

An Objective Communitarian Account of Semantic Content

Ásgeir Berg

January 10, 2025

“It is not only agreement in definitions, but also (odd as it may sound) agreement in judgements that is required for communication by means of language. This seems to abolish logic, but does not do so.”

Philosophical Investigations, §242

Kripke’s interpretation of Wittgenstein’s remarks on rule-following notoriously develops a sceptical argument against the possibility of semantic content. One possible response to the sceptical challenge raised by Kripke is the suggestion that the community is the background against which we ought to judge a speaker’s utterances—that “correctness consists in agreement with one’s fellows”.¹ The motivation behind such communitarian views about meaning is, among other things, that since we cannot specify a set of dispositions that can allow for an individual speaker to be mistaken in their utterances, we might instead judge what they say in the context of a speech community, maintaining that an analogous question would not arise for the community; that whatever the community is disposed to judge, under the right circumstances, is what its speakers mean.

Such communitarian views about meaning have been almost universally rejected, however, with very good reason. We might, for example, think that (1) if it is reasonable

1. See Boghossian 1989, p. 534 for this turn of phrase.

to suppose that a speaker can go wrong in their particular judgements concerning a given predicate, there is nothing incoherent in thinking that a community as a whole might similarly go off track. After all, if we think that *I* can be wrong in thinking that the grazing animals I see over there are cows, we should think that *we* all can be similarly wrong. Likewise, (2) communitarian views seem to struggle with getting the extension of our predicates right—if I’m likely to judge that those cows over there are horses, my fellows are equally likely to do so, in which case a communitarian will predict that for us, ‘horse’ means *cow or horse*.²

In this paper, I further develop a game-theoretic solution to the rule-following paradox, first presented in an earlier paper of mine (AUTHOR) and argue that the resulting communitarian solution is able to respond to the kinds of arguments canvassed above, and some besides—and further that such a game-theoretic account of semantic content is no threat to the *objectivity* of our linguistic practices, a worry which arguably lies behind the widespread rejection of such views, independent of specific arguments to that effect, such as the two just mentioned.

Along the way, I respond to a recent argument, structurally similar to (2), due to Andrea Guardo—the so-called privilege argument (see Guardo 2022).

I Basic constitutive practices

For the rest of the paper, I will argue that the semantic content of a given symbol *F* is determined by the game-theoretic equilibrium of what I will call the *basic constitutive practice* of using *F*. This equilibrium is determined by the dispositions to judgement of every agent taking part in the practice, and as such, the account is in essence a communitarian account of meaning whereby the correctness is defined by agreement with that equilibrium.

2. I will refer to (1) as the community problem of error and (2) as the ‘horsey cow’ case. The latter is due to Boghossian 1989.

In the first part of the paper, I will give an overview of basic constitutive practices and explain how they can be used to solve the rule-following paradox. The first part of this section follows Berg 2022 quite closely and thus covers much ground already trod, but is necessary to orient the reader before moving on to how the account can respond to the arguments I am interested in here, namely objections to communitarian accounts and their purported lack of objectivity. In order to save space, I must also refer to that paper for a detailed discussion of how I conceive of the problem.

In a nutshell, however, my focus in giving this account is to show how basic constitutive practices can provide correctness conditions for semantic content for indefinitely many cases, all the while ruling out deviant interpretations of what a given agent *S* might mean. As such, I do not really conceive of it as a problem about rule-following *per se*, but rather about the determination of meaning or content by finite means: How is it possible that *S* means *addition* by ‘+’ if any fact about *S* seems consistent with *S* meaning some deviant function we can call *quaddition*?

In order to make exposition easier, I will start by briefly mentioning the three main objections to communitarian accounts I will be concerned with giving a reply to, before moving on to giving my own account.

Problems for communitarian accounts The first problem (1) I will consider is essentially a re-statement of the rule-following paradox at the level of the community, instead of the individual. After all, just like an individual agent, the community has only calculated finitely many sums, and so if there is question about what determines correctness for the individual, there is a similar question for the community. A communitarian can thus readily explain what it is for an agent to make a mistake (it would be to fail to do what the community does) but has no such recourse in the case of the community itself.

Such accounts have thus only shifted the problem up a level, instead of solving it.³ I will refer to this as “the community problem of error”.

The second (2) is Boghossian’s ‘horsey cow case’.⁴ Boghossian points out that if agents are likely to make systematic mistakes in their judgments, communitarian accounts cannot get the extension of our predicates right. For example, I might be disposed to call horses ‘horses’, but at the same time, also be disposed to confuse them with cows on dark and foggy nights. These kinds of mistakes are *systematic* and arise because of features that every agent’s perceptual faculties have (we are all liable to see badly in the dark) and will therefore contaminate upwards into the community’s determination of meaning: if every agent has a certain disposition to judgement, for example, to judge that a cow in low lighting conditions is a horse, their practice should determine a concept where that is in fact correct, and so, there is no possibility of the community making a mistake of this kind. The communitarian is therefore forced to admit that ‘horse’ really means *horse or cow*, and not merely *horse*.

The third (3) problem I will address has not, as far as I know, been discussed in the literature before, even though it is related to both of the other two problems. This is what I will call “the logical problem of error”. The problem is this: if single agent can make a mistake, it should be *logically possible* that every agent makes a mistake. We can all be wrong, after all, and consensus does not entail correctness. *Prima facie*, however, one might think, that this is impossible: if every agent makes the same mistake, then that would be correct, according to communitarian accounts of meaning, and hence, there was no mistake, merely a shift in meaning. A communitarian account would therefore predict that it is in fact logically impossible that every agent makes a mistake and consensus would, *contra* common sense, entail correctness.

3. Examples of community accounts vulnerable to this objection would be e.g. Wright 1980 and Hauge-land 1998.

4. See Boghossian 1989, p. 536.

As we will see later, a solution to (1) does not entail a solution to (3), and so, these problems are distinct.

Basic constitutive practices, supergames and equilibrium paths

Now, the distinction between regulative rules and constitutive rules is a familiar one.⁵ On Searle's account of this distinction, constitutive rules are rules that "create and define new forms of behaviour" (Searle 1969, p. 33). The rules of chess, to take the canonical example, make it possible to play chess, since without the rules, there is no such thing as playing chess and no such thing as the particular moves of chess. The rules, we could say, constitute the game. We could also, following Rawls, say that the rules provide the 'stagesetting' by which we can evaluate a given action as a move in chess (Rawls 1955).

Consequently, I will say that P is a constitutive practice if it is only possible to say what it is to take part in P by referring to some kind of stagesetting that defines what P is. According to this definition, chess is a constitutive practice, because the rules serve as the stagesetting that defines chess. I will further define a *basic constitutive practice* as a constitutive practice which does not require some *other constitutive practice to serve as its stagesetting*.

However, since rules are among the thing to be explained, I will not require that stagesetting come in the form of rules.⁶ Instead, the *game-theoretic structure* of basic constitutive practices will serve as their own stagesetting, allowing us to evaluate actions relative to this structure.

Now consider the following coordination game due to Lewis (1969). Two agents, R and C , want to meet, but do not care where. The best place to go for each of them is to go where the other goes: it is best for R to go where C goes, and best for C to go where R goes. Now suppose they have a selection of three different places and that they receive a positive pay-off if and only if they meet. We can then represent all possible combinations

5. See e.g. Glüer and Pagin 1999 and Reiland 2020 for discussion.

6. I will not, however, discuss rules as such in this paper very much.

of meeting places by the matrix in Figure 1. The first number in the ordered pairs repre-

	<i>C1</i>	<i>C2</i>	<i>C3</i>
<i>R1</i>	(1, 1)	(0, 0)	(0, 0)
<i>R2</i>	(0, 0)	(1, 1)	(0, 0)
<i>R3</i>	(0, 0)	(0, 0)	(1, 1)

Figure 1: Meeting game

sents the pay-off for *R*, and the second that of *C*. We will say that a coordination game is in equilibrium if no agent would unilaterally change their move and receive a higher pay-off. The equilibrium therefore represents a state where no agent would do anything other than they in fact did, given what the others have done. In the case of the meeting game, we see immediately that the equilibrium lies on the diagonal.

