by the practice. So, just as the meaning of 'F' when *S* utters a sentence in which 'F' figures is determined by the basic constitutive practice of using 'F' in which *S* is embedded, so is the meaning of 'F' as tokened by *S*'s *intention* determined—again, by the basic constitutive practice of using 'F' in which *S* is embedded.[17] It follows that if the agents in the practice can have dispositions of an infinitary character, then the practice can determine such a content as well.

I do not, however, want to make a firm stand here on what the right account of intention in general is, and we should allow for a more complex account of intention to be consistent with our account of rule-following—one, for example, where *S* has an intention to φ without there being a corresponding state of *S*. For example, to borrow an example from Boghossian, suppose that a tennis player has formed the intention to

17. Although I should again stress that I do not claim to have solved Kripke's sceptical problem in *this* paper and answered all objections. It is rather that I'm taking a solution that I've developed elsewhere for granted, using *it* to solve the problem of rule-following proper, i.e. what it is to follow a rule.

For further elaboration and discussion of objections to my account of semantic content, see again Berg 2022 and Berg, forthcoming.

always hit the ball with a backhand if it comes to her left and with a forehand if it comes on the right. In the beginning, the tennis player consciously (and perhaps a bit clumsily) thinks about where each ball comes and acts accordingly. As she gets better and practices more, her reaction becomes unconscious and automatic. We don't need to suppose, when her playing has reached this level, that there is *any* state of the tennis player which can be identified as an intentional state. Nevertheless, it would be odd to say that she no longer has the intention to follow this rule or that she does not act with intention—or indeed that she is not following this rule. There might be other cases of this sort, where the overall description of an action and its circumstances would lead us to say that S has an intention, without there being a corresponding state of S that can be identified as an intentional state of S.

It would be preferable if the account allowed for such cases, and consequently, I will say that 'intention' is a family-resemblance concept, where in some cases there is some state of S which encodes their intention and other cases where there is no such state, but that S can nevertheless be correctly described as having an intention. Here, the idea would be to say that agents, through their linguistic training, develop dispositions to assent or deny claims of the form 'S intended to φ' and that we can posit that there is a basic constitutive practice of using descriptions of this sort which determines their meaning—and hence determines correctness conditions of whether such descriptions are correctly applied or not.

In the case of the tennis player, there might be some feature of the case that makes us say that she has that intention, even if there is no corresponding intentional state. For example, we might be inclined to to assert that the tennis player has the intention to hit the ball with a backhand because originally there was an intentional state that tokened that particular intention and there is a proper causal link from that state of affairs to the current one. The idea here is not to convert the account into a causal account of intention, but to point out that there might be other features of the case that we are

sensitive to when we ascribe intentions to agents, other than just an intentional state. We can also imagine more complex cases.

Either way, the two main theoretical claims about intention being made is that (1) the full content of *S*'s intention is determined by the game-theoretic structure of the basic constitutive practices in which *S* is embedded—the fact that *S* intends to play chess, hit with a backhand, or do arithmetic, rather than play chess*, hit with a backhand* or do quarithmetic depends on *S*'s being embedded in certain basic constitutive practices and not others, and that is in turn determined by the dispositions of all the other agents in the practice, and (2) that *intention* is a family-resemblance concept, whose features are tied together by our ascriptions of intention to agents. This does not mean that our ascriptions make it the case that *S* has a certain intention, but rather that they determine the meaning of 'intention' in our language. These claims should be consistent with a wide range of more substantial theories of intention.

The problem of finitude Since a second-order equilibrium path is determined by the dispositions of all the agents in the practice and their game-theoretic structure, the question of whether an intention can settle infinitely many cases hinges on the question of whether *S* can have dispositions with an infinitary character. One might therefore object: If dispositions are supposed to determine meaning (even indirectly, by constituting a structure of a practice) how can they possibly settle infinitely many cases? As Kripke points out, a number might be too large for me to grasp or take too long to be read out before the heat death of the universe. In such cases, I wouldn't have any dispositions, and hence no meaning would be determined. And yet, the rules of arithmetic have infinitary content.

