

Real Normativity

James Alexander Cross
UCL
PhD, Department of Philosophy

I, James Alexander Cross, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Saul Kripke's *Wittgenstein on Rules and Private Language* (1982) marked the beginning in earnest of the continuing debate about the connection of rule-following and meaning to normativity. This thesis continues my previous project (Cross 2013) on this topic.

That first instalment assessed a dilemma identified by John McDowell (1984) between two arguably unappealing alternatives: embracing something like Crispin Wright's (1980) global anti-realism, and facing up to the notorious 'sceptical paradox' Kripke found in Wittgenstein's *Philosophical Investigations* (Wittgenstein 1953). Where previously I focused on Wright, I will now focus on Kripke.

I begin with an outline of Kripke's problem and locate it within the context of my previous work, in turn locating the normativity of meaning within the challenge. I then critically assess the relative merits of several accounts of meaning-normativity that present as straight solutions to Kripke's paradox, before responding to arguments that the paradox can and should be resolved by rejecting the critical premiss that meaning is essentially normative.

I ultimately conclude that while meaning-normativists can plausibly resist this and similar assaults, the whole topic is in fact a red herring with respect to Kripke's paradox. I argue that despite his attestations to the contrary, the paradox turns on doubts about our ability to *know* our own intentional content, and hence that it should be resolved by a suitable account of epistemology.

Contents

Contents	6
List of Figures	7
1 Introduction: Kripke's Paradox	11
1.1 Chapter Breakdown	11
1.2 Kripke's Sceptic	12
1.3 Kripke's Commitments	13
1.4 Finding the Question	16
1.5 Thoughts from the Literature	19
1.6 Conscious Engagement	20
1.7 Summary	22
2 Normativity in the McDowell-Wright Debate	23
2.1 Normativity in McDowell 1984	23
2.2 McDowell's Dilemma	26
2.3 Normativity and the Illusion Argument	27
3 Externalist Normativity	29
3.1 Rules with Biological Purpose	29
3.2 Millikan and Kripke	30
3.3 The Justification Question	33
3.4 Meaning without Intention	34
3.5 What did Kripke 'mean'?	36
3.6 Conclusion	39
4 Non-intentionalism	41
4.1 Intentionalism	41
4.2 Intentionalist Practices	43
4.3 Conforming and Following	46
4.4 Non-intentionalism	50
5 Primitive Normativity	55
5.1 Dispositions Reconsidered	55
5.2 Some Key Commitments	60
5.3 Ginsborg's Kripke	61
5.4 Meaning in the Slab Game	63
5.5 Primitive Normativity and Anti-Realism	66
5.6 Haddock's Kripke	68

5.7	Haddock's Critique	70
5.8	Conclusion	71
6	Normativity, Correctness and Truth	73
6.1	Defining the Normativity of Meaning	74
6.2	Hattiangadi's 'Ought Implies Can' Objection	74
6.3	Normativity from Correctness	77
6.4	The Simple Argument	78
6.5	Analysis	81
6.6	In Defence of the Simple Argument	82
6.7	Conclusion	85
7	Defusing Kripke's Paradox	89
7.1	Intention without Knowledge Thereof	90
7.2	Understanding without Knowledge Thereof	93
7.3	The Normativity Question	95
7.4	The Constitutive Question	97
8	Conclusion	103
	Bibliography	105

List of Figures

5.1	Ginsborg's discussion of appropriateness	59
-----	--	----

Acknowledgements

Many thanks are due to my supervisors José Zalabardo and Paul Snowdon, and to the AHRC for generously funding this project. I would also like to thank the attendees and respondents of the 2013 and 2014 UCL Philosophy Graduate Conferences, whose feedback on early-days content from chapters 3,6 and 7 were both helpful and encouraging. As my MPhilStud thesis informed this project, I should repeat my thanks to all involved in that work — including to my reviewers, whose commentary inspired me to continue working on this topic. And of course, enormous thanks are due to my wife and family for their love and support during this project, without which it would not have been possible.

Chapter 1

Introduction: Kripke's Paradox

In 1984, Saul Kripke published *Wittgenstein on Rules and Private Language*, an 'elementary exposition' of Kripke's take on Wittgenstein's remarks on rule-following in his *Philosophical Investigations* (1953). While questions were raised about the interpretive accuracy of Kripke's gloss,¹ his account of Wittgenstein as presenting a sceptical paradox about the possibility of meaning has proved hugely influential. Kripke's discussion is ostensibly the origin of the debate about the 'normativity of meaning': the idea that the meaning of words is essentially bound up with certain prescriptions to action. Defenders of this very broad idea of 'meaning-normativism', who I refer to as 'normativists', are typically interested in the topic as a means to solving Kripke's sceptical paradox. Most commentators on the normativity of meaning thus have something to say about Kripke, and likewise, many commentators on Kripke's paradox will have something to say about the normativity of meaning. Speaking for my own part, my interest in the normativity of meaning also stems from the seeming insolubility of Kripke's paradox, and so I take the Kripke-Wittgenstein literature as the frame for this thesis. What we are looking for is an account of meaning-normativity that is at once palatable, defensible and a satisfying answer to Kripke's sceptic.

1.1 Chapter Breakdown

The present chapter focuses on providing an outline of Kripke's discussion and the sceptical challenge, and in the next chapter I tie the present debate to my previous work by looking more closely at the involvement of normativity in the McDowell-Wright debate on rule-following and objectivity.

The remaining dissertation weighs a series of possible responses to Kripke's discussion, attempting by various means to find an account that fits before finally settling on a position in chapter 7. This begins in chapter 3, where I explore the idea that meaning-normativity can be described in a paradox-dodging way by removing the dependency on anything 'internal' to the norm-bound agent. I note that this presents a quick defeat strategy for the paradox by locating meaning-facts out of harm's way, but conclude that it is unsatisfactory for precisely this reason: this approach fails to account for our ability to follow rules at a personal level, which is a major theme of Kripke's challenge.

¹Cf. for example (Blackburn 1984), (McDowell 1984).

Chapter 4 follows on, going into more depth on the involvement of intentions and mentality in rule-following and meaning determination. Here I push the limits of some critical assumptions about successful rule-following, weighing the plausibility of rejecting them ahead of later discussion.

Chapter 5 explores and discusses Hannah Ginsborg's 'primitive normativity' in the context of a response to Kripke.

Chapter 6 goes into detail on some critiques of the idea that meaning is essentially normative and offers responses.

Finally, in chapter 7 I draw on these considerations to assemble my own take on the sceptical challenge and its solution. I argue that, contrary to appearances, Kripke's controversial assumptions about the normativity of meaning and intention are not sufficient to generate the paradox outlined. This, I conclude, is because the sceptical paradox does in fact rest entirely on epistemological concerns.

1.2 Kripke's Sceptic

A significant obstacle, and one no doubt familiar to Wittgenstein scholars otherwise, is that Kripke's text is not as clear as it could be. This means that not all commentators have interpreted the task it presents in the same way, and as I progress in the thesis towards a way through the problem I will pick up on various nuances of the challenge that have emerged in the surrounding literature. However, to begin I will need to paint some kind of picture of Kripke's discussion, the paradox invoked and the challenges derived from it — at least, insofar as they relate to the normativity of meaning.

Kripke imagines a sceptic who asks me about the meaning of the word 'plus' and the plus sign, '+'. Asked "What is $68 + 57$?", and having never before performed this particular calculation, I confidently answer '125'. It seems that I answer as I do because I have previously grasped how to add — where 'how to add' is something like a rule or algorithm, certain instructions to be deployed in contexts of addition. Furthermore, I know what 'plus' means: I have grasped a rule for understanding that 'plus' generally refers to the addition function. Intuitively, both cases allow me to 'go on doing the same thing' with the word 'plus' as I have always done previously, even faced with addition problems I have never previously encountered. I can apply the rule to new cases, and if I successfully follow it, then my answer will be correct as determined by the rule.

The thought here is not one specifically about addition, but about my being able to give an answer that is mandated by something rigid. If I had grasped some other rule for using 'plus' then that rule might tell me to give a quite different answer: imagine that in some language, "What is $68 + 57$?" means the same as the English "What is the weather like?". If I follow the relevant rules for the language, then I should not give a numerical answer at all: but there are nonetheless correct ways of answering the question relative to that rule.

Kripke's sceptic now hypothesises that whenever I used 'plus' in the past I really meant 'quus', where x 'quus' y is equal to the sum of x and y if both are less than 57, otherwise it equals 5.

$$x \text{ quus } y = \begin{cases} x + y & \text{if } x, y < 57 \\ 5 & \text{otherwise} \end{cases}$$

Imagine now that I've never performed an addition with numbers as large as 57 before: then my past usage of 'plus' is consistent with my having meant either *quus* or *plus* by it. If the sceptic is right and I did always mean *quus* by 'plus', then my answer to 'What is $68 + 57$?' ought to be '5', not '125'.

Kripke's challenge is to disprove this hypothesis. However, Kripke argues that there are simply "no facts about me", mental or otherwise, that can be cited in favour of *either* the sceptic's hypothesis or my own hypothesis that I always meant *plus*. And of course, if nothing determines that I meant anything in particular by 'plus', then there is just no fact about what I meant. And it is easy to see from here that this applies just as well to all present uses of language: and not only to me, but to all language users. Hence the paradoxical conclusion "that all language is meaningless".

The best candidate for a fact about which rule I followed would be the fact that I also followed some further 'rule for following rules' when learning how to use 'plus'. It is this further rule that determines that I have always followed the *plus* rule, and thus have always meant *plus* by 'plus'. But the sceptic can always ask: "how is it that you were following one such further rule rather than another? For you only applied *that* rule in finitely many cases too: perhaps now the rule would prescribe the answer '5'..." and so on for any rule I care to imagine.

The problem thus seems to be inherent to the notion of rule-following. Any finite set of actions accords with infinitely many rules, yet the only thing that can tell us which rule applies seems to be a further rule — and then how are we to establish that *this* rule applies rather than any of infinitely many others?

1.3 Kripke's Commitments

Kripke presents his reasoning as flowing straightforwardly from intuitive premisses, to the conclusion that any competing hypothesis about what answer I should give, and any competing hypothesis about what I mean by 'plus' (or '+'), will be on equal footing with my own claim that I mean *plus* by 'plus' and that the question as I understand it accordingly merits the answer "125". The paradoxical conclusion is then that I do not mean one thing rather than another by 'plus', since there seems to be no plausible candidate for the fact of my meaning *plus* (or any thing else) by 'plus', and my meaning something in particular by this seems essential to determining how I ought to answer the question "What is $68 + 57$?".

However, in reality, there are a *lot* of these 'intuitive premisses' and not all of them are as intuitive as others. I will attempt to extract some here: others will emerge later on. I offer the following observations in a preliminary capacity only, as subtleties of phrasing are enough to suggest rather different sets of assumption for Kripke's challenge — this is also why I have kept relatively close to the text so far.

Extension

First, and without worrying too much about how they enter into the discussion, we should note that Kripke seems to treat *rules* as concepts of a sort, with an extension:

Extension Although I myself have computed only finitely many sums in the past, the rule determines my answer for indefinitely many new sums that I have never previously considered. (Kripke 1982, p. 7)

The rule Kripke has in mind here is something like a rule for meaning *plus* by 'plus'. The 'extension' of this rule is the set of all usage of 'plus' where 'plus' means *plus*. Whether the extension is *determined* by this rule is a further question; it may be that a given use of 'plus' means *plus* because of this rule — that accordance with this rule is what the meaning consists in — or else that the rule merely describes the circumstances where 'plus' means *plus*, this meaning being otherwise determined. I suspect that Kripke thinks the former, but I am not sure it makes much difference to his argument. What is important is that Kripke seems to take *following* the rule to make it the case that I myself mean what I do. So if I mean *plus* by 'plus', the background supposition during much of the discussion is that I must be following the rule for meaning *plus* by 'plus'; deliberately aligning my usage with the extension of that rule.

But this is to put things in a more complex way than necessary. We have been talking in terms of a rule for meaning x by y , but the more obvious kind of rule at work here is the arithmetical 'plus' rule, the addition function itself that takes (here) two arguments and outputs their sum. Kripke plays the thought that I appeal to this function in answering the example question, I apply the function to the question, and that is how I arrive at the answer "125". If I always apply this function when using 'plus', then that seems to entail my meaning addition by plus, and so to warrant my confidence in answering "125". The question being pressed is simply whether that *is* the function I always apply, given that the outputs of the quaddition function for all my uses of 'plus' up to his point would presumably have been identical.

Arithmetical and Metalinguistic Correctness

This brings out a second important point, that there is a firm distinction drawn between two senses in which "125" might be the correct answer to the example question:

It is correct both in the arithmetical sense that 125 is the sum of 68 and 57, and in the metalinguistic sense that 'plus', as I intended to use that word in the past, denoted a function which, when applied to the numbers I called '68' and '57', yields the value 125. (Kripke 1982, p. 8)

He identifies the 'metalinguistic' kind as the more immediately problematic, framing the subsequent 'sceptical paradox' in terms of accordance with one's past intentions.