If *R* and *C* keep meeting regularly, we can define a sequence of such meeting games

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

I will refer to such a sequence as a *supergame* and the index *t* as a period. Now suppose the agents adopt the following strategy:

f_i : If we haven't met, choose an arbitrary place, otherwise go to the last place we met.

The strategy system $\mathbf{f} = (f_R, f_C)$ is an equilibrium of the supergame Γ , since neither *R* nor *C* could get a higher pay-off if they deviated from f_i .⁷

It is important to note that when the agents choose a strategy for such a supergame, they are not thereby choosing a particular rule or regularity that fixes what their individual actions will be. It could happen, for example, that both *R* and *C* arbitrarily choose to go to 1 in the first period, and thus, according to f_i always go to 1 afterwards. It may also be that they do not manage to coordinate in the first period, in which case, they make

7. This game-theoretic framework is based on work by Peter Vanderschraaf (2018).

another arbitrary choice in the next, and so on. The choice is fixed, however, after they meet for the first time, but not at the time of strategy selection. The strategy does not, in other words, determine what any particular action will be.⁸

The actual actions taken at a given period t I will refer to as an *act profile* s_t of t . If we select an act profile which is in equilibrium from each period t of (Γ_t) , such a sequence is called an *equilibrium path* of (Γ_t) . As we just saw, the strategy system f is compatible with many different ways of being in equilibrium and so, defines infinitely many equilibrium paths through (Γ_t) .

The point of this definition is to give us a way to refer to all the different ways a strategy can be manifested in practice—for example, the sequence of act profiles where R and C always go to 1 is one equilibrium path, as well as the sequence where R goes to 2 in the first period and C to 3, and then they both go to 1 in every period afterwards.

The basic constitutive practice of adding

At the basis of my account are two claims about meaning. The first (i) is that meaning is a social phenomenon; the second (ii) that the way we learn and acquire concepts is constitutive of those very concepts. Basic constitutive practices are meant to cash out these claims in a way that makes them plausible and able to fend off well-known objections to (i).

My objective in this section is to use the definitions from above to define a *basic constitutive practice* for the use of a given term. I will follow convention and start off by using ‘+’ as my example, even if the idea is meant to generalise.

I will define the correctness conditions for the use of the symbol ‘+’ by building up the basic constitutive practice of using it in terms of coordination problems like the

8. When we go on to develop the idea of a basic constitutive practice, it is important that they have two properties: (1) that the content of the regularity is not specified in advance, and (2) that the agents are not required to grasp the content of the strategy they follow. In both cases, we would be presupposing semantic content, and consequently, the account would be circular. The meeting game has the first property, but not the second.

meeting game. The idea is that the practice of using the symbol in the language has this structure in basic cases and that this structure can serve as the stagesetting by which we can evaluate an agent's actions against. We will say that an action counts, for example, as a correct use of the term '+' if it lies on the equilibrium of the basic constitutive practice of using that term. *In this way, basic constitutive practices can define their own correctness conditions.*

We will make a number of assumptions about the agents. First off, we will suppose that the agents go through a process of linguistic acquisition—by which I mean any stimulus from the environment and others agents that shapes the agent's dispositions to judge that a certain object falls under a given concept or that a calculation should be continued in a particular way. For simplification, I will refer to this process as the agent's *training*, even if explicit training and instruction only forms a small part of the process. We will also, in order to keep things simple, assume that this process takes the form of a finite series of interactions with a set of exemplars of the concept they are acquiring. For example, *S* will have been shown a number of examples of the colour 'blue' and is then expected to identify blue things in novel cases.⁹ Defining correctness conditions for the use of a given term, for indefinitely many cases, is therefore equivalent to defining a sameness relation from the exemplars of that term to novel cases—what counts as the *same*.

We will also assume that the agents are so endowed that their dispositions are shaped in similar (but not necessarily identical) ways as they go through this process, which we do not need to suppose ever ends. For example, we can suppose that an agent is inducted into the practice of using the term '+' by being shown examples and having certain techniques explained, or maybe in the case of 'red', by being shown red objects, and that

9. In posing the problem like this, I've been influenced by Bloor 1997, pp. 9–14 and Kusch 2002, pp. 202–203.

consequently, the dispositions of the agent will be sufficiently similar to that of the other agents in how they react to new cases.¹⁰

Finally, we assume that the agents follow the strategy of judging what they are disposed to judge in a given case, or alternatively, of giving the reply they are disposed to give. We do *not* assume that the agents *prefer* to do so, since preference is an intentional state, and thus already has semantic content, but merely that they do so act.

Note that at this stage I do not suppose that ‘judging’ is an intentional act, i.e. I do not suppose that what I am calling ‘judgements’ here have semantic content; merely that the agents give certain replies, for example, that they write down the *string* ‘125’ or utter the *word* ‘blue’ under the right circumstances. I will revisit this point below.

The simplified ‘+’ game The basic building block of the basic constitutive practice of using the term ‘+’ is what we 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 2. Because we assume that the game can be repeated indefinitely, it is a supgame of the

	‘5’	‘125’
‘5’	1, 1	0, 0
‘125’	0, 0	1, 1

Figure 2: Simplified ‘+’ game for two agents

form $(\Gamma) = \Gamma_1, \Gamma_2 \dots$ where each period stands 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 June 17, 2044 at 12:00:01, for example, there’s a game for that, and so on for every possible in-

¹⁰. There is a worry that my use of *similar* is begging the question against the sceptic. A similar concern might be raised for the notion of ‘pay-off’. I address these concerns in Berg 2022) and hope to elaborate on them elsewhere.

stance. The point of the second-order equilibrium path is not just to have enough games to cover each sum, but *each instance* of the symbol being used, too.

Since we assume that the agents follow the strategy of replying with what they judge to be the correct answers and that their dispositions to judgement are very similar, only one equilibrium path will be selected through the simplified ‘+’ game—and because the agents acquire their dispositions through similar training, their choice of strategy is not probabilistically independent and thus the equilibrium is a *correlated equilibrium*.¹¹

The generalised ‘+’ game We can now generalise the simplified ‘+’ game by allowing indefinitely many agents to take part, in line with our definitions above, as well as allowing them any reply. We can also define a generalised ‘+’ game for any n and m by replacing the concrete numbers in the example above by any number. We then have a sequence of generalised ‘+’ games of the form

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

where each supergame in the sequence is indexed by period t , and is therefore repeated indefinitely many times. This sequence is thus an infinitely long sequence of supergames, each of which consists of games that are repeated indefinitely many times.

Here we make a key move. Notice that if we select one equilibrium path from each supergame in the sequence, that selection corresponds to one possible interpretation of ‘+’. In particular, the selection where the agents give an answer corresponding to *addition* represents *addition*. We can likewise make a selection corresponding to *quaddition*, if we chose the paths corresponding to *addition* for every generalised ‘+’ game up to $n = 57$ and the equilibrium path given by $s_t = \{('5', '5', \dots)\}$ in any game that comes after. I will refer to any such selection as a *second-order equilibrium path* through $(\Gamma)_+$.

11. This solution concept was first studied by Aumann (1974, 1987) and is more general than the more famous Nash equilibrium. See also Vanderschraaf 1995, 1998, 2018 and Gintis 2009.

We now have the stagesetting we require, since we can evaluate a given utterance in the basic constitutive practice of using the term ‘+’ as correct *if it falls on the second-order equilibrium path of that practice*. This gives us an answer to Kripke’s sceptic: *S* means *addition* by their use of the symbol ‘+’ and not *quaddition* because *S* is taking part in a basic constitutive practice of using the symbol ‘+’ where only the answers that correspond to addition lie on the second-order equilibrium path of that practice. The sceptic cannot make the further move of asking what makes it the case that *S*’s utterance takes place in the context of an *addition* practice and not the one that corresponds to *quaddition*, because such a practice is not actual—i.e. given the dispositions of *S* and the other agents only the second-order equilibrium path of *addition* will be selected. It is the *structure* of the practice itself that provides the stagesetting and there simply is *no* quaddition practice.