Some authors have thought that Kripke's discussion must be misguided: If a glass can have infinitary disposition to break when falling on the floor here, being struck there, and so on, for infinitely many ways, locations and times, why shouldn't human beings

have such dispositions?[18] There is something right about this. There is, after all, no particular reason to think that at an agent cannot have a disposition to add two very small numbers at any occasion, and hence have an infinitary disposition towards that particular addition. This doesn't really touch Kripke's point, however, which is rather about there being some cases where the requirements of having a disposition are too great for a finite agent in the first place: We cannot grasp or process numbers that are too large, for example, and thus cannot have dispositions regarding them, even if our disposition regarding a small number might have that kind of infinitary character.[19]

Fortunately, I think we can finesse the notion of a disposition in a way that makes these issues psychologically tractable. First of all, as we noted, there are cases where it does not seem impossible for our dispositions to settle infinitely many cases, if the inputs are simple enough. One such example would be the general disposition to say whether or not any given patch of colour is red. It does not seem implausible that a person might have such a disposition, even if nobody has the ability to sit through a near endless presentation of objects and saying of each one whether they are red or not. Thus, even if this is impossible, it doesn't follow that there is any difficulty in asserting that for *each* object, I have the disposition to give an answer. And if so, that is enough to determine a basic constitutive practice of using the word 'red' that settles infinitely many cases.[20]

Basic constitutive practices and compositionality So far, however, I've only explained how a basic constitutive practice can determine the content of a single term, like 'F', while the conditional content of a rule is a proposition. For example, if *S* intends to follow the email rule, what is tokened is something corresponding in mentalese to the expression "If I receive an email, then I will answer it immediately". The meaning of

18. Blackburn 1984 may have been the first to advance this point.

19. For that reason, Guardo 2022 refers to the problem as the "Cardinality Argument": The problem is not that we cannot have an infinite disposition, but that the dispositions we do have only cover a finite segment of the addition function.

20. The case of addition is more complex. For an extended discussion of how to solve the problem of finitude, see Berg 2022 and Berg, forthcoming. For a similar account, see Warren 2020. For general criticisms of Warren's account, see Guardo 2022 and Berg 2025.

this expression would have to be determined by the meaning of the sub-expressions that occur in it.

This is not entirely straight-forward, however, since we haven't built anything about compositionality into basic constitutive practices. What then is the semantic content of a full sentence? I propose to locate the source of compositionality in the training of the agents—and thus, it is not, as is common in the literature, that the meaning of a sentence is *metaphysically determined* compositionally by the terms which figure in it, but that the agents form dispositions towards the full sentence compositionally via their training. In other words, when the agents learn the meaning of propositions, their dispositions are formed in such a way by their training, in the general case, that their further dispositions depend compositionally on prior dispositions. That is, when the agents learn the meaning of particular terms or whole sentences, their dispositions with respect to those terms or the terms that figure in those sentences are thereby also being shaped.

For example, S might be embedded in the practices of using the terms 'kicked' and 'the ball' and consequently be disposed to assert the sentence "X kicked the ball" if and only if X kicked the ball. This might be because S is predisposed to combine them in this way—that there is something about S such that whenever S has these two sets of dispositions, they also have this further disposition, or it might be because S has learned that this is how these terms are combined in his linguistic environment, forming these dispositions after such training.[21] In short, S might form dispositions towards combinations of terms in just the same way S might form disposistions towards single terms.

So just as there is a basic constitutive practice of using a single term, there is also a corresponding basic constitutive practice of using a given sentence in which the term figures, and that practice is determined by the dispositions of the agents to assent or deny

21. Consequently, the account can also handle purported exceptions to compositionality, since S might lexicalise "X kicked the bucket" as "X died" as a result of learning it directly, and not composing it from its constitutent terms.

it on a given occasion, and those dispositions in turn are explained by the dispositions of the agents with respect to the terms which figure in the sentence. Thus, the basic constitutive practice of using a *sentence* determines truth-conditions for that sentence and compositionality enters into the picture at the level of *training* as we can assume that the agents, being finite beings, form their dispositions towards complex terms compositionally from simpler ones in most cases.