Since this is *ex hypothesi* my first ever attempt to calculate $68 + 57$, adherence to my previous intentions using 'plus' means looking at my previous uses of the word in different cases and trying to 'go on in the same way'. The challenge is to identify what 'the same way' might be. Supposing that there is some determinate answer for any computation 'x plus y', this being the 'arithmetically correct'

answer, and that there is equally some determinate answer that fits with my previous intentions in using the word 'plus', the 'metalinguistically correct' answer, how are we to establish that the two should coincide? The limited range of my past usage, viewed from outside, is consistent with my always having understood 'plus' to mean 'quus'. If no-one else can tell which I meant, can I myself? If I am to defeat the sceptic's claim that I always meant 'quus' before, "there must be some fact about my past usage that can be cited to refute it" (Kripke 1982, p. 9).

Internalistic Justification and Normativity

The sceptic invites a demonstration that I previously intended to mean addition rather than quaddition by 'plus', meaning that if I want (intend) to carry on using 'plus' in the same way, I should now answer '125' rather than '5'. The problem is, the facts about my usage and intentions seem insufficient to determine, in a way Kripke deems satisfactory, which function I previously meant, or indeed, which function I now mean, by the word 'plus'. If I meant *addition*, then we suppose I would have followed the rule for addition whenever asked to perform an addition, and consequently, it would have been a fact about me that I was following that rule rather than any other, and a fact about me that I meant *addition* rather than *quaddition*. But if I am to act on my intention to carry on in the same way, then it is not enough that there be some externalistic fact as to which rule I followed and which function I meant: I must be able to access such a fact myself, so that in this new case I will be *justified* in answering '125'. As Kripke puts it:

Ordinarily, I suppose that, in computing '68 + 57' as I do, I do not simply make an unjustified leap in the dark. I follow directions I previously gave myself that uniquely determine that in this new instance I should say '125'. (Kripke 1982, p. 10)

So now there are two new players: justification, and normativity. With the former the challenge apparently takes an epistemological twist, as we look for facts accessible to me (or any other agent, apparently) that would lend sufficient epistemic virtue to my claim to know what I mean now, and have meant previously, by 'plus' — as well as the corollary claim as to which rule I was following.

The latter, normative element of the challenge seems to be part of this justification: Kripke assumes that adherence to my previous intentions means that there is in this case a unique thing that I *should* say in answering '68 + 57', and it would seem that this 'should' is what justifies my answer. I look back, as it were, to my past usage of 'plus', and see that going on in the same way means sticking to the path, following the directions I gave myself (whatever form they took), and that I *should* consequently say one thing now rather than another or else fail to continue on in the same way and heed those instructions. But it is mysterious what those 'directions' might be, as they were *ex hypothesi* not explicit instructions to myself to answer '125' on this occasion.

So it seems the challenge might involve some or all of the following:

- Determining what rule, if any, I was following whenever I tried to add before;
- Determining what I meant by 'plus' whenever I used the word before;

- Determining what I should do now, given my past usage of 'plus';
- Determining what 'directions' I previously gave myself;
- Finding a justification for my answer '125';
- Justifying my 'certainty' that my answer should be '125'.

Answers to Kripke typically attempt to fulfil at least these tasks. But of course, interpreting even what some of these tasks demand has proven to be controversial, as will emerge in subsequent chapters.

1.4 Finding the Question

There is some divergence within the literature when it comes to interpreting precisely what kind of response Kripke's sceptic requires. My intention here is to document some of the different interpretations that have been examined, as this will help us to clarify in turn what sort of thing is meant by the 'normativity of meaning'.

Although Kripke's initial presentation of the paradox avoids pinning down the precise nature of the problem, he does elaborate. First, he says the sceptic's challenge "takes two forms":

First, [the skeptic] questions whether there is any fact that I meant plus, not quus, that will answer his sceptical challenge. Second, he questions whether I have any reason to be so confident that now I should answer '125' rather than '5'. The two forms of the challenge are related. I am confident that I should answer '125' because I am confident that this answer also accords with what I *meant*. (Kripke 1982, p. 11)

Now it's slightly confusing that the first of these refers to 'answering the challenge' as the challenge itself, but for now let us take it simply as a call for demonstration that there is some fact that I meant plus, not quus. Kripke later indicates that the absence of such a fact about what I meant in the past would mean the absence of such a fact in the present (1982, p. 13), this being the full blown form of the 'sceptical paradox', so the challenge clearly applies to facts about what I mean now as well as what I have previously meant. If we extend this to cases of meaning in general, we get:

Form 1 Show that there is some fact that I mean one thing rather than another when I make a particular use of some expression.

The second form of the challenge seems to relate to Kripke's talk of justification. The word 'should' here has conspicuously normative connotations, as does 'reason', indicating that the question of my confidence is in some sense a question of warrant, justification, prescription, or some such. This is confirmed by Kripke's subsequent statement of adequacy conditions for any putative answer:

An answer to the sceptic must satisfy two conditions. First, it must give an account of what fact it is (about my mental state) that constitutes my meaning plus, not quus. But further, there is a condition that any putative candidate for such a fact must satisfy. It must, in some sense, show how I am *justified* in giving the answer '125' to '68+57'. (Kripke 1982, p. 11, my emphasis)

That is, something about the nature of the fact in Form 1 must be such as to give me a justifying reason to answer '125'. This suggests that Kripke's talk of 'confidence' is really talk about justification (warrant, prescription, etc.). Hence, it seems fair to interpret the second challenge as follows:

Form 2 Show that I am justified in using some expression in a particular way.

Kripke adds that the 'directions' as to what I *should* answer (in the '68 + 57' case), i.e. the substance demanded by Form 2, must be 'contained' in any fact that constitutes an answer to Challenge 1 (1982, p. 11).

Kripke proceeds to examine some potential answers, before concluding that the sceptic's challenge cannot be met without fundamental revisions to our conceptions of meaning and truth. He thus favours an expensive 'sceptical solution', which involves the rejection of truth-conditional semantics in favour of a justification-based alternative, comparable in some ways to Crispin Wright's anti-realist reading of Wittgenstein.²

Dispositionalism and Meaning-Normativity

Kripke effectively considers two kinds of alternatives before resorting to the 'sceptical solution': the anti-reductionist option, which holds that the facts about what I mean are determined by my being in some *sui generis* state of meaning such-and-such; and the dispositionalist option, which holds that the facts about what I mean are in some sense determined by what I am disposed to say under certain conditions.³ The idea that meaning is normative emerges as a keystone in his rejection of the latter. He says, in a famous passage:

Suppose I do mean addition by '+'. What is the relation of this supposition to the question how I will respond to the problem '68 + 57'? The dispositionalist gives a *descriptive* account of this relation: if '+' meant addition, then I will answer '125'. But this is not the proper account of the relation, which is *normative* not descriptive. The point is *not* that, if I meant addition by '+', I *will* answer '125', but that, if I intend to accord with my past meaning of '+', I *should* answer '125'. Computational error, finiteness of my capacity, and other disturbing factors may lead me not to be *disposed* to respond as I *should*, but if so, I have not

²See Wright 1980 for his contemporaneous position. McDowell (1984) compares Kripke's solution to Wright's account of rule-following, as I have discussed previously (Cross 2013). For developments in Wright's position since 1980, see Wright 1989b, 1992, 2001, 2007. See also chapter 2 below.

³For more on the contrast see my discussion of Ginsborg's primitive normativity in chapter 5 below.

acted in accordance with my intentions. The relation of meaning and intention to future action is *normative*, not *descriptive*. (Kripke 1982, p. 37)

So the idea here is that the dispositionalist is somehow ruled out from meeting the challenge basically from the assumption here expressed that we are not interested so much in the facts about what I will do under any future circumstances so much as what I *should* do. It is clear from the context that something other than my disposition to answer as I do is supposed to provide this 'should' — and it seems likely that it is the 'instructions I previously gave myself' that are relevant here. What we need to uncover, it seems, is a) which meaning-rule I previously instructed myself to follow in this situation, and b) what that meaning-rule says I ought to do now, with respect to that past self-instruction.

The Rule-Following Machine

Kripke (1982, pp. 32–35) takes time to consider an important reply from the dispositional side, which he attributes in this instance to Dummett (1959, p. 331), to the effect that since we can programme a machine to mechanically yield the correct output of the plus function, there should be no difficulty in us doing so ourselves. If a machine is capable of 'going on in the same way', why not a human — who, after all, must surely have programmed the machine to go on in that way in the first place?

Kripke begins by trying to discern what precisely is meant here, and struggles for a while with the notion that the objector is assuming we might “draw up a program to embody our intentions” such that the machine becomes “*definitive* of my own intentions” (1982, p. 33). With the present ubiquity of computers, this is perhaps a little less difficult to envisage: I can write a computer program that, suitably coded, will output just what I intend in a way that seems 'rigid'. For example, executing the following JavaScript code would generate two random integers less than 100 and output their sum:

```
var n = Math.floor(Math.random() * 101);
var m = Math.floor(Math.random() * 101);
console.log(n + m);
```

Put the code in a suitable loop, and the computer will go on adding numbers together forever — or rather, until it breaks. Even if we imagine a computer system that is kept operational indefinitely,⁴ Kripke notes that two obstacles remain (1982, pp. 34–35). First, only a finite number of inputs can ever be processed by the machine, yet there are an infinite number of possible applications of 'plus'. Thus for any given correctly computed addition, it could always be the *next* one that reveals the machine to have been interpreting 'plus' as something divergent like 'quus'. This is the 'finitude' objection.

Second, and more troublingly, the machine is always at risk of malfunctioning — in our terms: the program may always be improperly executed for some particular loop. But this raises the question: what makes it the case that a malfunction

⁴The tendency of server administrators to replace large parts of their machines while still in operation, such that many processes are running continually for years on end, renders this possibility similarly conceivable. The limiting factor here is the continued existence of the universe.

occurs? If it is possible that the machine may malfunction, then there is some standard by which it can be said to malfunction. What is that standard? And then we face the question of ‘accordance with past intentions’ once again. The framing of this problem is complete when we consider the case where the system *does not* malfunction: for even then, it ‘goes on in the same way’ only on some interpretation or other. Whether any given output is a malfunction or not is only determined with reference to the intentions of the programmer, and keeping hold of those is precisely the problem at hand.

1.5 Thoughts from the Literature

I turn now to a few developments in the interpretive literature on rule-following and the Kripke paradox that are worth considering as we go forward.

Permissive and Prescriptive Rules

Boghossian (2012, §2) proposes a distinction between two kinds of rules. On the one hand there are *prescriptive* rules, those with imperatival content as per ‘if the light is red, stop the car’, dictating what *should* be done in a particular context; on the other are *permissive* rules, those that dictate what *may* be done in some context. Of the latter, Boghossian suggests (p. 29) a chess rule “If the configuration is C, you *may* castle” as an example.

This contrast is familiar, but it is interesting to present it as a substantive dichotomy. Boghossian supposes that when (Kripke 1982) talks about rules, he is really talking about the prescriptive kind more or less exclusively, and that consequently Kripke’s vision of the ‘rule-following problem’ is not necessarily definitive. I have some doubts about this attribution, but will remark in the first instance that it is controversial to draw such a firm distinction at all. For prescriptions and permissions are in an important sense interdefinable: as per a fairly mundane deontic logic, that one must ϕ means simply that one may not $\neg\phi$. Formally:

$$\text{Must}(\phi) \leftrightarrow \neg\text{May}(\neg\phi)$$

So the point cannot be simply that some rules are prescriptive and the rest are permissive, as this is essentially a matter of formulation. ‘If the light is red, stop the car’ is prescriptive, but can be expressed in terms of what one is *not permitted* to do; Boghossian’s castling rule describes a permission, but can be expressed in terms of what it is prescribed that one *must not* do. What, then, is this distinction supposed to amount to?

In fact, being able to explain permission in terms of prescriptions is not quite the same as actually converting one thing into the other. Suppose that ‘*S* must ϕ ’ means that *S* has some ‘reason’ to ϕ . Then its negation means that *S* has no such reason. In these terms, ‘*S* may ϕ ’ then means that *S* has no reason not to ϕ —another negative existential claim. Naturally, we can substitute talk of reasons here for talk of whatever philosophically significant thing we think is involved in being subject to prescription over being permitted to do something. And that is surely the point: we are naturally inclined to think that prescriptions and permissions are different in some important way, and it is that difference — whatever it is — that Boghossian is trying to indicate with this distinction.

ME and MD Norms

One major contribution that Glüer and Wikforss make to the debate is the introduction of a means of distinguishing two varieties of normativism.⁵ The two varieties are distinguished by how they answer a question of metaphysical priority: which comes first, meaning or normativity?

Meaning-engendered (ME) norms are norms that hold because of facts about the relationship between meaning and correctness conditions. An ME normativist consequently views meaning as having metaphysical priority. Meaning-determining (MD) norms, on the other hand, are norms that account for and determine meaning. An MD normativist thus considers meaning to be a product of MD normativity, which has priority.

The distinction is useful because, like most good distinctions, it helps us to pin down what a slogan like 'meaning is normative' might be supposed to mean. Do we mean: there is something called meaning, and its existence confers prescriptions? Or alternatively: meaning exists, and words mean what they do, because we are bound by certain norms? These accord with the two positions described above.