Given the game-theoretic machinery then, a basic constitutive practice will provide correctness conditions for any case that the agent’s dispositions settle, as soon as those dispositions are fixed.

The problem of finitude The aim of this paper is to argue that by appealing to basic constitutive practices, one can give a community solution to the paradox that can solve common problems for such solutions, as well as preserving objectivity. As we’ve just seen, dispositions play an important role in the definition of such practices, and so, such a solution would be a non-starter if it cannot answer more general problems for such accounts that have been raised in the literature. Three of the most prominent were raised by Kripke already, namely (a) the argument from normativity, (b) finitude and (c) error. Before moving on to the problems facing community solutions, I will briefly address the last two of these.¹²

12. In general, I adopt a very weak understanding of normativity, and side with the so-called anti-normativists (see e.g. Hattiangadi 2006, Hattiangadi 2007 or Wikforss 2001) and accordingly understand the argument as a challenge to provide correctness conditions for the use of terms (see e.g. Glüer and Pagin 1999, Hattiangadi 2006 or Verheggen 2011). And so, I am not side-stepping the issue, since that is what this

The problem of finitude is the objection to dispositionalist accounts of meaning that the requirements of meaning are infinite, but *S* only has the disposition to give a reply to an addition problem when the numbers are small enough (see Kripke 1982, pp. 26–28). One might then think that the stagesetting provided by basic constitutive practices is not sufficient to settle indefinitely many cases as soon as the dispositions of the agents are fixed, as it is doubtful that they can have the disposition to give a reply for infinitely many cases (as a proper solution would require). *Pace* Kripke, I believe this is possible.

Let's consider a different concept first, that of an object being *red*. It is true that I do not have the disposition to sit through an endless parade of objects and judge the redness of each object that passes in front of me. Eventually, I would get too tired or even die. But it does seem plausible, however, that I am able to say of *any* object in such a sequence whether it is red or not: take an arbitrary object from the sequence and I will be able to give some judgement about its colour (setting vagueness aside)—and why not? Thus put, we can see that the argument from finitude isn't really about the possibility of having infinitely many dispositions regarding a certain class of objects, since there are indefinitely many red objects and there is nothing implausible about me having a disposition with regard to *any particular one of them*.¹³ It is therefore possible that our dispositions can fill out the second-order equilibrium path of the basic constitutive practice of using the term 'red'.

But what about '+'? Matters are not as simple here, since we clearly do not have a disposition to reply with the sum of any two numbers. In fact, it is only plausible to suppose that we have a disposition to reply with the sum of very small numbers. However, our actual mathematical practice relies on calculations and techniques and our acquisition of arithmetical concepts cannot be separated from our mastery of these techniques.

whole paper is about. For a further discussion of the problem of normativity, and the idea of being guided by the meaning, see Berg 2022.

13. For this reason, Andrea Guardo prefers to call the argument 'the Cardinality Argument' (Guardo 2022). He points out that the problem really about showing that our dispositions can track more than just a finite segment of e.g. the addition function. Failing to do so is, as we've just seen, consistent with having infinite dispositions.

The problem then reduces to us having dispositions to sum small numbers, carry the one and continue the calculation—and that is quite similar to our dispositions in the case of *red*. After all, why shouldn't I have such dispositions in *any* case?

This kind of approach is reminiscent of Warren's recent solution to the problem of finitude and is, I believe, successful in showing that the problem of finitude can be solved (see Warren 2020). Warren points out that there is no particular difficulty in supposing that *S* has a disposition to add small numbers, carry and move on to the next step in the calculation, and that if that is so, there is a sense in which *S*'s dispositions do encode *S*'s "unbounded execution of the process" (Warren 2020, p. 8). *S* has, for example, a disposition to move on to step 18 in a given algorithm, after having executed step 17—and this is true, even if *S* has not completed the previous steps (imagine, for example, that someone else has already finished them, and then *S*'s shows up: why wouldn't *S* be disposed to sum two small numbers, carry and move on?).¹⁴

There is therefore a sense in which *S* is disposed to execute step 17, even if they are otherwise prone to never get there. This holds for any two steps n and $n + 1$ of the algorithm, irrespective of the size of n . We can therefore agree that even if *S* does not have the disposition to finish a sufficiently long calculation, *S* may nevertheless have the necessary dispositions for *any* step in the calculation, as they are all psychologically tractable ('add two small numbers', 'carry the one', 'move on', and so on).¹⁵ In this way, we can fill out the second-order equilibrium path of using the term '+' and solve the problem of finitude.

Note, however, that I do not claim that *S* is thereby *adding*, merely that *S* has dispositions that cover any case where *S* might be expected to add.¹⁶ We could even, if neces-

14. Cf. Turing's discussion of computation in his famous 1936 paper (see p. 253-4).

15. We could also say, following Peacocke, that we find such cases and transitions between them 'primatively compelling' (Peacocke 1990). For criticism of Peacocke, see Kusch 2006, p. 129, and for why this criticism does not apply here, see the next footnote.

16. I do not say that as a result of their training, *S* has a disposition to add. That would give Kripke's arguments against the algorithm view traction (see Kripke 1982, p. 15-17). I only claim that as a result of their training, *S* has dispositions to do *something*. The correctness conditions, and hence the identity, of the concept come later and are fixed by the practice. I will return to this below.

sary, weaken this requirement, since the game-theoretic machinery does not require that every agent has a disposition for every case. It is only necessary that a *plurality*—which need not even be the majority—of agents has a disposition to give the same answer in a given case for the equilibrium to settle there. The account is therefore very forgiving in this respect.

The problem of error

I’ve just briefly discussed the problem of finitude and will now turn my attention to the problem of error. The problem of error is relatively simple. It is the problem of explaining why *S*’s claim that $68 + 57 = 5$ should be viewed as an incorrect addition, rather than a correct quaddition. In other words, the account must allow for *mistakes*—it cannot predict that any purported mistake by *S* is really just a different calculation whereby he does the right thing, and thus indicates a difference in meaning; an account of meaning must explain how there can be a gap between what we mean and what we do.

Accordingly, we should also be able to explain mistakes that are not linguistic in nature, whereby a word is used correctly, but nevertheless misapplied. For example, if *S* says (and means to say) that particular animal in the distance is a cow, but it turns out to be a horse, then *S* has not made a linguistic mistake, but merely made a mistake about what they saw. *S* thought the horse was a cow, and so, relative to *S*’s intention of using the word, *S* used the word correctly. Likewise, *S* could use a word incorrectly, but apply it correctly (e.g. if *S* thought that ‘horse’ means cow and *mistakenly* calls some horses ‘horses’).¹⁷

Now, according to my account so far, *S*’s utterance is correct if it falls on the second-order equilibrium of a basic constitutive practice, and incorrect otherwise. This is not enough, however, since if *S* intends to say that the horse is a cow and does so, then *S*’s use

17. For this distinction between correct application and correct use, see Millar 2002.

was correct, even if it does not fall on the second-order equilibrium of the term ‘horse’. The correctness conditions of *S*’s utterances therefore seem to depend on what *S* intended to say, and I do not think *S*’s intentions are eliminable here—nor do I think we should want to eliminate them.

In the literature on rule-following, however, it has been taken to be mysterious how intentions could do this job, since the rule-following problem has been seen to infect all our mental content, including intentions.¹⁸ Since the view under consideration is an externalist one, however, we can do for mental content what we did for linguistic 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. In the case of ‘+’, we could, for example, give a simple account by supposing that *S*’s mental state somehow tokens ‘+’ and that its full content is then given by the practice. So, if *S* intends to add, the practice settles the correctness conditions of the term ‘+’—tokened by their intention.¹⁹

This explains how *S* could fail to hit the equilibrium: *S*’s mental state tokens ‘+’ but *S* does not perform the action that lies on the equilibrium of the practice, and was thus wrong—and not merely right relative to a different meaning. Similarly, if *S*’s intention tokens the term ‘cow’, but the animal was in fact a horse, then *S* was incorrect, since the second-order equilibrium path of using the term ‘cow’ wouldn’t have settled on that case.²⁰

18. See e.g. Boghossian 1989. Crispin Wright is a dissenting voice, see e.g. Wright 2001 and Wright 2002.

19. Here I’m picking up on a suggested reading of Kripke by Boghossian:

On another way of reading it, Kripke would be assuming not some controversial view of the relation between thought and language, but rather that thoughts themselves involve the tokenings of expressions (of mentalese) and that those expressions, too, get their meaning by our following rules in respect of them. (Boghossian 2008, p. 487)

The suggestion amounts to replacing the purported role of rules with that of basic constitutive practices.