In this sense, basic constitutive practices can compose, and if we assume that the dispositions of the agents can cover indefinitely many cases, this account implies that S can have an intentional state which encodes the full content of a rule, provided that S is embedded in a basic constitutive practice of using the sentence which that state tokens.

Quilty-Dunn & Mandelbaum's account of inference

This brings us back to Boghossian's regress. How can S follow a rule, if following the rule is a matter of forming an intention with conditional content and if in order to follow the rule, S must first follow an inference rule? How can we explain the necessary movement of thought without a regress—while keeping the distinction between following a rule and acting in accordance with it?

Here, I will assume that the mental architecture of the agents itself is such that it responds to the constitutive structure of the content being represented and that when things go well, the agent's movement of thought that is brought about by this architecture accords with modus ponens. I am thus taking the second option discussed in §2, denying that requirements of the inference rule need to be explicitly represented for the agent to have the required movement of thought, instead making the necessary structure be that of the agent's mental architecture. On this view, whenever a mental representation is tokened by the mechanism that satisfies the requirements of a rule, it thereby tokens the consequent. The idea is that in order to infer, there need not be some state, contra Boghossian, of S that explicitly encodes an inference rule, it is enough that S's

cognitive mechanism itself is such that it is responsive to the contents of the rule in the right way.

I take this view to be a weaker version of an account of inference put forth by Quilty-Dunn and Mandelbaum (2018). The starting point for Quilty-Dunn and Mandelbaum is the notion of a rule being built into the mental architecture of an agent, which holds "iff whenever a mental representation is tokened that satisfies the antecedent of the rule, then, *ceteris paribus*, the system will token a representation that satisfies the consequent of the rule" (Quilty-Dunn and Mandelbaum 2018, p. 7.). They then define a a movement of thought from some state P to a state Q to be inferential iff (a) P and Q have discursive content, (b) some rule is built into the cognitive architecture of the agent having the movement of thought such that P satisfies the antecedent of the rule in virtue of its constituent structure and Q satisfies its consequent in virtue of its constituent structure, and (c) there is no other factor that intervenes in the transition from P to Q (Quilty-Dunn and Mandelbaum 2018, p. 8–9).

The benefit of this view is that, because the rule is built into the agent's mental architecture, the agent need not hold any explicit beliefs about the rule's requirements for their movement of thought to count as inferential. This avoids the regress. Moreover, since the system's functioning is both sensitive to the discursive structure of thoughts and rule-based, the inferential transitions it produces are not merely associative, and are thus immune to deviant causal chains—like those in Boghossian's counterexamples—which are associative in nature.

However, as Warren (2022) points out, Quilty-Dunn and Mandelbaum's approach is grounded in cognitive psychology. Warren, by contrast, treats inference as a folk concept and offers a functionalist account. Indeed, as we will see below, I take inference to be a family-resemblance concept, and for that reason I think the latter approach has much to recommend it. My account could also be given emphasising this aspect. My focus here, however, is on the kind of transition between mental states that occurs in

rule-following, where I follow Quilty-Dunn and Mandelbaum. There are differences between my account and theirs, however. For instance, because they conceive of rules as built into the agent's cognitive mechanisms, they take inferential transitions to be fundamentally rule-governed. I, by contrast, am only willing to grant that the mechanism is such that the movement of thought is in accord with a rule.

But why not just accept Quilty-Dunn and Mandelbaum's account wholesale? There are two main reasons: The first concerns the intrinsic limitations of such a system. The modus ponens *rule* is an argument form that determines an output for any pair of propositions p and q. But no representational system can take just *any* p and q as inputs—as there would be propositions that are too large and complex for the system to represent, due to limitations of memory, attention and so forth. Now recall that Quilty-Dunn and Mandelbaum's definition of when a rule is built into a system says that "whenever a mental representation is tokened that satisfies the antecedent of the rule, then, *ceteris paribus*, the system will token a representation that satisfies the consequent of the rule". It follows that there will be some propositions p and q such that the system cannot represent p or q and hence, that the system will trivially encode *any* rule that is consistent with modus ponens up to the limit the system can represent. That is to say, because the system can really only encode a finite segment of modus ponens, given some ordering of possible inferences, the system is consistent with infinitely many rules.