However, the distinction must be wielded with caution, since ME and MD normativism need not be mutually exclusive positions. Perhaps we think the fact that 'snow' means *snow*, means that one ought to use 'snow' in only a limited range of ways. The meaning-fact engenders a norm of use, and this is an ME normativist position. But then perhaps we also think that this metalinguistic element of the theory of meaning, "snow' means *snow*", is itself true because it is governed by some more fundamental norms of meaning. This would be a kind of MD normativism.

On the other hand, it might then be objected that the whole thing then just becomes MD normativism. If "snow' means *snow*' is true by virtue of an MD norm, then it just follows that 'snow' means *snow* by virtue of an MD norm. But it could be the case that we conceived different ME-oughts as deriving from 'snow' from those prescribed by the MD norm that governs the meaning of "snow' means *snow*". This would still make the MD norm more fundamental, but it seems fair to say that the ME 'ought' here is directly caused only by the fact that 'snow' means *snow*. It is at most indirectly caused by there being certain more fundamental norms in place.

The idea of a heterodox normativism is not obviously unpalatable. It is not impossible that in some cases meaning is determined by norms, while in others it confers prescriptions itself. For example, let me define the new word 'horque' as meaning *truffle*. Anyone wishing to use the word 'horque' ought to use it in much the same way as they use the word 'truffle'. This is clearly a direct consequence of the fact we brought about that 'horque' means *truffle*: hence this is an example of ME normativism. However, perhaps 'truffle' means *truffle* because there are certain norms in force that direct us how to use that word. This would be an example of MD normativism.

1.6 Conscious Engagement

José Zalabardo's (1997) influential interpretation of Kripke's assault on dispositionalism argues that the invocation of the normativity of meaning to this end has

⁵Cf. for example Glüer and Wikforss 2010 and Glüer and Wikforss 2009a, p. 33

been widely misread. According to the ‘Standard Normativity Argument,’ which Zalabardo identifies as the misreading of Kripke in question,⁶ a dispositional account of meaning-facts is inadequate because such an account of what I will say under certain conditions determines nothing about what I *ought* to say. The challenge is that “dispositional facts are descriptive,” and since “descriptive facts cannot license evaluative claims,” we cannot account for the ‘ought’ simply by appeal to dispositions (Zalabardo 1997, pp. 469-470).

But this argument depends upon the substantive unjustified assumption that it is impossible to derive normative from non-normative (i.e. descriptive) facts. As Zalabardo observes (1997, §III), there is no reason in principle why a dispositionalist could not reply that how I ought to use some expression is just how I would, in fact, use it under ideal conditions. Perhaps there is some question as to what those ideal conditions, in fact, are: but to insist that this is indeterminate or otherwise inadmissible information is a substantive claim in itself, and is unsupported by the argument so formulated. Thus, the Standard Normativity Argument does not provide the definitive knock-down that was promised.

Zalabardo argues (1997, §IV) that attributions of the Standard Normativity Argument to Kripke are false, who in fact makes a quite different argument. The seat of this different line is the idea that the fact of what I previously meant by some expression *justifies* my present usage of it. If any fact that would satisfy Form 1 of the sceptic’s challenge must also answer Form 2, then in the addition case, for example, any such fact must be justificatory of my answer ‘125.’ We are not interested in the move from description to prescription, as it were, but in the possibility of justifying my actions by appealing to my dispositions.

The proposed form of this justification is as an instruction, as per Kripke’s insistence that a meaning-determining fact “should *tell* me what I ought to do in each new instance” (Kripke 1982, p. 24). For Zalabardo’s specific discussion of facts as to whether some object satisfies a given predicate,⁷ the requirement is that such facts “ground claims about how I should apply predicates by telling me which applications to endorse” (Zalabardo 1997, p. 284). The challenge against dispositionalism is simply that facts about my disposition to apply some predicate to certain objects at certain times cannot underwrite (justify) my endorsement of those applications. That I am wont to ϕ is not, we tend to think, a *reason* for me to ϕ : even when we say of some particular action that it ‘justifies itself,’ this is not quite what we mean. On the other hand, we do sometimes think about what we would do under ideal conditions, in order to justify using some expression — think of trying to identify a colour swatch in poor light. Might that be a way out for the dispositionalist?

Zalabardo’s Kripke thinks not. While in cases of deliberation perhaps something like this does happen, it’s not an option in the kind of fundamental case under discussion, that of simply of carrying on meaning the same thing by some expression, or doing right by the intentions I previously had as part of my understanding of it. I do not have to deliberate when I answer the sceptic’s question with ‘125’; I do not *hypothesise* that I ought to answer this way.⁸

⁶He attributes this interpretation particularly to Boghossian (1989) and Fodor (1990), agreeing with the former that Blackburn (1984), Wright (1984) and possibly McDowell (1984) also read Kripke this way.

⁷This is stipulated to be the case iff that object exemplifies a property to which the predicate in question stands in some particular relation (Zalabardo 1997, p. 468).

⁸At least, not based on concern as to whether I am using the word ‘plus’ correctly; of course, I

Taking a generalised view of the original scenario, Zalabardo provides the following analysis. Let q be a question of the kind that interests Kripke's sceptic (like 'What is $68 + 57?$ '), let F be the set of facts about how q should be answered, and let a be some possible procedure for answering q . Now whether a is a *justified* procedure for answering q will depend on which facts are in F , and on how a relates to them. If the facts in F determine how q should be answered, then if a is justified, it must be "suitably related to" (i.e. justified in virtue of) f . Conversely, if $a_1 \dots a_i$ are just those justified procedures for answering q , then the facts in F "will have to be related to those procedures in such a way as to render them justified" (Zalabardo 1997, p. 479). Zalabardo's Kripke appeals to this latter thought in his critique of dispositionalism: the idea is that no dispositional fact could be related to my actual answering procedure in this way.⁹

Let us formulate this idea in more general terms:

Conscious Engagement	If $a_1 \dots a_i$ are just those justified procedures for answering q , then the facts in F will have to be related to those procedures in such a way as to render them justified
---------------------------------	--

The relationship Zalabardo's Kripke has in mind is that of *conscious engagement*; that is, a would not be justified unless it involved my consciously engaging with F . So for example, if the correct answer to q were determined by certain physiological facts about my brain, then any justified procedure for answering q would have to involve conscious engagement with those facts.

Even without providing much further characterisation of what this kind of relationship involves, the line against the dispositionalist is clear. If we take the best variety of a dispositional answer to the sceptic as involving appeal to 'how I would be disposed to answer under ideal circumstances,' it is intuitively apparent that facts of this kind are not directly available to me in the right kind of way precisely because in order to get close to them I must speculate about ideal circumstances. The formulation of 'tentative hypotheses' about what I would do under such circumstances is not enough — I must engage directly with the actual facts as to what I would do under the circumstances that would be (as a matter of objective fact) ideal.

1.7 Summary

In this chapter I have attempted to broadly characterise Kripke's puzzle and introduce some key ideas that I will refer to and develop as we progress. Before we embark into new territory, I would like to connect this work to the previous one (Cross 2013) by looking at how the debate I examined there relates to Kripke's challenge, and to meaning-normativism. This is the topic of the next chapter.

may hypothesise out of concern about my arithmetical ability.

⁹That is, no candidate set of dispositional facts F determining how q should be answered could relate to any a for answering q in such a way as to render a justified.

Chapter 2

Normativity in the McDowell-Wright Debate

In previous work (Cross 2013) I attempted, with limited success, to unravel the connection between some of the best-known rule-following literature and the concept of normativity. Focusing on John McDowell's 'Wittgenstein on Following a Rule' (McDowell 1984), I discovered a vast web of concerns centring ultimately on the tension between a communitarian conception of truth, represented there in the views of Crispin Wright (Wright 1980) — and, to a lesser extent, Kripke — and an apparently indispensable view of metaphysical objectivity that holds facts about the world to obtain independently of our judgements about them. McDowell challenges Wright's revisionism about this, arguing that if a community of speakers ultimately determines the meaning-rules of our language then rule-following, and hence meaning, cannot be possible. Quite how he substantiates this argument, and what positive position he might prefer, became the subject of my analysis.

2.1 Normativity in McDowell 1984

McDowell's most forthcoming strategy in that paper is a transcendental argument against anti-realism. Urging the indispensability of the conception of objectivity it would sacrifice, McDowell suggests that anti-realism simply will not do. However, this does little to move the dialectic along, as Wright reserves the option of simply denying any such indispensability — indeed, something of the kind seems implicit in his claim that the global community of language practitioners simply has “no standard to meet” when it comes to the correctness or incorrectness of their actions (Wright 1980, p. 220).

So the weight of McDowell's position rests on his ability to show that Wright's anti-realism is deficient in some other sense. The general strategy is to outline some kind of important contrast that we take for granted, and then explain why Wright cannot account for it. This has some degree of force, and I attempted (Cross 2013, §2.3) to support McDowell's case with the observation that Wright makes use of suspiciously extra-communal sounding language to describe the operation of his communitarian model. Indeed, this theme of ‘mixed levels’ of descriptive fact is one McDowell introduces himself.

This is all very well, but as was subsequently pointed out to me, this entire discussion of rule-following and meaning seems to make very little direct appeal

to *normativity* as such.¹ To say that a community of language users is or is not the source of meaning-determining rules and practices is not necessarily to say a great deal about agential relationships to those rules and practices. It is one thing to establish that rules of some particular kind exist in virtue of some particular metaphysical arrangement or other, but it is quite another thing to describe the means of operation and governance proper to those rules. So in a debate that is presented as revolving around the normativity of meaning, the disputants apparently become sidelined in the theory of truth and ontological analysis of rules.

In fact, there is some significant content on normativity to be found in McDowell et al (and Kripke, of course, was largely responsible for injecting the idea into discussions of the rule-following considerations). In his (1984), they figure most prominently in McDowell's attempt to formulate a cogent transcendental argument against anti-realism. McDowell's case seems to be that the communitarian anti-realist's ideological commitments force him to describe language practices, and hence meaning itself, in causal, dispositional, and certainly non-linguistic terms. This, he claims, comes at the expense of the *oughts* that underpin language practices as normative enterprises. Thus, somehow, there is a problem: the proper description of any situation where some expression is sanctioned over another is no longer one that involves normativity at all. Instead, the proper description here is one of brute, causal, blurring out of the kinds of things our peers are disposed to find agreeable. McDowell appeals to our better instincts, emphasising that this cannot be the ultimate truth of the matter. But as mentioned above, he struggles to justify that intuition in a way that engages effectively with the anti-realist challenge, and retreats to safer ground in order to show us the less contentious limitations his opponents face.

The example of this he supports most effectively is the distinction, inspired by Wittgenstein, between asking for something yellow, and asking for something all competent language users would *call* yellow (McDowell 1984, p. 335). Wittgenstein says:

Could a justification of an action as the execution of an order run like this: "You said 'Bring me a yellow flower', whereupon this flower gave me a feeling of satisfaction; that's why I've brought it"? Wouldn't one have to reply: "But I didn't tell you to bring me a flower that would give you that sort of feeling in response to my words!"?

(Wittgenstein 1953, §460)

McDowell claims it is in the spirit of this thought, that a feeling of correctness is distinct — indeed, *distinguishable* — from correctness itself, that we might object to the communitarian view of correctness as determined by communal assent, here conceived as a shared 'feeling of satisfaction'. If what it is to follow a rule is just to secure communal assent on the matter, then we naturally face a problem in accounting for the common-sense intuition that these two notions can come apart; that to many kinds of judgment, assent is one thing and correctness, or the matter of what *ought* to be done, is another. As observed, this is not a fatal concern for Wright, who considers his position revisionary anyway. However,

¹I would like to thank Daniel Whiting and Bill Brewer for their comments highlighting this issue.

the more interesting point is that appropriate fulfilment of the command ‘bring me a yellow flower’ seems to involve something normative. McDowell switches to discussion of commands here not arbitrarily, but apparently because there is something to be had in the clarity of the normative import of this situation. It is an analogy: in language, the thought goes, one necessarily adheres to prescriptions of a kind, and this adherence is a normatively charged phenomenon. McDowell seems to be saying that one is properly *obliged* rather than merely *conditioned* to interact linguistically with others in certain kinds of ways, and that running the two descriptions together entails loss of content. So in this respect at least, the most significant shortcoming of a communitarian anti-realism like Wright’s is that it “has no room for norms, and hence — given the normativeness of meaning — no room for meaning” (McDowell 1984, p. 336).

This reveals that the seat of McDowell’s challenge to Wright here is in the lack of normativity at the ‘basic level’: at root, Wright’s universe is properly understood as non-normative, and because we (supposedly) source meaning in normativity, Wright is thus guilty of stripping all meaning from the world presented to us in veridical experience. But this does depend on a further thought: that the disposition of Wright’s agents to ‘march in step’ in linguistic matters as a result of their cultivated attitudes is not in itself constitutive of the relevant kind of normativity.