In general, I just have in mind the well-known distinction between “representational vehicles”, whatever they are, and the content that fills them. The suggestion here is that the content is fixed by the practice *S* is embedded in.

20. This feature of the account may be of use in explaining why the epistemology of meaning is authoritative for the speaker, despite being ultimately analysed in terms of dispositions: *S* has a ‘meaning intention’ and when *S* reports on the content of that intention, the content of *S*’s utterance describing the intention

A similar externalist account could be given for ‘judgement’. With all of this apparatus in place, we can now say that when S *judges* that a is F , the content of that judgement is given by the practice S is a part of. This is not viciously circular, as the same sentence token can serve as input to the game-theoretic machinery while also expressing a proposition with content; we need not assume that content for the former to go through.

One might worry that S ’s utterance falling on the second-order equilibrium of the practice of using the term ‘+’ is consistent with S meaning *addition* by ‘+’ and being correct, as well as being consistent with S meaning *quaddition* and being incorrect.²¹ However, because the account is an externalist one, we can rule this out. Suppose for example that S ’s mental state tokens ‘+’ and that S is embedded in an addition practice. If S ’s answer accords with *addition*, then S meant *addition* and was correct, but if S ’s utterance does not so accord, then S meant *addition* and was incorrect. There is, in this case, no question of S meaning *quaddition* and being incorrect, because S is not a part of a *quaddition* practice, and thus the content of S ’s intention cannot be filled out that way. This is because, given the dispositions of all the other agents, the second-order equilibrium path being selected is only consistent with *addition* and not *quaddition*, even if such a practice is possible. S meaning *quaddition* is therefore not consistent with S being a part of an addition practice, as the content of S ’s intention is determined by the practice S is a part of.

Does this mean that S cannot mean *quaddition* by ‘+’? It might seem that the account rules out this possibility, which would, in my opinion be too strong a result, since seemingly we can mean deviant things with our symbols (speaking in code might be one example). If we suppose, however, that when S ’s intends to mean *quaddition* by ‘+’, their intention somehow tokens the *definition* of *quaddition*, where the terms in that

depends on the same basic constitutive practice as the intention itself—both the intention and the report derive their content from the same practice.

S ’s willingness in giving *that* report could in turn be explained by S ’s dispositions.

21. This is, I believe, Miller and Sultanesu’s objection to Warren’s solution to the problem of error. See Miller and Sultanesu 2022.

definition get their meaning in turn from the practices regarding their use that S is a part of, then S can mean *quaddition* by ‘+’ by intending to use the term in a deviant way. What is ruled out, however, is that the deviant meaning is *basic* for S —that S can mean *quaddition* by ‘+’ without the definition of *quaddition* being involved in some way. S can, however, do so in the case of *addition*.

The problem of community error This leads us to the (1) problem of community error. This was the problem of explaining, given that an agent can make mistakes, how then the community can make a mistake. After all, the community has, just like an individual agent, calculated finitely many sums, and so if there is question about what determines correctness for the individual, there is a similar question for the community.

This objection depends on the assumption that the dispositions of the agents translate into dispositions at the level of the community; that there is such a thing as ‘the disposition of the community’ in a given case. However, on the current account, the community has no dispositions (correctness is determined by the structure of a practice, not dispositions directly) and as soon as the dispositions of the agents are fixed, the correctness conditions of the practice are settled immediately for indefinitely many cases—even for a novel case nobody in the community has yet seen.²² The second-order equilibrium path of using a given term F fixes the reference of F —that is to say, it fixes which concept F refers to—and so, the problem does not arise in this form. There can therefore be no question about the community ‘going off-track’ in the use of a term, since every case is already determined in advance, as soon as the agent’s dispositions to judgement are fixed by their training.

We can also put the point slightly differently: On this view, only agents make judgements and those are what can be true or false—i.e. particular judgements or propositions that express judgements. The way meaning is fixed by the community here is such that

22. This claim depends on the finitude problem being solved—that the agents can have dispositions that cover every case. See above.

questions of truth and falsity do not even arise, as the community never judges or expresses a proposition. There simply isn't anything for the community to be wrong *about* and thus to say that the community got things wrong in this sense is simply a category mistake.

In his celebrated paper on the rule-following paradox, however, Boghossian (1989) suggests that doubts about the coherence of communitarian accounts arise because communal consensus, the very thing meant to secure meaning on such accounts, does not have the characteristics of our ordinary notion of truth—presumably, though Boghossian doesn't explicitly say, that of being *objective*. Boghossian himself does not take this to be an objection to the view in itself, since for him, the communitarian account should not be seen as “an analysis of our ordinary notion of truth, but a displacement of it” and for *that* reason unlikely to be correct (Boghossian 1989, p. 535).²³

I would, however, resist this further step from Boghossian: I am not offering a displacement *nor* an analysis of our ordinary notion of truth and the current account leaves it quite as it is. Here, there are two cases to consider. The first is where meaning is ascribed to *S*. The ordinary notion of truth, I take it, entails that if *S* meant *red* by ‘red’ then the proposition that *S* meant red by ‘red’ is true. That is to say, according to our ordinary notion of truth, the proposition

(P) By ‘red’, *S* means *red*

is true if and only if *S* means *red* by ‘red’. On the analysis of meaning offered here, *S* means *red* by ‘red’ if *S*'s utterance lies on the second-order equilibrium path of the basic constitutive practice of using ‘red’ and so, the proposition *P* is true if this situation

23. Boghossian's worry could thus also be put like this:

For the believer in objectivity, human opinion in no sense constitutes truth; truth is in no sense supervenient upon human opinion. (Wright 1981, p. 99)

It is true that on my account, the agent's dispositions to judgement about particular cases fix the meaning of terms. Nevertheless, I deny the inference to the further claim that their *opinions* are thereby constituting the truth of anything. In the next section, I will argue that even if every agent had the *opinion* that a particular object falls under a given concept, it would not follow that it did—that it is possible that every agent makes a mistake.

obtains. There is therefore no threat to our ordinary notion of truth in this case, nor to objectivity, since there is a fact of the matter what *S* means.

The second case is where *S* asserts a proposition. This is quite similar to the first case: If *S* says for example, that ‘the letterbox is red’ and *S* means *red* by ‘red’, then *S* has spoken truly, if the letterbox is in fact red. As before, it is an objective fact what *S* meant and there is nothing in the account that implies that the letterbox’s redness is not an objective matter (after all, the agents’ dispositions are shaped by their training and this training presupposes an external reality and interactions with it). The agents’ dispositions to judgement fix the reference of ‘red’—its meaning—but it does not follow that they are thereby deciding *that* a particular object is red. The world, and hence external facts (in some sense) play a necessary role in constituting the truth of the statement.²⁴ This outcome is merely the old chestnut that the truth of a sentence depends on its meaning on one hand, and the facts on the other, and so there is, again, no threat to objectivity.

Of course, it is quite possible that agents in a particular practice would apply two predicates where we apply only one (for example, one for ‘dark red’ and another for ‘light red’) or one where we use two (for example using one word for the whole blue-green spectrum) but the fact that different communities might slice reality up into different concepts undermines neither objectivity nor truth.²⁵ But even if there is no sense in which the community can go off-track on this account, we still lack a response to the other two problems, however, the logical problem of error and the horsey cow case. To properly reply to this pair of challenges, we will require further stagesetting, however,

24. Claudine Verheggen makes a similar point about the possibility of objectivity being the result of linguistic training that occurs in “in specific circumstances, in relation to [children’s] activities and objects and events in their environment” (Verheggen 2003, p. 305).