This would of course not impact any particular inferences that human beings actually make, since all of those would be below the limit of what the stipulated inference system in us could represent—and as such, this limitation will probably not worry Quilty-Dunn and Mandelbaum very much. Nevertheless, by definition, that system would still encode infinitely many rules, and hence we would be following infinitely many rules when reasoning, if the functioning of such a system determined which rules we follow. Furthermore, to modify a point made by Kripke, if we tried to add another *ceteris paribus* clause to the definition, saying that if we were to add enough capacity to the system, it

would encode the modus ponens rule and not other rule, that counterfactual may be true, but only on the presupposition that the system encodes modus ponens and not some other rule. For these reasons, I do not believe that Quilty-Dunn and Mandelbaum can account for rule-following without modifications.

The other reason is the problem of error. The ceteris paribus clause in Quilty-Dunn and Mandelbaum's definition of when a representational system encodes a rule is supposed to rule out what they call *performance errors*—since, as they recognise, the system will not always function in such a way that "the transition successfully occurs". They go on to define a performance error relative to a system S as a

> behaviour or mental event caused either by an intervention by another system that interrupts the normal functioning of S, or by some factor one or more psychological level down (perhaps to a neural level). (Quilty-Dunn and Mandelbaum 2018, p. 11)

This definition seems circular from our point of view, however. The *ceteris paribus* clauses in the definition of what it is for a rule to be built into a system S isn't really spelled out by Quilty-Dunn and Mandelbaum further than being such that they rule out performance errors—a rule is built into the architecture iff "in a world where there are no performance errors, the rule will accurately describe any transition within its scope" (2018, p. 9). But if a performance error is always defined relative to normal functioning of S, both definitions seem to rely on each other, since surely, normal functioning depends on what rule is built into the system, and what rule is built into the system depends on what counts as a performance error. Suppose for example that a neuron misfires, leading to a purported misinference—something that Quilty-Dunn and Mandelbaum allow could happen. Doesn't the very description of it as misfiring depend on what rule is in fact built into the system? And doesn't what rule is built into the system depend on what neurons should fire and how?

These considerations should probably not cause Quilty-Dunn and Mandelbaum to

lose any sleep, since they are content with the functioning of the system not being normatively evaluable (2018, pp. 2, 5 and 15). They would probably not be much moved by such Kripkean underdetermination arguments regarding which rule the system is actually encoding either. For our purposes, however, the preceding arguments show, I believe, that we cannot say that a representative system such as they propose accounts for the difference between following a rule and acting in accordance with a rule. Granted, the functioning of such a system is not merely associative and thus is not vulnerable to the kind of counterexamples provided by Boghossian, but it can also not settle every case nor is it normatively evaluable, and hence, by our lights, does not count as rule-following proper.

In the next section, I will argue that by placing inferences in a basic constitutive practice of inferring, we can fill this gap—with the only cost being that we cannot say that the system itself encodes a rule, but merely that it is such that inferential transitions are in accordance with a rule. Consequently, to infer is not to follow an inference rule but a movement of thought which falls of on the second-order equilibrium path of the basic constitutive practice of inferring.

Inference as a basic constitutive practice

In the last section, I claimed that we can stipulate that the mental architecture of the agents is such that whenever the requirements of modus ponens are represented by the system, the consequent is thereby also represented, and hence that the output of the system accords with modus ponens. The difficulties left over were (1) that since the system cannot represent every possible antecedent and consequent, the functioning of the system is actually consistent with infinitely many rules being built into it, and (2) that we lack a clear criterion for saying when such a system is not functioning properly, and hence that we lack a solution to the problem of error.