I return to this theme in chapters 3 and 5, but in the present context this thought raises a question for McDowell: the most apparent way to justify claiming that communal action cannot constitute meaning or normativity is by demanding that meaning be located in some sense ‘beyond’ those actions. That is, there must be some kind of ‘independent’ standard of meaning by virtue of which an agent can be said to have responded as they ought in answer to the question “What is $68 + 57$?”, and crucially, *also* with respect to which that agent’s community can be said to have judged correctly or incorrectly that he responded as he ought. But, it might be replied, why would we expect this?

The answer McDowell has in mind is clearly that we want to preserve an intuitive picture of objectivity. That means that whether “5” is the right answer to a maths problem has to depend on more than just communal assent. But then, while Wright is genuinely advocating anti-realism about arithmetic and the objection seems to connect here, Kripke seems fairly clear that his own challenge does *not* question the objectivity of mathematical truth; his sceptic does not question the arithmetical sum of 68 and 57, but rather, whether the plus function is the one I mean when I say ‘plus’. As such, it is not immediately clear that McDowell has the intuitive high-ground. We may be happy to say that Wright’s account is lacking as a result of its ‘global’ anti-realism because we want to preserve a realist, objective view of facts about mathematics, and yet we may still be less enamoured with the idea that facts about what I mean by ‘plus’ are also objectively determined in this way.

As such, McDowell’s position would benefit from strengthening: he needs to motivate the idea that the facts about meaning are suitably determined above and beyond the realm of mere assent.² It has to be demonstrated either that there is some ‘mechanical’ problem with the possibility of communally-determined meaning, or else that the latter fails to meet our expectations at some finer conceptual

²This position has a few things in common with what Hattiangadi (2007) calls ‘semantic realism’ — see also my chapter 6 below.

level. And this is where the appeal to PI§460 comes in: we want to preserve some vestige of possibility that the community reaches an agreement about what I meant previously by ‘plus’, but where their decision is also *wrong*. Wright’s account robs us of this possibility, and hence also of the possibility that I ever *could* have meant *quus* by ‘plus’ if to all observers it looks like I meant *plus*: but the force of Kripke’s challenge surely comes from the fact that I could have meant this, an idea we would seemingly want to preserve. In normative terms, what is at stake is the possibility of distinguishing what we all agree I ought to do from what ‘in actual fact’ I ought to do.

The essential normativity of meaning was taken largely for granted in the decade or so following Kripke’s ‘Wittgenstein on Rules and Private Language’ (1982), and this goes some way to explaining McDowell’s inference from an absence of normativity to an absence of meaning. Since a core aspect of McDowell’s overall strategy was to show that Wright could not account for meaning, as a step on the way to showing that he could not account for common-sense objective judgments of investigation-independent fact, it should be apparent that consideration of normativity and its relation to rule-following is essential to McDowell’s strategy.

2.2 McDowell’s Dilemma

McDowell (1984, pp. 341–342) describes a dilemma that Kripke’s challenge seems to present, and uses this dilemma to frame the debate. I have given this topic considerable attention elsewhere (Cross 2013, §4.2) so here I shall try to be concise.

McDowell thinks that Kripke’s challenge gets off the ground because we permit the assumption that understanding anything is always some kind of interpretive enterprise, where an interpretation counts as correct by (possibly explicit) reference to some kind of standard. The alternative seems to mean accepting that nothing is normative, assuming that this is entailed by Wright’s account. McDowell aims for a third way — something that navigates us safely between Wright’s account and Kripke’s paradox.

If we accept the assumption of the first horn then the threat of Kripke’s regress looms. If we allow that understanding is always a kind of interpretation, and that whether any interpretation is correct is itself an interpretation, which therefore may itself be correct or incorrect, there seems to be nothing concrete we can cite that would justify any interpretation of anything, and thus mandate one understanding over another as ‘correct’. Whether I mean *plus* rather than *quus* is subject to interpretation, a process which is itself dogged by the same question: perhaps it is the case that I mean *plus* by ‘plus’ except in precisely the present circumstances, in which case I mean *quus*. No ‘facts about me’ seem available to indicate otherwise, for they too must be interpreted in one way rather than another. . . and so on. McDowell thinks that, having rejected Kripke’s ‘sceptical solution,’ the only way out from here is by appeal to some kind of interpretation that never fails, consistently yielding a determinate result for all new cases, a ‘super-rigid’ interpretation (McDowell 1984, p. 332).³

³See also Cross 2013, §3.2.

The problem with this is not the claim that understanding some expression means that we always have a determinate answer to the question ‘what is the appropriate response?’, but rather that the answer is supposed to come in the same form as the answer to any other question: an interpretation.⁴ In the context of the present model of understanding, that just means something that could otherwise have been misinterpreted, but by some grace of God, turns out never to be problematic in this way. Of course, this is a ludicrous and dangerously *ad hoc* sounding response to Kripke’s challenge. It is significant mainly as a direction as to what kind of solutions would *not* be acceptable.

Denying the regress seems to mean embracing the second horn of the dilemma, denying the irreducibility of normativity and certain norm-governed notions (truth, meaning, etc.). Thus we agree that something is true because of a complicated interaction between various merely causal phenomena, and we get something that looks quite a lot like Wright’s communitarianism. We are disposed to react in certain ways to certain stimuli; we share this disposition uniformly enough across the species; we impulsively seek the assenting behaviour of our peers, and so on.

As I hope is clear, McDowell objects to this option, and I have previously (Cross 2013) presented my best attempt at understanding why. The concern, I gather, can be understood as follows. The legitimacy of our normal claims to know objective facts about reality depends on it *not* being *fundamentally* the case that what we call an ‘objective fact’ is determined as a matter of brute causal dispositions and non-normative facts about us. If this is the ‘real truth’ about cases where we call something ‘true’, then it follows that there is no question whether we *ought* to call something a fact, only a question whether we *will*, as a pure dispositional fact about behavioural propensities in response to stimuli. Yet intuitively, we *ought* to call something ‘true’ if and only if it *is true*: not, if and only if *we would call it true under certain ideal conditions*. . . , or anything of the kind.

2.3 Normativity and the Illusion Argument

Perhaps we might be attracted by the conception of normativity as a causally governed, brute fact-of-nature type phenomenon, such that the best explanation of a true claim that S ought to ϕ is an explanation focusing on facts about the wiring of S’s brain, the gregariousness of human beings, their propensity to conform to the behavioural patterns of their peers, and so on. Undoubtedly, such facts are going to be relevant.

However, we might also find this thought uncomfortable. If it’s really true that S ought to ϕ , we might think, then we shouldn’t be able to characterise the ‘ought’ operative here in non-normative terms like this. For the ‘ought’ here to be genuine would be for it not to be reducible to something un-ought-like. And this is precisely the kind of worry that Kripke appealed to when he announced, persuasively at the time, that “the relation of meaning and intention to future action is *normative*, not *descriptive*” (Kripke 1982, p. 37). For if the normativity at stake were straightforwardly reducible to dispositions, this objection would carry no weight — but the undertone here is perhaps that the ‘real story’ about what I mean is *fundamentally* one about normativity, *not* about my dispositions.

⁴Cf. McDowell 2009, p. 83.

As indicated, it is not that the anti-reductivist is averse to the thought that normativity involves facts about human nature, psychology, and other *dispositions* of a non-normative variety. For if this were the case his position would surely be hopelessly at odds with the natural sciences. Rather, he takes issue with the reductivist enterprise itself. There is something about *reducing* the normative to the non-normative that he finds uncomfortable. However, in McDowell's case it is never quite articulated why this is, beyond a desire to sustain a 'base level' picture of reality that contains normativity — which is not exactly uncontroversial.

So in my previous work I came to the conclusion that McDowell doesn't get very far in developing a sceptic-resistant anti-reductivist account of normativity. His debate with Wright confuses issues of normativity and objectivity to the extent that it is unclear which McDowell thinks comes first. However, what does shine through is his objection to Wright on the basis of what I called the Illusion Argument (Cross 2013, pp. 39-40). He argues that if the normative is properly reducible to the non-normative, then our propensity to think we are genuinely subject to norms looks like a propensity to the *illusion* that we are subject to norms. Then comes a slightly odd, phenomenological turn in the reasoning:

... once we have this [norm-free] picture, it seems impossible simply to retain alongside it a different picture, in which the openness of an individual to correction by his fellows means that he is subject to norms. The first picture irresistibly claims primacy, leaving our openness to correction by our fellows looking like, at best, an explanation of our propensity to the illusion that we are subject to norms. (McDowell 1984, p. 347, my brackets)

McDowell reasons that once we accept the idea that what we *ought* to do with language is a matter of staying in step with our communal fellows, we are left with no psychological barrier to accepting that normativity is 'really' an illusion rather than a genuine phenomenon to be explained (McDowell 1984, p. 336, p. 347).

This, it turns out, is quite possibly the linchpin in McDowell's reasoning. The aim is to demonstrate that Wright and the reductionists must have it wrong, since they take normativity to be a genuine phenomenon – more or less – and certainly not illusory. But the trouble is, even if the idea that we can't resist thinking of normativity as an illusion were true, it would seem this should stimulate a further rush of phenomenological reflections on the theme of *why* we are also inclined to desire that normativity be more than illusory... and so on. The point is, the argument seems to depend a lot on us sharing some inclinations with McDowell that go beyond the standard (though perhaps no less questionable, at base) philosophical appeal to intuitions. It seems, then, that McDowell's transcendental reasoning lacks some of the required intuitive force.

Chapter 3

Externalist Normativity

Ruth Garrett Millikan (1990) has proposed an account of rule-following in response to Kripke's paradox that appeals to normativity of a biological variety. The regress of rules is stemmed by one's competence at conforming to a certain kind of (unexpressed) *biological purpose*. This purpose is the standard against which one's performance is ultimately judged, and is the root source of the normativity involved in rule-following and the activity of meaningful linguistic exchange.

In this chapter I critically examine the basis of this view, concluding that the sort of normativity appealed to is a) not the kind of thing Kripke had in mind in his presentation of the challenge, and b) not the kind of thing we would want to call 'real normativity.' I argue that the externalist sourcing of this root kind of normativity makes it difficult to wield as an action-guiding 'ought' in the kind of way that applies to language practices, since it means that the seat of our justification for answering, in Kripke's case, '125' rather than '5,' is unacceptably distant. I argue that this 'externalist normativity' is a few steps removed from the intuitive picture of meaning-normativity as justification for action with respect to our semantic intentions. I propose that Kripke's paradox can be resolved with a far simpler account of meaning-normativity than Millikan's if we're willing accept the same degree of intentional disconnect, via what I call the 'quick externalist solution'.

3.1 Rules with Biological Purpose

Millikan sees Kripke's sceptic as challenging us to account for

[...] the *normative* element that is involved when one means to follow a rule, to account for there being a *standard* from which the facts, or one's dispositions, can diverge. (Millikan 1990, p. 329, my ellipsis)

The problem as presented is supposed to be that such a 'standard,' whether internal or external, could not ultimately be something 'represented' to me. It might be that I think of an explicit rule (say, '+2' for Wittgenstein's example) and try to follow that; but then, the possibility of my doing so depends on, it would seem, the interpretation of some further rule for following the '+2' rule. The troublesome element here is the 'interpretation' involved, and this is taken to be an inevitable feature of the fact that the rule is some kind of representational 'given'. Millikan's goal is to prevent this regress by inserting a stage in the justificational

chain that is non-representational. This is the ‘standard’ she mentions: and the standard she has in mind is a biological one.

Say the heart has a purpose. This can be expressed by a rule like ‘pump blood around the body,’ or ‘expand and contract at a regular interval,’ or even, ‘keep the organism alive.’ This is the kind of biological purpose Millikan has in mind. It is a result of evolutionary design, being a reason for the proliferation of the organism in question. Such purposes can, however, be somewhat more complicated than this: Millikan provides the example of the male hoverfly, which has a propensity to dart suddenly at a mathematically precise trajectory in response to certain visual stimuli. She describes this as a rule, say: ‘dart suddenly in such and such a direction whenever you get such and such a visual stimulus.’ This serves the hoverfly well from an evolutionary perspective: for in many cases it enables the hoverfly to find a female and pass on the same behavioural trait to a new generation of hoverflies. Moreover, the hoverfly clearly has no knowledge of this rule, nor any ability to form an intention to follow it — rather, it is a behavioural propensity of the hoverfly with an externally characterisable ‘biological purpose.’ Thus, a given hoverfly *follows* this rule in the sense of displaying a competence to conform to it; with reference to this rule, we can pick out the hoverflies that do ‘the right thing’ from those that are defective in their behaviour.

Furthermore, this affords us a means of saying determinately that the hoverfly is following *this* rule rather than some other ‘quus-like’ variant, which might have the hoverfly darting in a trajectory relative to certain visual stimuli except on the third Tuesday in May (or some such thing), because the quus-like alternative serves no biological purpose. There can be no evolutionary advantage to following a quus-like rule, and so if a hoverfly is in fact disposed to act in a way that coincides with *this* rule, we can feasibly characterise this as defective with reference to a legitimate, biological standard of self-preservation and genetic propagation.¹

A lot more could be said about how this is to operate, especially, for example, how we are to determine which purposes are the gold-standard biological ones. But for the sake of expediency, my focus will be solely on the way that Millikan takes this to be a solution to the task as she herself states it: to account for the *normative* aspect of rule-following, and of meaningful language use insofar as that is a paradigmatic rule-following activity.