25. The account could even accommodate correspondance theories of truth, since the truth of the relevant statements depend on certain facts obtaining. The solution is thus a *straight* solution and does not, like Kripke’s sceptical solution, require throwing standard truth-conditions overboard and replacing them with assertibility conditions.

There is not space here, however, to get into the weeds of Kripke interpretation. I am, however, in broad agreement with Wilson’s reading of Kripke (see Wilson 1994 and Wilson 1998).

and I will respond to them after a small detour. After that, I will return to the subject of objectivity.

II The privilege argument

Before addressing the last two arguments left on the table at the end of the last section, I will briefly discuss the so-called ‘privilege argument’. It has its roots in Kripke’s discussion of ideal conditions (Kripke 1982, p. 28), as well as some remarks by Paul Boghossian (1989, §23) and Anandi Hattiangadi (2007, pp. 106–8 and p. 117). Recently, it has received a systematic treatment in a paper by Andrea Guardo (2022).

After responding to this argument, we will be in a position to answer the two aforementioned arguments.

Outlining the privilege argument The motivation for the privilege argument is very similar to Boghossian’s challenge to community accounts of meaning, except levied at the individual: *S* might judge that $68 + 57 = 125$ when they are well rested, but not when they are tired. *S* thus has two different sets of dispositions that conflict and give different meanings, and so we must be able to explain why one set, which we can call *C*, should count as meaning-determining, instead of some other set of conditions *C**. This must be done in a way that is neither question-begging nor *ad hoc*. The problem is thus quite simple: Why are my dispositions when I’m well-rested and sober meaning-determining, and not the ones I have when I’m tired and drunk? What privileges the former over the latter?

The most naïve way is to simply claim that the dispositions in *C* determine meaning because *C* are the conditions under which *S* is adding and not quadding. As Guardo points out, this is patently circular (Guardo 2022, p. 872–3). According to this view, *S* means *addition* by ‘+’ because *S*’s dispositions are privileged and *S*’s dispositions are privileged because by having them, *S* means addition. This is of course bad enough, but

this circularity is not the ultimate reason why this proposal does not answer the sceptic. One of the main challenges of the sceptical paradox is to explain what it is that constitutes the correctness of certain utterances and the incorrectness of others—meanings establish norms.²⁶ And by appealing to the fact that certain dispositions give the correct answer and others do not, the dispositionalist has taken such a norm for granted, not explained how it is constituted in the first place.

One might try to characterise *C* as ‘ideal’ conditions where the dispositions that count are those that *S*’s *would* have if, for example, *S* wouldn’t get tired, would live long enough to hear the whole problem, etc. One might suspect, however, given Kripke’s arguments against attempts to solve the finitude problem by appeals to *ceteris paribus*-clauses and idealisation, that such attempts would likewise be question-begging or circular (Kripke 1982, p. 27–28). I believe this is enough to rule ideal conditions out as a non-starter.

In his paper, Guardo also discusses ‘standard’ conditions, which are those conditions under which we most often produce our utterances—where this may be cashed out in terms of agents being e.g. well-rested, sober, not distracted, and so on, as well as the environment being optimal for their perceptual faculties, e.g. well-lit, quiet and not filled with a dangerous gas, and so on. For lack of space, I will not reproduce his arguments against them here.

One might ask, however, how the standard dispositionalist will explain how our meaning can come apart from what we say when we are *not* in standard conditions. That is to say, if I’m tired and looking at a cow, thinking that it is indeed a cow, the standard dispositionalist does not accept that my disposition to utter “That’s a horse” entails that I meant *horse* since that is not what I would have said if I had not been tired. But it cannot be that we mean in non-standard conditions what we *would* mean if we were in standard conditions, since this would reduce standard conditions to ideal conditions. It is equally

26. By this claim I do not mean to say that if *S* means addition by ‘+’, *S* ought to give the reply ‘125’ when asked about the sum of 67 and 58, only that this reply is *correct*.

implausible that we don't mean anything when conditions are not standard—surely I meant *cow* when I said 'horse', even if I'm tired?²⁷

A 'Copernican' revolution Fortunately, I believe that the game-theoretic nature of basic constitutive practices lets us answer the privilege argument without going down one of these three paths, as well as solve the outstanding problems for community accounts, by turning the problem on its head: Instead of specifying which dispositions are meaning-determining in advance, we can let the dispositions that come out of the agent's linguistic training be meaning-determining, whatever they are, thus allowing *any* dispositions that the agents actually have to determine meaning. Thus, the notion of standard conditions is not something we put in, but something we get out.

Above, I argued that the basic constitutive practice of adding requires of the agents that they master a technique, learning how to add small finite numbers, to carry the one and move on to the next step in the algorithm. If we accept that *S* can have dispositions that cover enough cases, one might *still* worry that the sceptic could question that *S* is in fact *carrying*, for example, or *S*'s really moving on to the next step in the same calculation. *S* has only performed a finite number of carrying operations, after all, and it could very well be that *S* is actually disposed to carry the one for *n* steps of the algorithm and do something different afterwards. A similar deviant interpretation can be found for *S*'s moving on.

We can answer the sceptic by focusing on *S*'s training. We cannot say in advance that *S* has been trained to *carry*—since that would be circular—but we can say that *S* has been shown a finite number of examples of what we would call *carrying* (in the metalanguage) and that consequently *S* has formed a disposition to give *some* reply when presented with a suitable case. This kind of disposition is not very psychologically demanding and it is,

27. For further discussion of standard dispositionalist accounts especially Warren's, see Berg, Forthcoming.

as I argued above, plausible that as a result of their training, *S* has such a disposition for *any* given step in the algorithm.

We can then say that every agent that has gone through such training and takes part in the practice has a disposition to give *some* reply when presented with a suitable case. These dispositions then define a basic constitutive practice of using the term ‘carry’ for the agents that take part in it and thus defines the correctness conditions of the term; what it is for these agents to ‘carry’ is therefore for them to perform the action that lies on the second-order equilibrium of the practice. This is not circular, because nowhere have we tried to demarcate which dispositions are meaning-determining and which are not, but take any disposition that the agents have as a result of their training as meaning-determining.

The agents do not mean *carrying* (as that word is used in the metalanguage) unless the equilibrium of the practice tracks actual *carrying*, however, which it may or may not do. But this is not a problem, because if we were to suppose that we ourselves belong to such a basic constitutive practice, then our dispositions would pick out some equilibrium and that would be, by stipulation, that of *carrying*.²⁸ The question of whether the agents mean *carrying* is thus just a question of whether they do what we do.²⁹

What I have said so far is not enough to solve to answer the privilege argument, nor to solve the two outstanding problems for community accounts, however. In both cases the problem is that agents are likely to have different dispositions under certain conditions than others, for example when they are tired, as opposed to when they are well-rested, and so we need to explain without circularity why the former kind of dispositions are not meaning-determining.

28. The question: “How do we know that this is *carrying*?” is misguided: On this account, some second-order equilibrium path would be defined and *whatever* that is, that is what carrying is—the meaning of the word ‘carrying’.

29. Here I merely have in mind the same kind of distinction between our language and *S*’s language as D. Lewis 1974.

I believe that we can solve this problem by appealing again to our assumption that the agent's meaning-determining dispositions are formed by their linguistic training. Until now, we've only been considering first-order dispositions, dispositions to judge that a certain object falls under a given concept or that a calculation should be continued in a certain way. But *S*'s linguistic training is such that they do not only form such dispositions, but also higher-order dispositions about the circumstances under which they give certain replies. For example, *S* might have learned to use colour words by being shown examples. This training would privilege certain conditions, since if *S* were to examine a necktie under unusual electric lightning, for example, *S* might then judge that the tie is blue, but if they were to make the same judgement in broad daylight, they might judge that it is green. But because of the way *S* has learned how to use the terms 'blue' and 'green', *S* would be disposed to withdraw their first judgement, but would stand firm with respect to their second.³⁰

I will therefore say that a disposition is stable if there is no higher-order disposition that overrides it and stipulate that *S*'s disposition to judgement in a given case counts as determining a second-order equilibrium of a basic constitutive practice if and only if *S*'s disposition is stable. This is not circular, because *S*'s meaning-determining dispositions are not defined in terms of standard conditions, but rather, standard conditions are defined as those conditions in which *S*'s dispositions are stable. The explanation for stability is thus not standard conditions, but rather which higher-order dispositions *S* has acquired through their training. The way the agents learn the concepts is thus constitutive of the meaning of the words they use.³¹

30. The example is from Sellars 1997, p. 37–39.

This notion of stable disposition is similar to the one found in Johnson and Nado 2014 (see also Johnson and Nado 2016, Johnson and Nado 2017). For Johnson and Nado, *S* means *red* by 'red' iff *S* would be disposed to apply 'red' to red things, if *S* had all the relevant information. For further criticism and discussion, see Andow 2016 and Nyquist 2020.