In this section, I will argue that by seeing our inferential practices as basic consti-

tutive practices, we can give an account of inference that has all the benefits of Quilty-Dunn and Mandelbaum's account, but where inferences are also robustly normatively evaluable. The simple thought, suggested by the analogy with 'red', is that a movement of thought counts as inference if we, as agents in a basic constitutive practice of using the word 'inference', are disposed to so call it. That is to say, just like our dispositions to judge which things are properly described by the word 'red' determines to which concept the word 'red' refers, our dispositions to judgement about which movements of thought are to be described as 'inference' determines the reference of the concept 'inference' for us.[22] And so, given our linguistic training, we are disposed to call movements of thought such as "He's late, so something must have happened" or "It rained last night, therefore the streets are wet" inferences, and others not.

However, we would also like to give a similar account of our use of the inferential words that figure in such statements, such as 'therefore', 'so', 'hence', 'thus', etc. Each of these words would have an associated practice of using it, each with its own second-order equilibrium path determining the correctness conditions for its use. Unlike the concepts I've used as examples so far, however, namely *red* and *addition*, these kinds of words link two propositions (or more). This means that, through their linguistic training, the agents are forming dispositions about when it is appropriate to assert something of the form "p therefore q" and the second-order equilibrium is determining when such a use is correct. It follows that a correct inference is one that lies on the second-order equilibrium path of the relevant basic constitutive practices, and hence we have a way of normatively evaluating a given inference.

Suppose for example that S accepts that a conditional of the form $p \to q$ is true and that p is true. The thought would be that because of how the structure of the agent's mental architecture is constituted and how they learn to use inferential words, the agent would now be disposed to assent to propositions of the form "$p \to q$, p, so q" (and sim-

22. For the objection that judgement is a contentful state, and thus that my account is circular, see Berg, forthcoming.

ilarly for the other agents in the practice and other inferential words). *S*'s movement of thought would thus be an inference if it falls on the second-order equilibrium path of the relevant practices—that is to say, both those of the inferential words and the term 'inference'. This does not mean, however, that a given movement of thought is an inference just because *S* thinks it is, since the equilibrium path is determined by the dispositions of all of the agents in the practice, and *S*'s disposition can deviate from that of others without the equilibrium path changing.

Inference is not just about using words correctly, however, but to infer what really follows. Here, I would make two points: We are stipulating that the agents have a mental architecture which is sensitive to the requirements of modus ponens, and further that their linguistic training does not occur in a vacuum, but is embedded in concrete circumstances and activities.[23] Through the training, then, agents would be picking up the pattern of the particular circumstances in which it is safe to utter '*p* therefore *q*' and so on, constantly applying the pattern to novel circumstances. Over time, because of their linguistic training and the fact that the agents have a mental mechanism that helps them to think in accordance with modus ponens, the agents' inferential utterances would approach what really follows more and more.

The second is that the claim is not that a certain movement of thought is an inference *because* we say it is, but rather that the correctness conditions of the relevant words are so determined, and hence their meaning. And since meanings are, among other things, a way of classifying things,[24] this would give us a way of distinguishing inferences from other movements of thought which are not, but in a fairly deflationary way. The idea is, like in the case of intention, that 'inference' is a family-resemblance concept, unified by

23. And for this reason, the temptation to refer to basic constitutive practices as 'language games' is almost irresistible. I don't want to over-emphasise the connection with Wittgenstein, however, so I will refrain.

24. Suppose I said: "Let a plane figure with three straight sides and three angles be called 'a triangle' ". I now have a way of categorising things into triangles and not triangles, but that doesn't entail that my speech act somehow created triangles. See Berg, forthcoming for an extended discussion of the objectivity of the underlying account.

the practice of inferring.[25]

This account of inference avoids the rule-following regress, because inference is not seen as following a rule at all, but rather as an output of the agent's built-in mental architecture mediated through the relevant basic constitutive practices. Inferences would be normatively evaluable and there would be no deviant causal chains, since those would not fall on the right second-order equilibrium paths. On this account, then, 'inferential words' like 'therefore', etc. are not a mere incidental feature of inferences—mere clothing in language, so to speak, but rather an essential feature of inference that enables agents to learn what it is to infer.[26] The upshot, however, is that there is no longer any objection to thinking that in order to follow a rule, the agent must therefore infer what to do from the requirements of the rule and the fact that the antecedent obtains.