3.2 Millikan and Kripke

In Millikan’s discussion of Kripke’s critique of dispositionalism, she identifies 2 main reasons offered for rejecting it: a) that people are disposed to be mistaken, and b) the finitude of our dispositions. As we have seen, though, where others (e.g. Ginsborg) have identified a third and separate ‘objection from normativity,’ i.e. that dispositionalists cannot account for the normativity of meaning, Millikan has Kripke presenting the famous thought that ‘whatever in fact I (am disposed to) do, there is a unique thing that I *should* do’ as a further distillation of points (a)

¹It seems to me that we could go a step further, though I am not sure Millikan intends this, and say that the biological standard somehow ‘rules out’ the analysis according to which the hoverfly whose behaviour conforms to a quus-like variant is following the quus-like rule properly rather than following the biologically purposeful rule badly. Perhaps we could argue that such non-biologically-purposeful rules do not apply to hoverflies.

and (b). The problem is located in the fallibility and limited scope of an individual human's capacity for self-regulation — weaknesses not found in the purely natural process of evolution by natural selection — and so Millikan identifies that something similarly unbending, without these weaknesses and variance, must underpin our capacity for rule-following and represent a suitably rigid standard of correct action.

What we'd like to say is that there is supposed to be something against which my performance is *measured* — and this is surely something like what Kripke has in mind. But whatever my meaning something by an expression amounts to, “its nature must lie in part in what is *not* simply given to consciousness” (Millikan 1990, p. 326): it is not against some ‘given’ object that my performance is measured, since the measuring itself would be a rule-following activity of the same kind, and would thus require some further standard to be measured against. And this is clearly how Kripke's regress begins.

Millikan argues that the posited given-ness is the problem here, rather than the subsequent notion that what is given should be some sort of internal or mental item, since it makes no difference to the generation of the regress whether the standard is something ‘in my mind’ or some physical thing that I can hold in my hand. Either way, appealing to it to judge my performance means appealing to some further standard. . . and so on. So instead, the standard she has in mind is naturally not something which need be given to us in whatever manner. She favours a biological — which is to say, in one way or another, *evolutionary* — standard that performance is ultimately judged against. There are two such kinds of ‘rules’: those of a *proximal* nature, “the following of which depends only upon incoming sensory information and an intact physiology,” and those that are *distal*, giving an (often multiply realisable) characterisation of what the organism ought to be doing that will help it survive, and the following of which depends also “upon the actual state of the world”. So a proximal rule for humans might be ‘run away when you see large carnivores,’ while the corresponding distal rule might be ‘don't get eaten.’

But the striking thing about her account overall is that it seems to reject out of hand the very internalistic flavour others have found in Kripke's challenge. For the standards in question are applicable (in Millikan's own example) to the activities of apparently non-intentional, and certainly non-intelligent life-forms, for whom there is simply no question that they might ever know or comprehend what it is that directs their action. This seems at odds with Kripke's talk of justified action. The question whether the actions of some hoverfly, Millikan's chief example, are justified ‘with respect to its past intentions,’ seems empty: she does not seem to be implying that hoverflies have the kinds of intentions or intentional content that would be relevant to Kripke's challenge, hence no useful talk of justification can emerge. And yet, the whole line Millikan takes revolves around the simple ‘scaling up’ of the hoverfly case: humans follow rules of these two varieties also, and it is this that grounds the fact of their meaning one thing rather than another by what they say, and justifies what they do in fact say. If hoverflies have no intentions, then Kripke's challenge apparently does not extend to them, since his question about rule-following clearly solicits explanation of our capacity to embark wilfully on a certain course of action in order to fulfil the demands of some rule.

To be justified is surely to be part of some institution, with respect to which

one is justified. That is, justification is itself an institution: it does not apply, at least not in the sense that interests Kripke, to those that do not partake of it. This is not to render normative talk about animals meaningless, since claims like ‘the hoverfly shouldn’t have flown so close to the hungry frog’ are wholly legitimate without reference to human intentions and desires — there are purely biological obligations at work. The point is just that it would seem confused to describe this or any other error on the part of the hoverfly in terms of justifications with respect to its past intentions, which is apparently quite a different concept of error and one that does not belong here. The trouble is that if she is to satisfy Kripke’s challenge, Millikan must have us leap from the former, plausible account of *external* biological imperatives into an account of *internal* justification with respect to mental content.

There are indeed cases where *explanatory* justification of animal actions is a legitimate topic — so, for example, the hoverfly that darts suddenly towards a car may be justified in doing so with respect to Millikan’s proximal hoverfly rule. The rule justifies the action in the sense that it explains the purpose behind it — and indeed, it is in these terms that Millikan aims to conduct the debate. But the purpose of the hoverfly’s actions is akin to the purpose of a spark plug’s firing in the car’s engine, which is justified by the proximal spark plug rule: ‘whenever the ignition is turned, fire.’ The purpose relevant to human rule-following activities, on the other hand, is that of *agency*.

This is, in a sense, the trouble with Millikan’s otherwise very appealing account of rules in the world. When Kripke insists on the availability of the rule or justifying fact to me in some immediate way, such that it can *tell* me what I ought to do (and so on), he is not playing devil’s advocate of an intuition he intends to revise in a game of ‘here’s what we thought, here’s why it fails’ (as Wittgensteinian as that might be). Rather, his sceptical solution is precisely that: an admission of defeat, and a regrouping on a new and less proud footing. The idea that rule-following involves my intentions directly, that it has some footing in my conscious activity as opposed to the purely external, purely biological basis Millikan affords it, is one that we should not be willing to part with at such short notice. It seems that either Millikan fails to take Kripke’s challenge seriously, and so by his lights fails to give a satisfactory account of rule-following, or else presents only a ‘sceptical solution’ as Kripke does: the concession in her case being that meaning-facts and justification-facts need not be accessible in the way Kripke thinks essential.

Whether we agree with Kripke’s internalistic premisses or not, something significant emerges from these considerations. Millikan’s discussion takes biological purposes, or ‘purposing’ as the source of the normative element of rule-following (and so, meaning). However, the solution she presents demands that the ‘purposing’ in question be something externally-determined: that is, the standard by which success is measured is, first, not in fact necessarily measured (or measurable) by anyone, and second, it is quite possibly not made plain to me at all.

The consequence of this is that the image of normativity at work looks unfamiliar. The source of Kripke’s normativity intuition was presumably the thought that there are certain ways that I ought to use some expression — but it is troubling to think that the oughts in question may not be plain to me at all. On Millikan’s view, the oughts — the normativity of meaning — are sourced in nature, are determined in line with evolution perhaps, though no less sophisticated for it: and yet, this means that some such oughts are to be *discovered* by empirical enquiry. And

not, indeed, enquiry of the kind that Kripke discusses, where I encounter some new case of expression use, and with it the ‘ought’ pertaining to that particular scenario: rather, it is the kind of enquiry that leads to the complex and precise ‘proximal hoverfly rule’ Millikan outlines, and is properly the domain of the natural sciences. Rather, as Kripke suggests, it seems that if I can follow a rule then I can be confident that my actions are justified with respect to that rule, and without consulting the expertise of anyone else.

3.3 The Justification Question

A clear assumption of Millikan’s solution to Kripke’s challenge is that having a suitable standard against which your actions may be measured is sufficient for normativity. In the light of our present considerations, this seems close to question-begging. As Glüer and Wikforss indicate (2009a, cf. also chapter 6 below), this is far from obvious in the case of *truth*, where an assertion is true iff it meets the standard of being the case, but nothing normative seems to follow straight off about whether we *ought* to speak the truth. By the same reasoning, that one has a biological function is not sufficient to establish that one ought to fulfil it; that one has a biological purpose is not sufficient to establish that one ought to conform to it. It may, perhaps, provide some *reason*, if only an *explanatory* one, to underlie some action you perform: but it is not obviously the motivating source of that action, in the proper normative respects. It seems to me that the only way to escape this line of questioning, if defining a universal standard for normativity, is to begin work from the perspective of an agent who is subject to norms, not from the objective perspective being proposed here.

What we would like to know, it seems, is either what justifies (in Kripke’s case) my answer ‘125,’ or what makes it the case that I ought to say ‘125’. Millikan’s answer seems to be that, ultimately, my capacity to conform to a biological standard in a way that has a normal explanation (compared to the explanation of the same conformity exhibited among my ancestors) is both what makes it the case that I follow one rule rather than another, and what justifies my actions. On the latter point, we are supposed to take comfort, or so it seems, in the thought that my continued survival to the point of passing on my genes is the justification of my saying ‘125’. Whether I am successful at this particular instance of rule-following (following the meaning-rule for ‘plus’) is a matter of whether I am successful in achieving certain inexplicit biological purposes; whether I have done what I ought to have done, is just the same thing.

There is surely no denying that, from a commonsensical perspective, one ought to do what preserves one’s life; and also, from a somewhat stark evolutionary perspective, that as an organism one ought to do what maximises the proliferation of one’s genetic material.² What is troubling is that this should be held aloft as the source, or at least the grounding, of all normativity. The expression ‘biological imperative’ seems applicable in this case: but the operation Millikan envisages is the derivation from them of all other imperatives, and consequently, of all other kinds of justification.

²I hasten to add: this does not mean constant and widespread procreation, which can clearly overburden an ecosystem to the ruin of all descendants.

This is troubling precisely because it inserts a point in the justificatory chain where the very question of our justification is neither asked by us nor answerable within the scope of our practices. That is, the justification we were looking for has been conceded as unavailable, and located outside the realm of our own institutions. What we should have liked to say was that the reason one ought to say ‘125’ has a mathematical answer, or a linguistic one, but something that could be located within a realm that is familiar to us as providing justification. Millikan grounds justification outside any such realm: the true justification of my saying ‘125’ may clearly be not only unknown but incomprehensible to me and all other members of my species. It seems that, at its very roots, Millikan’s account denies us the kind of internally available justification for action that Kripke sought — but if we are willing to make such a radical departure from the requirements specified, I think we could actually answer the challenge much more straightforwardly.

3.4 Meaning without Intention

Kripke’s sceptical reasoning seems to uncover a paradox inherent to the notion of following a rule: any course of action accords with an infinite number of conflicting rules, yet nothing about my actions or mental activities determines that I am following any particular one of those rules; nor, consequently, that I ever mean one thing rather than another by my words. Kripke thus draws an “incredible and self-defeating conclusion, that all language is meaningless” (Kripke 1982, p. 71). If it is this conclusion that is the real ‘paradox’, then our challenge on the face of things is to falsify it while preserving the key assumptions of the debate: that meaning is normative, and that my meaning on any given occasion of language use is determined by semantic rules.

Kripke’s invokes the normativity of meaning to reject any straight dispositionalist solution to his paradox. He says that a dispositionalist account can at best describe what I *will* do if I am to accord with my previous intentions for use of the word ‘plus’, whereas the challenge was to say what I *should* do in light of those intentions. So he summarises, famously, that “the relation of meaning and intention to future action is *normative*, not *descriptive*” (1982, p. 37).

While I agree with the sentiment, the phrasing here paves the way for a confusion about how meaning and intention interact. We might happily grant that both my past meaning and past intentions relate normatively to my future action, but we should be clear that the two are *separable* phenomena. That is: what I previously meant, and what I previously intended, are not the same thing.

I’ll begin by arguing for the separability claim. A fairly traditional thought in the philosophy of language is that the meaning of an uttered expression is partly determined by some mental event on the part of the speaker, even if Frege rejected the supposition that meaning is determined by the holding of some private idea.³ What we at least tend to think is that my utterance gets the meaning it does as a function of my intentions. So for example, in the literature we hear thought experiments where ‘S means *F* by *x*’, where the idea is that S partially determines the meaning of *x* by some kind of mental effort.

³Cf. Frege 1918.

The door is open for simple confusion of two senses of the word ‘mean’. “What do you mean by this?” is a question with two possible meanings. Perhaps you wrote a word the speaker does not know and he is asking what it means. This is a question about semantics. Alternatively, the speaker might be asking why you have done something. That is a question about your intentions. However, it may also be a question about what you *intended* the word you wrote to mean. Let’s adopt the convention of using the word ‘mean’ and its cognates exclusively for the semantic sense, and say ‘intends to mean’ for the intentional sense.

The point here is that if we can sustain the idea that meaning and its (putatively) associated obligations are in no way dependent on the private intentional content of the meaner, then Kripke’s apocalyptic conclusion about the possibility of meaning one thing rather than another loses its bite. That is, if meaning-engendered norms do not depend metaphysically on the intentional content of speakers, then what I mean by a word is not up for debate any more than is the arithmetic sum of 68 and 57.

The central claim under consideration, then, is that there is no necessary connection between meaning and intentions. It is possible that a word means F when no-one intends it to mean anything. Conversely, it is possible that S intends his utterance of x to mean F , but his utterance of x does not mean F . I will proceed by examining these two possibilities in greater depth.