31. This notion of stability is not the same as that of Warren 2020. For him, stability is a statistical notion, whereby *S*'s answer α to a given problem is *stable* if and only if the ratio of *S*'s non- α answers tends zero as the number of *S*'s attempts to give an answer increases.

One might worry that we can only characterise these higher-order dispositions by presupposing the notion of semantic content—that these are dispositions to *doubt* or at least *assess* a prior response, and that such judgements have semantic content. I do not believe, however, that what we’ve said requires such contentful judgements in the general case, but merely that the agent has a disposition to stand firm in some cases and a disposition to withdraw a response in others. In practice, how this will look is the agent exhibiting doubting behaviour (hesitating, seeking confirmation, correcting themselves, etc.) under certain circumstances and not others, and *some* of that will be in the form of linguistic behaviour and contentful judgements.

Such cases, however, will only be of a mature agent who has already undergone much linguistic training already. Here, a similar point holds as above: Such judgements will get their content from the practice *S* is embedded in, but that is not circular since the same response can serve as input to the game-theoretic machinery while also expressing a proposition with content. At bottom, what matters are the dispositions to either stand firm or withdraw a given judgement—and how *S* acquires those is explained by the training process.

If that’s right, this immediately (dis)solves the privilege argument: By turning the problem on its head, there is no reason to try to demarcate some dispositions as special, since the meaning-determining dispositions are not specified as inputs to the account, but taken to be those that as a matter of fact come out of the training.³²

Solving the logical problem of error and the horsey cow case Now consider a case where the agents are subject to an optical illusion, being presented with two lines on a piece of paper such that the lines are in fact unequal in length, but because of how they are drawn, they appear to the agents as being of the same length. If the agents were

32. One might worry that the problem recurs, that we can now ask the question: Why are our trained dispositions meaning-determining and not our untrained dispositions? The answer to this worry is that training is a temporal process, and untrained dispositions are replaced by trained dispositions. If there were any untrained dispositions after the process, those would still count as meaning-determining.

taking part in a basic constitutive practice of using the term ‘length’, we might think that they would then judge that the lines were of equal length, and thus that their use was correct—and hence that they could not make a mistake, even if intuitively, we would want to say that they got it wrong.

However, if we suppose that the agent’s training and practice with regards to the term ‘length’ is the same as ours, in that they also measure length with rulers, lay things on top of each other to see which is longer, and various things like that, then they would in fact have higher-order dispositions with regards to their judgement about the length of the two lines. In this case, the agent’s stable disposition would be to judge that the lines are in fact of unequal length, and thus the practice would determine the right second-order equilibrium after all. If, on the other hand, their practice is such that they only go by visual impression, they would simply mean something other than we do by the term ‘length’ and there would be no reason to say that they made a mistake here in the first place.

Now recall that we analysed *S*’s making a mistake as *S*’s mental state tokening a given term but their actual response missing the second-order equilibrium path of the corresponding basic constitutive practice. For example, *S*’s mental state might token ‘cow’, because *S* thought that those horses are cows, but the basic constitutive practice of using the term ‘cow’, of which *S* is a part, does not settle on an equilibrium in that case. In the case of the lines, there is no particular difficulty in supposing that this happens to every single agent simultaneously, since it is possible to give a reply that does not correspond to one’s stable disposition, and thus possible for anyone to do so. It is therefore possible that every agent makes a mistake on this account, and thus it solves the logical problem of error.

The horsey cow case is similar. Boghossian is certainly right that agents are wont to make systematic mistakes that might contaminate upwards and result in a community account of meaning giving the wrong verdict. But by appealing to the agent’s training

and how it shapes their higher-order dispositions, we can give a reply: the nature of systematic mistakes is such that the dispositions that lie behind them are not stable and agents would always have learned to have further dispositions to doubt them in the case of horsey-looking cows on foggy nights. Their stable dispositions, however, do settle in the right place, even if no one actually gives the correct reply.

III Objectivity and judgement-dependence

I've argued that the account on offer can (a) answer Kripke's sceptic without circularity, especially with regard to the privilege argument, and (b) answer a number of difficult problems for such accounts, including the community problem of error, the logical problem of error, and Boghossian's horsey cow case. I've also claimed that the account is not a threat to our ordinary concept of truth nor to the objectivity of our linguistic practices.

There are, however, still reasons to be sceptical of this last claim. The determination of a second-order equilibrium path of a basic constitutive practice depends, as Crispin Wright put it in a different context, on there being "a shared uptake, a disposition to concur in novel judgements involving the concepts in question" (Wright 2007, p. 487), and so, one might worry, as Wright does, that

the requirements of a rule, in any particular case, are simply whatever we take them to be. For if the requirements of the rule are not constituted, as the platonist thinks, independently of our reaction to the case, what can be available to constitute them but our reaction? But that idea effectively surrenders the notion of a requirement altogether. (Wright 2007)

If we transpose this concern to the account under consideration here, it is quite simply that if the constitution of meaning is not independent of our dispositions to judgement about particular cases, then what is true or false about a given case is not independent of

our dispositions to judgement either. And this would seem to undermine objectivity—if certain facts are judgement-dependent, how could we be wrong about them?

There are no prospects of *solving* this problem on this account. But can we live with it—or perhaps better, stop seeing it as a problem? I believe we can, and that there are reasons to accept this as a benign feature of our actual linguistic practice, rather than a knock-down argument against communitarian accounts of meaning.

The case of the Waywards Suppose, for example, we encounter an alien race, the Waywards, whose practices are in many ways similar to ours.³³ They have a number system and a counting practice that looks a lot like ours, as well as an arithmetical operation ‘+*n*’. The Waywards, however, are such, perhaps for cultural and biological reasons, that their trade and commercial activities tend to revolve around the diminishing marginal value their exchanges have for the average Wayward. These facts cause Waywards to react to the teaching of arithmetical concepts differently than we do. For them, the most natural way of successively applying what they call the operation ‘+2’ past 1000 is to continue “1004, 1008...”. Likewise, when the numbers become greater than 2000, they continue “2006, 2012...”. As a result, the average Wayward stably judges, for example, that “ $1008 + 2 = 1012$ ”.

How should we react to our encounter with the Waywards? Wouldn’t it defy credibility to claim that they are making a mistake in almost every elementary calculation they make? And thus more plausible to think that they are not adding at all, or at best, that they ‘add’ differently, and hence that the symbol ‘+’ does not refer to *addition* in the Wayward’s language, but rather a different function, say, $x + ((\lfloor \frac{x}{1000} \rfloor + 1) \times n)$? But that assessment seems to be based on the fact that the Waywards react differently to training than we do, and hence that it does not seem problematic, in the case of the Waywards, to ascribe differences in stable judgement to a difference in meaning. But, if radical in-

33. This is a simplified version of a thought experiment due to Shaw 2022, 119–124. He defends a similar kind of linguistic relativism as I do here.

terpretation begins at home, it should seem equally plausible to ascribe a difference in stable judgement by ourselves to a difference in meaning, too. And why shouldn't it? After all, the Waywards would be likely to say the same thing about us.