Revisiting the finitude objection The account of rule-following defended in this paper is not a dispositionalist account of inference in the sense that the agent's dispositions determine what rule they are following. However, dispositions play a central role, since the dispositions of *every* agent determines a second-order equilibrium path of a basic constitutive practice. One might, therefore, worry that a version of the finitude objection comes back, since, as Boghossian points out (Boghossian 2008, 496–497), we do not have the capacity to grasp infinitely long propositions, and thus could not have a basic constitutive practice of inferring that covers infinitely many cases. This is of course essentially the same as my objection to Quilty-Dunn & Mandelbaum's claim that modus ponens could be built into the mental architecture of the agents—and my arguments for

25. The account therefore has certain affinities to that of Miller 2015, who thinks that is

> an appropriate emphasis on training and custom that gives us the right to think of the causal transactions involved in my encounters with the sign post as instances of genuine rule following rather than episodes in a 'mere' causal process.

Here, the customs and practices that *S* are a part of are given in terms of a game-theoretic structure, basic constitutive practices.

26. This account thus has some features in common with Neta 2013. There is not space here to develop it further, but if pressed, I would give an answer in the same vein as Neta (and indeed Quilty-Dunn and Mandelbaum) concerning Boghossian's so-called 'taking condition' on inference. See Boghossian 2014 and Boghossian 2018.

why we could have dispositions of an infinite character regarding 'red' do not seem to apply.

Fortunately, I don't believe we need to solve this problem. There is a distinction, after all, between what really does follow from what, and what our inferential practices are able to capture. Boghossian points out elsewhere (Boghossian 2014) that even if Fermat's Last Theorem follows from the Peano axioms, we cannot *infer* that the theorem is true from the axioms alone. The account of inference just given explains why: There would not be a second-order equilibrium path for such cases, and hence, they would not count as inferences. We can give this same answer in general: When the propositions are too large for us to grasp, we will *eo ipso* not have any dispositions with regards to them, and hence no second-order equilibrium path will be determined. Consequently, such cases would not count as possible inferences—and that seems the right result, since that is the premise of the problem in the first place, that certain inferences are impossible. We simply cannot infer everything that follows and our account of inference should reflect that.

5 Final remarks

The account of what it is to follow a rule that I've been concerned with defending in this paper is that for *S* to follow a rule *R* is for *S* to act on their intention to act in accordance with *R*. The content of this intention is determined by the relevant basic constitutive practices that *S* is embedded in, where the use of a term 'F' is correct if and only if it falls on the second-order equilibrium path of the practice of using 'F'. And so, if *S*'s intention tokens 'F', S's action accords with their intention iff it falls on the second-order equilibrium path of the practice of using 'F'. It follows that *S*'s utterance can be normatively evaluated with respect to the practice and on the assumption that agents can have dispositions towards infinitely many cases, *S*'s intention would thereby have infinite content.

Furthermore, since the second-order equilibrium path of the practice of using 'F' is determined by the dispositions of the agents taking part in it, and since the having of dispositions does not require following a rule, the determination of an equilibrium does not require that the agents follow rules with respect to the terms they are disposed to use. *S* may even *intend* to utter 'F' without being required to thereby follow a rule, because the content of the intention is determined by the practice, not by *S* following rules with respect to the content.

On this view, then, neither the content of our utterances nor our mental states is determined in virtue of us following rules with respect to those contents. The widespread assumption that the content of our intentions get their meaning in virtue of us following rules is thus false on this account. After all, in order for semantic rules to determine semantic content in virtue of us following such rules, we would have to have the intention to follow them and such intentions would themselves have semantic content.