First, the case of meaning without intention. It is possible to write a computer program that, when executed, will output a random string of 5 letters to the computer’s screen. We could nest this program inside a loop, and have the computer keep outputting 5-letter strings as long as we like. Now, it is possible that the computer will eventually output the string: GREEN. This will be insignificant to the computer, but to us it is clearly the word ‘green’. Therefore, words can be produced without intention. Indeed, we can imagine less controlled cases with the same output. For example, imagine a shipment of fridge magnets falls off the back of a lorry, and some of the magnets land on the car behind, landing in the order ‘green’. It is hard to imagine a case for interpreting this as anything other than the word ‘green’, yet it is clearly produced in the absence of any determining intentions.⁴

The more important question is whether a randomly produced word is meaningful. Again, I think there is not much case for denying this. Imagine that we extend our program to allow the computer to continuously output random alphanumeric and punctuation characters. In line with a famous thought-experiment about monkeys, the computer could possibly output characters identical in positioning to the whole text of *Hamlet*. Would this output be meaningful? I think so; indeed, the computer could produce an entirely new novel (as could sufficient quantities of misplaced fridge magnets), and the words in the novel would be meaningful too, perhaps even compelling. The fact is, if we read the string ‘Emeralds are green’ on a computer screen, whether it was generated by an intentional communicator or by a computer randomly picking ASCII characters will make no difference to the plain fact that the sentence is meaningful.⁵ Thus I conclude that words can be meaningful without there having been an intention that they mean anything.

⁴If you are concerned about the provision of a prefabricated set of roman letters, imagine instead that the computer randomly turns certain pixels black on the screen, and happens to print the shape: GREEN; or that the lorry is carrying a shipment of ball-bearings that happen to fall into that shape.

⁵In a way this is an extension of Searle’s ‘Chinese Room’ thought-experiment (Searle 1980).

Now we should look at the reverse case. Can S intend to mean F by his utterance of x , and yet fail to mean F by x ? Imagine that when Shakespeare wrote the text of *Hamlet* he actually intended to rewrite *Macbeth*. So for example, when the text reads,

“To be or not to be, that is the question.”

Shakespeare actually intended this to mean the same thing as:

“Is this a dagger I see before me, the handle towards my hand?”

That is, for every word in the text of *Hamlet*, Shakespeare had totally different intentions for its meaning than what it has been interpreted to mean. Would this imply that the true meaning of each word in the text is thus what Shakespeare intended, not what we have thought it to mean for hundreds of years?

Surely not. If Shakespeare intended to mean *hand* by ‘question’ then he simply failed to do so. Similarly, if I intend to mean *red* by ‘green’ in my utterance of the sentence ‘Emeralds are green,’ then unless I provide sufficient contextual information to determine otherwise, ‘green’ in that sentence simply means *green*. I think we should conclude from this that, intuitively at least, the meaning of a word is determined by semantic facts, e.g. the fact that ‘green’ means *green*. Of course, these facts need not be so simplistic: they could be contextually indexed, such that ‘green’ used to describe a non-green-coloured but seasick person means *ill*, or such that if S states his intention to mean *red* by ‘green’ at t , then S’s utterance of ‘green’ at t means *red*. All we want to deny is that a speaker’s *private intentions* directly determine meaning.

As such, we are considering a realist semantics that holds the meaning of any instance of a word to be entirely determined by which (if any) semantic rules it satisfies, taken as entities that are not dependent on anyone’s intentions. This is not to go so far as endorsing what Hartry Field has called ‘semanticalism’, the doctrine that there are primitive, irreducibly semantic facts (Field 1972, p. 358). Rather, it is the weaker claim that there are semantic facts, but that whatever their ontological status, they are not constitutively dependent on determinate mental content.

3.5 What did Kripke ‘mean’?

How does the premiss that meaning is not determined by intention impact on Kripke’s sceptical reasoning? We will need an account of what that reasoning was in the terms we have defined. However, we should begin by using the term ‘mean’ ambiguously as Kripke does, indicating the ambiguity by appending ‘*’ to the term and its cognates as it appears in the reasoning. We can subsequently try to interpret whether the various instances of the term communicate something semantic or intentional.

What precisely does Kripke’s sceptic ask for? The first challenge is to indicate some fact that determines that I ever meant* one thing rather than another by

If the person in the Chinese Room just sends through characters at random, the Chinese person outside may nonetheless find themselves reading the *Spring and Autumn Annals*.

‘plus’, when nothing about my past usage of that term, or anything about me at all, seems to determine that I meant* just one of an infinite number of things by ‘plus’.

A further challenge is to indicate some “instructions I gave myself in the past that justify the answer ‘125’ rather than ‘5’” to the question: “What is $68 + 57$?” (Kripke 1982, p. 13). The sceptic asserts that when I presently answer the question with ‘125’ this response is just an “unjustified leap in the dark” (Kripke 1982, p. 15), since my past mental history is compatible with the hypothesis that I meant* one of any number of things by ‘plus’. Having never performed this calculation before, what is to distinguish this unhesitating response from the mere brute, unjustified inclination to answer this (or any other) way?

The preferred solution is to say that I followed and continue to follow a particular rule for using ‘plus’, and that this is what determines what I meant* and still mean* by ‘plus’. Yet I must then be able to justify the claim that I follow and continue to follow one particular rule, when it looks (as per the first challenge) as if nothing will justify this claim.

Kripke’s discussion of how what I previously meant* by plus relates to how I answer questions involving ‘plus’ serves to confuse matters somewhat. Contrary to what Kripke seems to assume, answering ‘125’ or ‘5’ does not involve meaning* anything by ‘plus’ or ‘+’. What it does involve is *understanding* ‘+’ in the question ‘What is $68 + 57$?’ to mean* one thing rather than another, and answering based on that understanding. My answer is justified, we hope, by the fact that I understand what ‘plus’ in the question means*, that I grasp some rule for interpreting ‘plus’ that determines that I ought to answer as I do.

We may thus be inclined to think I answer ‘125’ because there is some kind of algorithm which always instructed me how to answer ‘plus’ problems in the past and which now guides my actions when answering the question “What is $68 + 57$?” (Kripke 1982, pp. 15–16). Call this algorithm R1. The trouble is, my actions in the past are also consistent with my having been guided by the deviant algorithm R1*, where R1* guides me to act as R1 does in all but cases of addition with any arguments greater than 57. Postulating some further rule or algorithm R2 that I followed to guide me to follow R1 rather than R1* is unhelpful, as all my past actions are also compatible with my having been guided by R2*, which guides me to follow R1*, since there would still be no fact about my past usage that would distinguish my following R2 from my following R2*, and hence R1 from R1*, and hence meaning* *plus* rather than *quus* by ‘plus’. If there can be no fact about what I meant* by ‘plus’ in the past, then Kripke quickly infers that there can be no fact about what I mean* by it in the present either. From this he infers, as if trivial, that this entails the general meaninglessness* of language.

However, it should be clear from our distinction that the sceptical line of reasoning is flawed right from the outset. If ‘plus’ means (semantically) *plus*, then whatever I in fact intended to mean by ‘plus’ in the past, it follows that I in fact meant *plus* — and the same is true in present and future cases. The hypothesis that I meant *quus* is straightforwardly either proven or refuted by the semantic facts about what ‘plus’ means in such circumstances. Problems with determining my intentions are irrelevant.

Perhaps Kripke simply conflates the notions of ‘meaning’ and ‘intending to mean’ when he concludes that language is meaningless. What he ought to have

concluded is at most that language is not ‘used intentionally’, not that it does not in fact mean anything.⁶ However, in the interest of charity, we should assume that the explanation for what would otherwise constitute an enormous oversight lies elsewhere. There are surely other assumptions at work that serve to generate the paradox.

Kripke aims to reason from intuitive considerations to sceptical conclusions. But if it is intuitive that some semantic rule or fact (of whatever ontological status) exists that determines what I mean, then Kripke apparently must permit a premiss that immediately defeats his argument. And our suggested intuition here is not so far from what Kripke himself proposes as intuitive: he suggests we might think that when I grasp the meaning of ‘plus’, I grasp a rule that determines for infinitely many cases how I ought to use ‘plus’ in order to mean *plus* by it.⁷ This just seems to mean that the rule determines the meaning of any application of ‘plus’. Either my usage of ‘plus’ accords with what the rule dictates it ought to be, or it does not: either I succeed in meaning *plus* by an application of ‘plus’, or I do not.

Of course, this is not quite what Kripke has in mind: what he intends is that for the rule to come into effect I must deliberately *follow* it, and that to fail to *follow* such a rule is to fail to mean anything, even if my actions incidentally conform to one. Nonetheless, it is a slightly suspect notion that a rule only becomes operative on my utterances once I intend it to. If there are rules that determine meaning, then it is surprising to think they should operate on an ‘opt-in’ basis: we do not think this about the rules that determine the truth value of what we say, so why the semantic value? If this slightly odd premiss is the source of the problem then Kripke must do more to motivate it.

In any case, if Kripke does reject our proposed ‘factualist’ view of meaning determination then he at least owes us some kind of motivation for what seems a radically revisionist move, since ours may seem to be fairly run-of-the-mill assumptions about semantics. Indeed, the semantic facts we envisaged need not even rely on any kind of *realism*, hence need not be open to objections on that basis: thinking that there are facts about what words mean does not prevent us from taking those facts to be dependent upon communal consensus, for example, provided this does not involve dependence on the content of anyone’s intentions (assuming this would be problematic).

However, the more pressing problem for Kripke dialectically is that in rejecting the notion of semantic facts he immediately undermines his own later argument that any truth-conditional semantics is falsified by the paradox. Rejecting the existence of semantic facts entails rejecting truth-conditional semantics, but if this rejection is needed to generate the paradox that motivates the rejection, then the whole argument is clearly circular. Kripke must find independent reasons to deny our factualism, for we cannot be motivated to reject it by a paradox it does not cause. Indeed, the opposite is the case: Kripke’s paradox turns out to provide excellent motivation for *adopting* semantic factualism.

⁶It may be tempting to think that this is what he *does* conclude, and that the conflation is on my part. However, this would render mysterious Kripke’s subsequent proposal for a ‘sceptical solution’ that dispenses with conventional truth-conditional semantics (1982, Ch. 3). He clearly thinks that there is a problem with the possibility of facts about what words *mean*, not simply what is intended by them.

⁷Cf. Kripke 1982, p. 15

Kripke remarks early on in his discussion that the sceptic is not questioning any facts about arithmetic. It turns out, however, that for similar reasons he is also not questioning any facts about semantics, unless we assume that the meaning of a word on any particular occasion is determined as a direct consequence of the speaker's intentions, which I have argued is possibly not the case.

What the sceptic does seem to call into question effectively is the matter of my intending to conform to the demands of one rule rather than another: the determinacy of my *intentions* is called into question. If we think that meaning F by x necessarily involves having the intention to mean F by x , the sceptic's reasoning seems to invite the conclusion that meaning is impossible. Still, I have argued that if we permit the assumption that it is a semantic fact that 'plus' just means *plus*, then using the word 'plus' is sufficient for meaning *plus* by it. After all, if a trained parrot suddenly blurts out "Idiot!" in the direction of a visitor, it is likely that no intention to mean one thing rather than another is involved: and yet what the parrot says is clearly not only meaningful but rightly judged offensive.

3.6 Conclusion

The real trouble with this approach, aside from its somewhat unorthodox premiss that intention does not determine meaning, is that it leaves intentions and their relation to justified action entirely mysterious. Indeed, Kripke discusses a semantic case of rule-following (i.e. following rules in such a way as to produce meaningful speech and written language), but the problem is not specifically about meaning — the real focus is on accurately interpreting one's own intentional content, and this is simply not addressed here, or indeed, by Millikan's strategy, both of which explain semantic value and its putatively associated obligations without any reference to the intentions of the meaner.

In short, it seems we can easily falsify Kripke's claim to have demonstrated the impossibility of meaning by locating all the meaning-determining facts about an agent's actions outside of the intentional engagement of that agent. But this falls short of accounting for a really central intuition at play in Kripke's discussion: that a human agent has the ability to choose arbitrarily from infinitely many rules to follow, and the facts about which rules they follow — whether the 'plus' or the 'quus' rule or something not at all to do with the determination of meaning — *are* determined, by and large, by the intentional activities of that agent.

Chapter 4

Non-intentionalism

In this chapter I continue the discussion of the involvement of intentions in the notion of meaning-determination and rule-following. In particular, I explore the idea that a rule can be said to be ‘in force’ without the explicit or even possible involvement of mental content — thus testing an aspect of the ‘conscious engagement’ assumption Kripke seems to make (§1.6 above).

4.1 Intentionalism

In §3.4 we considered two possible thoughts about the determination of meaning. I argued (§3.4) for a notion of meaning as determined by semantic rules or facts, according to a process that is nothing to do with anyone’s intentions. So perhaps it is just a semantic fact that ‘by default any instance of the symbol ‘green’ means *green*’, or what might be a consequent semantic *rule*, that if ‘green’ is to mean *green* then it ought to be presented in such-and-such a way.

The alternative, endorsed by Kripke, was to think that what an expression means actually is essentially something to do with determinate mental activities. Kripke seems to think that the meaning of what I say is determined by my *following* a meaning-rule, which is not only for what I say to conform with the rule, but for me to *intend* this, in at least the minimal sense of deliberately having what I say conform to what the rule demands.