The colourblind tribe and the super-detectors Now imagine that long time ago, a faraway tribe was exposed, unbeknownst to them, to dangerously high radiation, leading to a genetic mutation. This mutation was hereditary and caused their offspring to become red-green colourblind. Now further suppose that this change did not lead to any loss of vocabulary. It seems quite plausible that the members of the tribe would now use the words 'red' and 'green' as synonyms, stably judging that things are both red and green, where we say that they are either red or green. It would seem, as with the Waywards, strange to say that these people were now consistently wrong about their colour judgements, at least if the practice had gone on for some time, and its hard to locate the source of the apparent conceptual shift in anything but their change in stable judgements.

This intuition is strengthened, if we imagine that a neighbouring tribe had developed a superhuman ability to detect colour, being able to make clear and stable judgements about more primary colours than we are able to. Here, however, the change lead to an explosion of new vocabulary, not stasis; leading them to make many distinctions in their language that we are unable to make. If we are tempted to say that the colour blind tribe is consistently wrong, we should presumably say that the possibility of such super-detectors means that we could be consistently wrong about *our* colour judgements, too; that it could be that most things that we say are red are in fact not so. After all, they are better, by stipulation, at detecting colour than we are.

Maybe this outcome is welcomed by someone who thinks that the meaning of a given expression is *always* independent of our judgements regarding our use of that expression, the 'platonist' as Wright calls them, but by framing the dialectic around the

cases just considered, the possibility of our practices becoming completely divorced from our dispositions to judgement seems to be the phenomenon to be explained, rather than the default assumption.

After all, these cases do suggest that differences in stable judgement point to a difference in meaning, and while it is not necessary to explain this fact by stable judgements somehow *constituting* meaning, it does suggest that stable judgements across a community of the form ‘*a* is *F*’ co-vary with *a* being in the extension of ‘*F*’ for that community. This covariance thus seems to be an expected *feature* of our actual linguistic practice.

Furthermore, we can readily explain this covariance, given the general contours of the account under consideration, by our judgements adjusting to inputs from the environment through the process of linguistic training. That is to say, if this covariance would not obtain, our practice would be out of touch with the environment in which it is embedded, and our judgements would change to compensate, until co-variance were restored. In this sense, then, what is true or false about a given case is not independent of our dispositions to judgement, because our practice *requires* a certain amount of covariance of truth and meaning facts.

Infallible judgements and worldly facts Whence the worry that linguistic relativism of this sort is a threat to objectivity, then? I believe there are at least two sources of this concern, both of which have a certain degree of *prima facie* plausibility. The first is related to what is referred to in the literature on conventionalism as ‘the argument from worldly fact’ or ‘the master argument’.³⁴ This is the claim that, while conventions might establish the meaning of a given sentence *S*, for example that *S* says that *p*, a convention cannot make it the case that *p*. For instance, it might be a convention among English speakers to use the word ‘vixen’ for all and only female foxes, but it does not follow that the true proposition ‘all vixens are foxes’ is true in virtue of that convention—since,

34. The terminology is due to Topey 2019 and Warren 2015. Among those who defend this argument are Boghossian 1996 and Sider 2003. It is sometimes called “the Lewis-Lewy objection” after its early proponents, C. I. Lewis 1946 and Casimir Lewy 1976 (see Glock 2003 for discussion).

surely, *that* statement is true because every individual vixen is, as it happens, a fox. The proposition ‘All vixens are foxes’ would thus be true even if there were no languages at all.

As applied to the account under consideration, the worry would be that our stable agreement in judgement is somehow supposed to make it the case that *a* is *F*, and that this explains why our judgement in that case is infallible: The account predicts, one might think, that stable dispositions to judgement ground the *truth* that *a* is *F*, or otherwise create or make that *fact*, and thus we cannot be wrong about such cases. Or alternatively, that for such reasons, *a*’s nature as a thing which is *F* cannot be independent of our judgement that *a* is *F*. This would be a devastating objection, but fortunately, it misses the mark.

The general reason is that when we state this objection, we are referring to the predicate *F* in the metalanguage, but the judgement that *a* is ‘*F*’ is partly constitutive of the meaning of *F* in the object language. Suppose for example that we would all have a stable disposition to judge that ‘a stick *S* is one metre long’—in the sense that we’d all utter this sentence in response to the right stimulus. It follows that the *sentence* ‘*S* is one metre long’ would be true in our language, but it does not follow that our agreement has made it the case that *S* has any particular length or otherwise *made* or *created* any facts about *S*—other than the fact that the predicate ‘one metre’ (in our language, not the metalanguage) applies to *S*. We could, for example, have judged that ‘*S* is two metres long’ without any change in *S*’s length. The stable agreement in judgement has only fixed the *meaning* of the relevant expressions and nothing beyond that, even if the corresponding sentence ‘*S* is two metres long’ would now be true.

Similarly, scientists among the color-blind tribe might come up with a scientific theory that explains why an object *a* is what we would call “red or green” and exactly agrees with our corresponding theory, for example that *a* either reflects light with a wavelength of 500–565 nm or 625–750 nm; that their stable judgements about *a* being ‘green or red’

are partly constitutive of the meaning of ‘red’ and ‘green’ in their language is thus no more mysterious than the fact that in Italian, some blue things are “blu” and some are “azzurro”. In a possible world, otherwise the same as ours, where everyone speaks Italian, the sky is blue, even if the inhabitants would stably judge that it is azzuro and not blu. From the Italian perspective, of course, the sky is and would be azzuro, even if everyone spoke English and stably judged that it was blue. This kind of linguistic relativism is quotidian and no threat to objectivity.

We still have to account for the seeming infallibility of stable dispositions to judgement, however. First recall the problem: The meaning of a term ‘F’ is determined by the second-order equilibrium path of the basic constitutive practice of using ‘F’ and this equilibrium is determined in turn by the agent’s stable dispositions to judgement. This means that some dispositions to judgement of the form ‘*a* is *F*’ constitute the meaning of *F* and thus that it looks as if there are infallible judgements: if we would all judge that *a* is *F*, then *a* is *F*.

I maintain that in the light of the previous discussion, this phenomenon is not as strange as it first might seem: we should expect that a difference in stable judgements leads to a change in meaning, and that this is best explained by certain judgements *constituting* meaning—being judgement-dependent in the sense discussed above. After all, stable judgements are judgements that are made in extremely favourable conditions *relative to the linguistic training*: if the agents stably judge that *a* is *F*, it is thereby given that they will not (stably) accept any reasons to judge that *a* is not *F*, simply because that is how they have learnt to judge *F*-ness. For example, if I stably judge that *a* is red and you want to convince me that it is not, you would have to point to some external factor that could have interrupted my judgement—strange lighting, for example. If the judgement is stable, there are *eo ipso* no such factors in the general case, since otherwise I would have had an overriding disposition; so if stable judgements are meaning-constituting, it is not surprising that our judgements about them co-vary with the meaning facts themselves;

that follows directly from the nature of linguistic training and how our dispositions are formed in response to our environment.

It doesn't actually follow from this co-variance, however, that there are any infallible judgements, since any given judgement is always by a particular agent at a particular time, and not by the totality of agents. And for any such judgement it is the case that it could have been different: if *S* stably judges that *a* is *F* and hits the equilibrium, *S*'s judgement is not infallible, since *S* could have judged that *a* is not *F* and thus missed the equilibrium. Any judgement that *S* could make thus could or could not hit the equilibrium, and is therefore not infallible, even if *S*'s judgement is partly constitutive of the meaning of '*F*'. And since no particular judgement is infallible, there are no infallible judgements.

Finally, even if it were *true* on this account that *if* we would all judge that *a* is *F*, *then a* is *F*, it doesn't follow that this conditional should be construed as the judgments in the antecedent causing or grounding the truth of the consequent; that is not how the material conditional should be read. It only says that these facts co-vary—and the preceding discussion shows why this is no surprise: the judgements that *a* is *F* in the antecedent (partly) constitute the *meaning* of *F* in the consequent. Nothing about the nature of *a*, as it exists independently of us and our linguistic practices follows from this fact. The covariance of dispositions to stable judgements and meaning facts is thus no threat to the objectivity of meaning.