We can likewise avoid Boghossian's regress by stipulating that inference is a transition between thoughts such that whenever the agent's mental mechanism tokens the requirements of modus ponens, it thereby tokens the consequent—within the limits of *S*'s cognitive ability. The requirements of modus ponens are thus not an explicit belief of *S*, but rather a part of *S*'s mental mechanism that tokens mental representations in accordance with it. This transition is thus not strictly speaking rule-governed, but merely acts in accordance with modus ponens. This transition can be placed in the basic constitutive practice of inferring, thus making it normatively evaluable. Therefore, supposing that *S* intends to follow the email rule and receives an email, *S*'s movement of thought from recognising that they have received an email and that this matches the requirements of the rule is a result of this mental mechanism. There is no regress, because the requirements of the inference rule are not explicitly represented by *S*.

Finally, *S* is following the email rule and not any other rule, because, due to the structure of the basic constitutive practice that *S* is embedded in, *S*'s intention only represents

the email rule, and no other rule.

Acknowledgements

I would like to thank the audience at the University of Iceland's Institute of Philosophy, as well as Elmar Unnsteinsson and Fenner Tanswell, for helpful discussion related to the subject of this paper, and the anonymous reviewers for invaluable feedback and criticism.

References

Berg, Ásgeir. 2022. "Rules as Constitutive Practices Defined by Correlated Equilibria." *Inquiry* 65:1–35.

———. 2025. "Problems for 'Standard' Dispositionalist Accounts of Semantic Content." *Synthese* 205 (51).

———. Forthcoming. "An Objective Communitarian Account of Semantic Content." *Ergo.*

Blackburn, Simon. 1984. "The Individual Strikes Back." *Synthese* 58 (3): 281–301.

Boghossian, Paul. 2008. "Epistemic Rules." *The Journal of Philosophy* 105 (9): 472–500.

———. 2012. "Blind Rule-Following." In Coliva 2012, 27–48.

———. 2014. "What is inference?" *Philosophical Studies* 169 (1): 1–18.

———. 2018. "Delimiting the Boundaries of Inference." *Philosophical Issues* 28 (1): 55–69.

Coliva, Annalisa, ed. 2012. *Mind, Meaning, and Knowledge.* Oxford: Oxford University Press.

Glock, Hans-Johann. 1996. *A Wittgenstein Dictionary.* Cambridge, USA: Wiley-Blackwell.

Glüer, Kathrin, and Åsa Wikforss. 2009. "Es Braucht Die Regel Nicht: Wittgenstein on Rules and Meaning." In *The later Wittgenstein on language,* edited by Daniel Whiting. Palgrave-Macmillan.

Guardo, Andrea. 2022. "Yet Another Victim of Kripkenstein's Monster: Dispositions, Meaning, and Privilege." *Ergo* 8.

Hacker, PMS, and Gordon Baker. 2009. *Wittgenstein: Rules, Grammar and Necessity.* Second Edition. Wiley-Blackwell.

Kripke, Saul. 1982. *Wittgenstein on Rules and Private Language.* Harvard University Press.

Miller, Alexander. 2015. "Blind Rule-Following and the 'Antinomy of Pure Reason'." *Philosophical Quarterly* 65 (260): 396–416.

Neta, Ram. 2013. "What is an Inference?" *Philosophical Issues* 23 (1): 388–407.

Quilty-Dunn, Jake, and Eric Mandelbaum. 2018. "Inferential Transitions." *Australasian Journal of Philosophy* 96 (3): 532–547.

Searle, John. 1969. *Speech Acts: An Essay in the Philosophy of Language.* Cambridge: Cambridge University Press.

Vanderschraaf, Peter. 2018. *Strategic Justice: Convention and Problems of Balancing Divergent Interests.* Oxford: Oxford University Press.

Warren, Jared. 2020. "Killing Kripkenstein's Monster." *Noûs* 54 (2): 257–289.

———. 2022. "Functionalism about inference." *Inquiry,* 1–25.

Wright, Crispin. 2001. "Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics." In *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations,* 170–213. Cambridge: Harvard University Press.

———. 2007. "Rule-Following Without Reasons: Wittgenstein's Quietism and the Constitutive Question." *Ratio* 20 (4): 481–502.

———. 2012. "Replies Part I: The Rule-Following Considerations and the Normativity of Meaning: Boghossian, Peacocke, Horwich, and Bilgrami." In Coliva 2012, 379–401.