Of these two pictures of meaning, call the latter intentionalism and the former *non-intentionalism*, in light of the most apparent difference between the two. The non-intentionalist thinks that if an utterance conforms to the right semantic rule, or satisfies the right semantic conditions, this is sufficient for the utterance to be meaningful. The intentionalist, on the other hand, thinks that if semantic rules determine the meaning of an utterance then they do so only in part. Meaning is necessarily a consequence of a speaker’s following those rules; meaning *green* by ‘green’ is an inherently intentional action.

This ‘rule-intentionalism’ aligns roughly with Kripke’s preferred (but ultimately rejected) view of meaning as determined by the following of semantic rules. But by ‘intentionalism’ *tout court* I shall mean the broader idea that meaning is ultimately determined by some intention or other, whether by the act of intentionally conforming to (i.e. following) rules or otherwise. This broad intentionalist might simply hold that meaning-determining semantic facts all contain some intentional

condition, say, that an utterance of ‘green’ means *green* only if the utterer intends such-and-such.¹

That is a speaker-focused example, but beyond that, a weaker intentionalism might only require the involvement of *someone’s* intention in the constitution of meaning. So perhaps a weak intentionalist might hold that an utterance “green” means *green* in virtue of some commonly held intention that utterances of that kind should be understood to mean *green*, or perhaps that the utterance’s meaning *green* is conditional on the possibility that someone, anyone, understands it to have this meaning.

Weak it may be, but this latter form of intentionalism still cannot benefit from the paradox avoidance strategy discussed in §3.5. The involvement of determinate mental content of any kind in the constitution of semantic facts or rules opens us up to Kripke’s questions about how we know that they are determinate, and how we know what their determinate content is. This is unfortunate, as I suspect a weaker form of intentionalism is more broadly appealing than non-intentionalism. Supposing it is: should we actually prefer weak intentionalism?

Non-intentionalism seems to entail that any meaning-determining facts or rules are platonic entities that exist beyond the cognitive access of mere mortals. This amounts to a radically realist position: facts about language, an intuitively socially-constituted phenomenon, obtain irrespective of anything necessarily accessible or mind-dependent. By contrast, the intentionalist seems committed to the possibility of our having cognitive access to whatever determines meaning, since meaning is in an important sense determined by *us*.

However, this difference runs less deeply than it may seem. So far, we have painted non-intentionalism as enabling meaning ‘in a vacuum’, such that at some possible world with no living beings, if the right semantic facts obtain and some pebbles roll into a particular shape then result is something that means *green*. But whether non-intentionalism is consistent with this possibility is determined by how we think the relevant semantic facts come about, or how the relevant semantic rules come to be ‘in force’, not by the bare notion itself. The non-intentionalist is not committed to the ejection of human minds from the picture. The only commitment here is that determinate intention is not what ultimately determines meaning. For example, it could be that the linguistic community shares a common propensity to a non-cognitive feeling of ‘belongingness’ towards certain usages of language, similar to what Hannah Ginsborg has called ‘primitive normativity’ (Ginsborg 2011b).² The semantic fact that ‘green’ applies correctly to green things might then consist in the fact that the community is in such a way disposed towards certain usages of ‘green’. This disposition would not be dependent on our thoughts about how to use ‘green’, which would instead be a consequence of this more primitive notion.

Thus the non-intentionalist essentially denies that facts about meaning are determined by the precise cognitive mental content of any language user. In this sense it is then possible to discover the facts of language to be at odds with what we think. Is that a strange idea?

¹A further distinct view might be that no intentions are necessary *per se*, but that a speaker must understand (correctly interpret) some sort of semantic rule as a condition of his meaning *green* by ‘green’.

²See also Ginsborg 2012, and chapter 5 of the present work.

Perhaps not so strange on reflection. For it is exactly this point that Kripke's thoughts about rules brings out: we think that referring terms and predicates apply beyond our ken, that they have extension beyond the acknowledgement or experience of anyone living or dead. There are new cases to which they apply, with consequences that nobody has considered. Either we think that these consequences are there 'to be found', or that they are brought into being by our considering them — this is roughly the realism/anti-realism distinction. The latter position here is a product of concerns that we could not have access to facts independent of us in the way described by the former. The realist *seems* to be saying that the facts in question are by definition beyond our grasp. But this is merely a very strong kind of realism: to say that we do not determine the facts about meaning is not to say that language is essentially nothing to do with us.

4.2 Intentionalist Practices

How does intention relate to the participation in a rule-governed practice? Let us suppose, with Kripke, that participants in such a practice must all be rule-followers. If speaking meaningfully is a rule-governed practice, then to speak meaningfully is to follow the rules of the practice — an *intentional* activity. Of course, one may not be consciously aware of all the rules or able to list them, but one is at least not acting without *any* guidance — not making 'unjustified leaps in the dark,' as Kripke puts it Kripke (1982, p. 15) — but perhaps tacitly following rules of quite some complexity.³ However, in this section I would like to put pressure on this idea.

Let's take a game example. Football has *constitutive* rules without which it would be impossible to play the game: the rules make it the case that certain actions count as the practice of football-playing. As it happens, versions of the rules are written down somewhere, and although they can vary depending on the set-up, for now let's focus on one particular variant and call that 'football'. Now, the intentionalist is committed to the idea that this game, football, did not exist and consequently could not have been played before the rules were constituted by someone's mental content. That is: because playing football essentially consists in (at least) conforming to the rules of football, if those rules are constituted by mental content, and that mental content does not exist, then there are no rules, so no football, so no playing of football.

Yet it is clear that no particular player of football is required to hold these rules in mind as determinate content, whether in the form of knowledge or something else. The rules of football plausibly (perhaps necessarily!) say nothing about what the players must *think*, only what they must *do*: so the right action in the right place counts as playing football, even if the player in question turns out merely to be a convincing automaton with no inner life.

Of course, what we are describing here is *conforming* to the rules of football, but not *following* them. Still, there is no useful doubt here that if I am performing actions that conform, I am playing football. We can imagine all sorts of crazy things going on in my head while I am acting as I do — perhaps I think I am writing a symphony, flying through space, or preparing an omelette — but the right physical

³For tacit rule-following see in particular Millikan 1990, Peacocke 2012.

actions and responses at the right time are sufficient for playing football. And those actions may be very complex, not just running here, kicking there, but making strategic formations and replying to the captain's orders in kind. None of this is dependent on the contents of my head, as the possibility of replacing me with a sophisticated mechanical duplicate illustrates. What can be doubted is that I am playing football *intentionally*: but as we discussed earlier, the whole text of *Hamlet* could be produced without intention.⁴

Now if one player can do this, why is it not in principle possible that *all* players of football should not know or follow the rules? We can surely extend the case to the 22 players necessary for a single match. And why should that match not be the *first* ever match? Thus it seems, contra the intentionalist's thesis, that football may have been played before anyone knew or mentally sustained the rules.

This misses the point. Granted, a player can play without knowing the rules, or even intending to follow them; perhaps also, a whole two-team contingent of players could do so. But it still follows that one only plays football if there are rules *somewhere*. And the intentionalist will argue that those rules are at least partially constituted by some determinate mental content *somewhere*. In our example where the hypothetical first game of football involved no-one who knew the rules, we can recognise it as a game of football because *we* do know the rules, and in a sense bring the rules with us through a window of retrospect. These rules were not around then, but they are around now, and it is only with the benefit of a backward-looking analysis that the activity that occurred on that day counts as football.⁵

But why is *that* a problem? Some rules are time-constrained. 'Fasten seat-belts' on an aeroplane is a time-limited rule, or else we'd never get off it. Moreover, rules are invented: there were no rules about not keeping people as slaves in the 18th century. So of course when we look back we can say "he is not supposed to keep slaves," but this is inaccurate because the rules were not applicable then. Our thought that people were playing football before anyone invented the rules is mistaken for the same reason.

But hold on: that's more than a little bit controversial! Certainly there were no *laws* about not keeping slaves in the 18th century, but we like to think there was a little-acknowledged moral imperative not to do so, a moral rule that was in force. We think so because the rule in question is deemed *timeless*. It is not relative to a time or place or context because it is categorical in nature. If so, does this affect the football example?

What I'd like to suggest is that the rules of a game are similarly timeless — because in fact, all rules are. The slavery law was clearly invented, say in 1792, and consequently not applicable before its inception: but this will simply be reflected in the complete and timeless formulation of the rule, which will be indexed to the

⁴See §3.4 above.

⁵It might be objected that I have missed an obvious point, which is that this putative first game where everyone is acting non-intentionally was *not* football because it hadn't been invented then. However, I think this position is harder to defend than it may appear. It is natural to apply our own conceptual framework to the activities of those who do not share it — we may be reminded of Davidson's 'principle of charity' — and moreover, to think that it genuinely does apply there. It is not unreasonable to claim that the first person to boil rabbit, potatoes and rosemary in a pot of water was conforming to the rules for making a rabbit stew, even if the intention was just to clean the ingredients, or something else totally bizarre.

relevant temporal domain. For example: “It is illegal for S to keep slaves at any time after 1792”. This rule applies before 1792, though vacuously.

Still, it will be objected that we have missed the point again. The non-intentionalist may be right that the relevant rules are timeless, but their inception, persistence and validity at any one time is the product of human mental activities. The point is that at some point football was invented, and this invention made it the case for all past practices that they conformed or did not conform to its rules. But the rules themselves are sustained by the mental activities of someone or other.

So now imagine that a game called ‘football’ was invented by one person, Fred, who lives on a desert island. On day D Fred was feeling inspired, and he spontaneously composed, in his head, a list of all the rules of football. The intentionalist holds that it is by virtue of Fred’s mental content at D that anyone’s activities at or before D, anywhere in the world, must count as conforming to the rules of football, as *playing* football. The question now is this: if Fred suddenly died at D+1 after producing the rules, and no-one else spontaneously produced the same rules, could there be football at D+100?

I concede that this must be true in a sense, though we must be careful to say which sense. There is such a thing as a lost art, a practice whose rules and techniques are no longer known; and of course, it could be that the lost art was lost in a bubble, such that no-one now living knows or could ever know that it existed. There is a sense in which no-one outside the bubble could engage in such a practice, because no non-bubblers could intend to conform to (i.e. follow) its rules, because all *possibility* that they might gain knowledge of those rules is lost. It seems that Fred’s football is just such a case.

But as we have argued, following rules is not a necessary condition of engaging in a practice, for which, at least in some cases, non-intentionally-directed conformity is plausibly sufficient. I suppose we should call this a weaker sense of ‘engagement’ in a practice: but if it is not a constitutive rule of the practice that the practitioners intend to follow its rules, unlike in our football case, then the conformity of their physical actions to the rule is surely sufficient — think again of the substitution of automaton footballers. It is hard to see how the impossibility of the players having cognitive access to the rules, by tracking down and talking to Fred, could have anything to do with whether they conform to Fred’s rules or not. Concerns about the nature of time notwithstanding, how would giving Fred a crystal ball at D so that he can see the putative football players in the far future make the difference between their actually playing football and doing nothing in particular? What about if he has a crystal ball, but never uses it — or only looks at one putative football game, when in fact there are two to be seen? To me, this addition to the scenario just seems insubstantial. Perhaps the future is not any different from the past here.

Consider a case where the rules are in the past, and the putative rule-follower is in the present. An experimental archaeologist wants to know how a particularly unusual Grecian urn he found was produced, so he experiments with techniques in an attempt to replicate it. In doing so, he unwittingly conforms to the rules of a secret long dead urn-making practice, ϕ -ing, that was known only to the original craftsman. Or rather, to avoid begging the question, we should say: if he had acted in the way he does now while the craftsman was alive, it would certainly have conformed to the rules for ϕ -ing. Now the question: does the archaeologist ϕ

in the present? The craftsman is long dead, and kept no record of his rules. Yet if he were present, he would recognise the archaeologist as engaged in ϕ -ing.

If the intentionalist concedes that the archaeologist does ϕ , then it would seem that the constitutive rules of the practice are themselves constituted by the mental content of a dead man. This seems unlikely — the dead are not known for their cognitive abilities.

Rather, it seems the intentionalist should deny that the archaeologist actually does ϕ . This means either a) rejecting the timelessness of the constitutive rules or b) denying that conformity is sufficient for ϕ -ing. However, both of these options face difficulties. Option (a) means either claiming that no rules are timeless, to which I have presented counterexamples, or else making a case for thinking that the craftsman's rules in particular expired with him. The latter does *not* mean saying that the rules for ϕ -ing were indexed to some period in his lifetime, for in this case they would have a timeless reach. The rule 'to ϕ , S must F between t and $t + 1$ ' still applies at $t + 2$, and simply says that from that point onwards S cannot ϕ . Rather, the claim would have to be existential. The rules for ϕ -ing actually cease to exist when the craftsman dies. We will explore this possibility, which I consider doubtful, in §4.3.

Similarly, option (b) means either claiming that no practices can be practised by merely conforming to their rules, to which I have presented counterexamples, or else making a case for thinking that this is at least true of ϕ -ing. Yet it is tough to see how the latter could be achieved without simply stipulating that ϕ -ing requires rule-following, an *ad hoc* move that can be easily countered.