IV Concluding remarks

Communitarian accounts of meaning have been widely rejected in the literature on the rule-following paradox, for good reason. In this paper, I've offered a game-theoretic account of semantic content and focused on how it can give replies to some of the most difficult and important objections to communitarian accounts, namely Boghossian's horsey cow case, the community version of the problem of error, as well as what I called the logical problem of error.

More importantly, however, I've tried to show how communitarian accounts of meaning do not necessarily threaten the objectivity of our linguistic practices, neither because such accounts must replace our ordinary notion of truth (the one offered here leaves it quite as is) nor because they result in infallible judgements; for it is possible that stable dispositions to judgment co-vary with meaning facts without there being any infallible judgments.

Finally, this covariance of judgements and meaning facts is benign and what we would expect from our linguistic practices.

V Acknowledgments

I would like to thank Elmar Unnsteinson, Fenner Tanswell and Crispin Wright for comments and discussion on the topic of this paper. I'm also grateful to the anonymous referees for Ergo for their detailed and thoughtful comments.

References

- Andow, James. 2016. "Zebras, Intransigence & Semantic Apocalypse: Problems for Dispositional Metasemantics." *Philosophia* 44 (1): 53–62.
- Aumann, Robert. 1974. "Subjectivity and Correlation in Randomized Strategies." *Journal of Mathematical Economics* 1:67–96.
- . 1987. "Correlated Equilibrium as an Expression of Bayesian Rationality." *Econometrica* 55 (1): 1–18.
- Berg, Ásgeir. 2022. "Rules as Constitutive Practices Defined by Correlated Equilibria." *Inquiry* 65:1–35.
- . Forthcoming. "Problems for 'Standard' Dispositionalist Accounts of Semantic Content." *Synthese*.
- Bloor, David. 1997. *Wittgenstein, Rules and Institutions*. London: Routledge.
- Boghossian, Paul. 1989. "The Rule-Following Considerations." *Mind* 98 (392): 507–549.
- . 1996. "Analyticity Reconsidered." *Noûs* 30 (3): 360–391.
- . 2008. "Epistemic Rules." *The Journal of Philosophy* 105 (9): 472–500.
- Gintis, Herbert. 2009. *The Bounds of Reason*. Game Theory and the Unification of the Behavioural Sciences. Princeton: Princeton University Press.
- Glock, Hans-Johann. 2003. "The Linguistic Doctrine Revisited." *Grazer Philosophische Studien* 66 (1): 143.
- Glüer, Kathrin, and Peter Pagin. 1999. "Rules of Meaning and Practical Reasoning." *Synthese* 117 (2): 207–227.
- Guardo, Andrea. 2022. "Yet Another Victim of Kripkenstein's Monster: Dispositions, Meaning, and Privilege." *Ergo* 8 (0).

- Hattiangadi, Anandi. 2006. "Is Meaning Normative?" *Mind & Language* 21 (2): 220–240.
- . 2007. *Oughts and Thoughts: Rule-Following and the Normativity of Content*. Oxford: Oxford University Press.
- Haugeland, John. 1998. *Having Thought: Essays in the Metaphysics of Mind*. Cambridge: Harvard University Press.
- Johnson, Michael, and Jennifer Nado. 2014. "Moderate Intuitionism: A Metasemantic Account." In *Intuitions*, edited by A. Booth & D. Rowbottom. Oxford: Oxford University Press.
- . 2016. "Intuitions and the Theory of Reference." In *Advances in Experimental Philosophy and Philosophical Methodology*, edited by J. Nado. London: Bloomsbury.
- . 2017. "Actual vs. Counterfactual Metasemantics: A reply to Andow." *Philosophia* 2 (45): 717–734.
- Kripke, Saul. 1982. *Wittgenstein on Rules and Private Language*. Harvard University Press.
- Kusch, Martin. 2002. *Knowledge by Agreement: The Programme of Communitarian Epistemology*. Oxford: Oxford University Press.
- . 2006. *A Sceptical Guide to Meaning and Rules: Defending Kripke's Wittgenstein*. Chesham: Acumen.
- Lewis, C. I. 1946. *An Analysis of Knowledge and Valuation*. La Salle: The Open Court Publishing Company.
- Lewis, David. 1969. *Convention: A Philosophical Study*. Cambridge: Harvard University Press.

- Lewis, David. 1974. "Radical interpretation." *Synthese* 27 (3-4): 331-344.
- Lewy, Casimir. 1976. *Meaning and Modality*. Cambridge: Cambridge University Press.
- Millar, Alan. 2002. "The Normativity of Meaning." *Royal Institute of Philosophy Supplement* 51:57-73.
- Miller, Alexander, and Olivia Sultanescu. 2022. "Rule-Following and Intentionality." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University.
- Nyquist, Mons. 2020. "On Complete Information Dispositionalism." *Philosophia* 48 (5): 1915-1938.
- Peacocke, Christopher. 1990. "Content and Norms in a Natural World." In *Information, Semantics, and Epistemology*, edited by Enrique Villanueva. Blackwell.
- Rawls, John. 1955. "Two Concepts of Rules." *The Philosophical Review* 64 (1): 3-32.
- Reiland, Indrek. 2020. "Constitutive Rules: Games, Language, and Assertion." *Philosophy and Phenomenological Research* 100 (1): 136-159.
- Searle, John. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Sellars, Wilfrid. 1997. *Empiricism and the Philosophy of Mind*. Cambridge: Harvard University Press.
- Shaw, James. 2022. *Wittgenstein on Rules: Justification, Grammar, and Agreement*. Oxford: Oxford University Press.
- Sider, Theodore. 2003. "Reductive Theories of Modality." In *The Oxford Handbook of Metaphysics*, edited by Michael J. Loux and Dean W. Zimmerman, 180-208. Oxford: Oxford University Press.

- Topey, Brett. 2019. "Linguistic Convention and Worldly Fact: Prospects for a Naturalist Theory of the a Priori." *Philosophical Studies* 176 (7): 1725–1752.
- Turing, Alan. 1936. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* 42 (1): 230–265.
- Vanderschraaf, Peter. 1995. "Convention as Correlated Equilibrium." *Erkenntnis* 42 (1): 65–87.
- . 1998. "Knowledge, Equilibrium and Convention." *Erkenntnis* 49 (3): 337–369.
- . 2018. *Strategic Justice: Convention and Problems of Balancing Divergent Interests*. Oxford: Oxford University Press.
- Verheggen, Claudine. 2003. "Wittgenstein's Rule-Following Paradox and the Objectivity of Meaning." *Philosophical Investigations* 26 (4): 285–310.
- . 2011. "Semantic Normativity and Naturalism." *Logique Et Analyse* 54 (216): 553–567.
- Warren, Jared. 2015. "The Possibility of Truth by Convention." *The Philosophical Quarterly* 65 (258): 84–93.
- . 2020. "Killing Kripkenstein's Monster." *Notus* 54 (2): 257–289.
- Wikforss, Åsa. 2001. "Semantic Normativity." *Philosophical Studies* 102 (2): 203–226.
- Wilson, George. 1994. "Kripke on Wittgenstein and Normativity." *Midwest Studies In Philosophy* 19 (1): 366–390.
- . 1998. "Semantic Realism and Kripke's Wittgenstein." *Philosophy and Phenomenological Research* 58 (1): 99–122.
- Wright, Crispin. 1980. *Wittgenstein on the Foundations of Mathematics*. Cambridge: Harvard University Press.

- Wright, Crispin. 1981. "Rule-Following, Objectivity and the Theory of Meaning." In *Wittgenstein: To Follow A Rule*, edited by Steven H. Holtzman and Christopher M. Leich. Routledge.
- . 2001. "On Making Up One's Mind: Wittgenstein on Intention." In *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*, 116–142. Cambridge: Harvard University Press.
- . 2002. "Meaning and Intention as Judgement Dependent." In *Rule-Following and Meaning*, edited by Alexander Miller and Crispin Wright. Montreal & Kingston: McGill-Queen's University Press.
- . 2007. "Rule-Following Without Reasons: Wittgenstein's Quietism and the Constitutive Question." *Ratio* 20 (4): 481–502.