4.3 Conforming and Following

Still, the idea that one might engage in any practice just by exhibiting the right external behaviour is an uncomfortable one, no matter how complex that pattern of action. Some kind of intentional relationship is needed to connect the agent to the practice. However, merely appealing to the intention 'to conform to the practice' looks like the intensional fallacy: it is not sufficient for my following the unknown rule that produced a number sequence that I intend to conform to 'whatever the rule is' and happen to do so. That cannot be following a rule, since it is comparable to a sort of blindness. I might intend to get through a maze via the fastest route, but if I then also happen to take that route by chance it would be wrong to say I *followed* it in the relevant sense unless I also knew what that route was.

So if S had somehow heard that Fred had a practice with some determinate rules, it would seem insufficient for following those rules that S intends to conform to just 'whatever those rules were' while happening to act in accordance with them. The fact that he has heard of Fred makes no difference.

However, something has been overlooked here. What if, in line with his intentions, S *guesses* Fred's rules? What if, moreover, he guesses *correctly*? In the number sequence case, someone asked for the next n will tend to make an educated guess based on certain intuitive principles, but as Kripke's discussion established, no-one presented with just '5, 10, 15, 20, ...' can establish with certainty that this was produced by the rule $n + 5$. Yet when someone confidently answer that the next n is 25, we would like to say (sceptical thoughts aside) that he is following just

that rule. The question now is what makes this different from someone who looks at the numbers, is completely baffled, but intends to conform to ‘whatever rule produced them’ and arbitrarily elects to answer ‘25’.

One difference is that the former person intended to conform to both ‘whatever the rule is’ *and* the rule $n + 5$. Conversely, the person who guessed did not have the intention to conform to the rule $n + 5$ even though he incidentally did. Consequently, he did not *follow* $n + 5$.

However, things get complicated if we imagine that the latter person guessed the rule was $n + 5$, formed the intention to conform to it, and answered ‘25’. Perhaps he did so at random, without drawing any connection between this rule and the number sequence. Then we should have no choice but to say that, although this was not an *educated* guess, this person still followed both the rule $n + 5$ and ‘whatever rule generated the sequence’.

It follows that in the future football case, if S guesses Fred’s rules correctly and follows them, then he is playing football. Moreover, he may actually have never heard of football or Fred: perhaps he just guesses correctly that there was some practice with such-and-such rules.

It will be objected, of course, that S is now the person whose mental content sustains the rules and practice of football. He has not rediscovered football, but merely invented it separately and in coincidentally just the same way, as the practices and tools of agriculture were by tribes of ancient humans having no contact with one another. But then, S certainly does not think this is what he has done! He believes, perhaps irrationally, but perhaps also truly, that he is playing an ancient game with long-forgotten rules.

This is now a close analogue of the Grecian urn case. The archaeologist guesses the technique for producing the urn, and by replicating it in practice he seems to engage in ϕ -ing, a practice of which he has no knowledge with rules he does not know.

It seems that rules cannot be both timeless and constitutively dependent on the mental content of finite beings. Yet it is likewise difficult to resist the thought that the archaeologist succeeds in ϕ -ing despite the absence of any mental content to sustain that practice. The thought that the mental content existed at least once in the past does not seem to help: what difference does it make? Perhaps the archaeologist has no idea that there was such a special secret practice as ϕ -ing: instead, he is trying to replicate what he takes to be a pottery technique common among the masters but of which his specimen is the only remaining example. The proper conclusion seems to be that such practices are *not* dependent on mental content for their subsistence.

But *again* it will be objected: the only reason that *we* think the archaeologist counts as ϕ -ing is that *we* know about the practice. If we were not privileged to this narrative information, we would not think it!

Perhaps we would not. But does this change any facts about the scenario? Again, as in Kripke’s sceptical challenge,⁶ the itching temptation is to say that a lack of knowledge counts for a lack of facts. But why should we think that our observation of what the archaeologist does as ϕ -ing is what *constitutes* his ϕ -ing? If we do not observe him, does that really make it the case that he does not ϕ ? What if we look, but are somehow physically prevented from assessing whether

⁶See my chapter 7

he is ϕ -ing or not? Clearly our blindness would not be enough to change the *facts* about whether he does ϕ or not.

Perhaps what is important is that someone who knows of the practice *could* observe the archaeologist's activities under normal conditions. Thus if it is the case that if we were to observe his actions then we would observe this as ϕ -ing, then this is sufficient for its being the case that he actually does ϕ .

But again, this seems to place a strange and unnecessarily heavy burden on the observers of an activity to determine the facts about what that activity amounts to. Assume that it is not a rule of the ancient potter's practice that something only counts as ϕ -ing if someone who knows the rules *judges* it to be so: then why should the archaeologist's ϕ -ing or not be contingent on the actual or potential judgments of actual or potential observers?

The objector points to our involvement as observers because we know the rules of the ancient practice, but it is not our knowledge of or judgment about this practice that matters. Rather, it is the fact that the rules are held in our minds as a point of reference. We do not have to be present to witness the archaeologist's efforts, we do not have to judge whether they count as ϕ -ing or not, but we do have to mentally sustain the practice of ϕ -ing if anything is to count as ϕ -ing.

What is it to mentally sustain a practice in the requisite way? If the statements just made are true, then it seems the practice could be sustained by a brain in a vat, so long as it knew the rules. So imagine that, through an admittedly unlikely B-movie-like turn of events, the ancient potter's brain is replicated from traces of his DNA, with all its mental content still intact, including the rules of ϕ -ing. Now suddenly the practice exists again. Moreover, perhaps we only replicated enough of the brain for it to function minimally and retain the relevant items of knowledge. Still, this should be sufficient for the reinstatement of the practice.

I could go on, but I suspect this picture is already ludicrous enough. Why would the operativeness of a rule be contingent on the existence of a certain configuration of grey matter in a tank somewhere? It seems that a piece of paper with the rules written on it would do just as well.

So let's change our example to remove actual mental content from the picture entirely. The rules that constitute the practice of ϕ -ing now consist entirely in what is written on a certain piece of paper that the ancient potter left behind (obvious anachronisms notwithstanding). However, as before, the piece of paper is not a constitutive part of the practice. Rather, the paper merely 'bears' the rules that do constitute the practice. Now it seems we can make sense of activity that conforms to these rules after everyone who knows them is dead.

It seems that what matters about having this piece of paper rather than not having it is that somebody could read it. Yet this is not a condition we specified. Perhaps instead the paper is locked in a box, and because it is so old, any attempt to read it (even via penetrative imaging techniques) would cause it to crumble instantly to dust. Put simply: imagine that there is no possible way to read the piece of paper, though it is in fact such that if someone *were* to read it, they would learn the rules of the practice. I suggest that in this situation, we are still happy that if the archaeologist makes his urn in the right way, he will count as ϕ -ing due to the fact that it is logically possible (though not *actually* possible) for someone to read the rules.

This is odd, however. We are not interested in the actual or potential judgments

of anyone about the practice, nor their actual ability to access the rules. All that matters is that there is some intelligible record of them.

Now what if the rules were written by the potter in some kind of idiolect? Now they are intelligible only if we have the potter's mental rules available to interpret them. That is: the rules remain in concrete record, but they are not conceivable intelligible in actuality because the standard by which they must be decoded (assuming they are otherwise indecipherable) is lost forever. Yet on reflection it seems, somehow, that we might *still* be happy that their concrete existence on paper is sufficient to sustain the practice!

This is very odd, but it is also telling. For of course, the thought that the written rules have meaning if only we could get hold of the no-longer-existent mental content for interpreting them as such reveals that we think practices are nothing to do with hard records or actual mental content: we think that the paper has meaningful things written on it when there is no currently existing record of the rules that could determine whether any given interpretation of what it says is correct or not.

What this seems to suggest is that the metaphysically possible involvement of someone with the right mental stuff is sufficient for a rule's being in force, even if such involvement is not possible under current circumstances. But this is tantamount to permitting that because there is no necessary obstacle to the potter's being around at the same time as the archaeologist, the rule counts as being in force even though he is dead and in fact has no mental stuff for the rules to consist in.

What is important is apparently just that somebody's determinate mental content constituted the rules *at some point*. But is that really what this entails? We seem to be helping ourselves to some view of individual times as locations in 4D space, so that in some sense the potter still exists, but in the past. Just as the spatial proximity of the potter to the archaeologist has no effect on whether something can count as ϕ -ing, so (we think) the *temporal* proximity of the two is also irrelevant to that matter. Yet that claim is dubious. The attendant symmetrical supposition that one can conform to rules *yet to be* invented is troubling, as it raises familiar concerns about temporal asymmetry: perhaps the future is indeterminate though the past is determinate. So maybe all we should really commit to is the idea that for ϕ -ing (or at least conforming to the rules thereof) to be a possibility at t , the constitutive rules of ϕ -ing must exist no later than t .

Still, in more pragmatic terms it does seem like the important thing about the potter's having existed is not that both he and his mental content in some sense persist, but that he *could* have been around at the same time as the archaeologist, even if this is a non-actual scenario. That is, the point is essentially modal rather than temporal. But if so, then we open up a whole new can of counterfactual worms. For imagine now that there never was a potter, but that there could have been, and that there never was a secret pot-making practice called ϕ -ing, but that there could have been. The next question we might ask, assuming the archaeologist's actions remain unchanged, is whether what he does *actually* conforms to the rules of ϕ -ing, or whether it merely *would* have conformed to those rules if there *had been* a potter who created and mentally sustained them.

However, this is a red herring. Whether some action conforms to a non-actual rule, a rule that is not actually 'in force', is nonetheless a matter of actual fact. If the

possible rule is that S ought to F, and S in fact does F, then the proper conclusion is that S conforms to a possible rule, not that S possibly conforms to a possible rule. An essential part of Kripke's reasoning is just that any action actually conforms to an infinite number of rules, whether anyone goes so far as to entertain those rules or not. The sceptical challenge is clearly not dependent on the fact that the sceptic *actually* proposes an alternative hypothesis for which rule it is that I follow when I use 'plus', but on the possibility thereof. Yet if we insist that rules are sustained in some way by mental content, then the most we can say here is that this infinite number of rules is a set of *possible* rules, any of which we *could* entertain.

Consequently, it follows that the archaeologist's behaviour satisfies all the non-actual rules of ϕ -ing, and as such, he does indeed ϕ .

4.4 Non-intentionalism

The principle objection to this conclusion is that our example supposes, implausibly, that to conform to the rules of a practice is sufficient for engaging in that practice. It will be objected that even if the constitutive rules of a practice can be 'in force' without anyone's mental content, we should hold off on the supposition that conforming to them will be sufficient for engaging in the practice in all but the most trivial cases; for normally, to engage in a practice involves *intentional* engagement, *rule-following*, not the mere coincidental satisfaction of such-and-such conditions.

The supposition here is that in a game of football, if we discovered that each of the team members were acting more or less at random and had no idea of what they were doing, then we should conclude that despite appearance they were *not* playing football. Rather, they were acting in a way that looks exactly like football-playing to someone ignorant of the contents of the minds of those involved. No matter how complex the activity, it can be observed that if the actor does not intend to engage in it, then he does not. Certainly this raises difficult questions with regards to artificial intelligence — if we cannot ascribe intention to what looks like a football-playing robot, does it play football? — but the objector might simply defer that question, confident that in the human case intention is necessary.

At first glance this objection seems impressive, but it relies on an overly simplistic view of what it is to engage in a practice. To engage is to follow the constitutive rules of the practice, it is assumed, so to lack the relevant intentions is to fail to engage. However, this places quite a substantial intentional burden on practitioners. If a practice is conceived as a complex of constitutive rules, then on this view, to engage in the practice is to follow all the rules. Yet many real-world practices are likely to include many rules of which the practitioners are not even aware. For example, if speaking English consists in following all the syntactic and semantic rules of English, then it is debatable whether anyone speaks English at all — not for Kripke's reasons, but simply because to follow (intend to conform to) all those rules would seem to entail knowing them all in some sense or other, and that seems doubtful.⁷

Perhaps we should draw a distinction between the properly constitutive rules of a practice and its 'supplementary' rules. It might be contended, for example,

⁷Thanks are due to my colleague Jonathan Wolf for suggesting this example.

that although it is a constitutive rule of chess that bishops move diagonally, the *en passant* rule is less fundamental to the game. This would explain why complete ignorance of this rule, characterised by surprise and doubt on the player's part when his opponent makes tactical use of it, would not normally prevent one from playing chess. In a case where only one player knows the rules, the difference between them would not be in which games they are playing, but in their knowledge of and ability with respect to one and the same game.

By contrast, we might want to say that someone who does not know that bishops move diagonally, or *a fortiori* does not intend to conform to this rule, does not play chess. However, this is less obvious than it sounds. Imagine a very poor chess player who never moves his bishops, but knows how he should move them if he wanted to. Now imagine a putative chess player who does not know how to move his bishops, so he does not. Both conform to the rule 'If you move your bishop, you must move it diagonally'. However, the contention here is that the former plays chess, but the latter does not, because he has no knowledge of this rule, or at least because he does not follow it. If he plays a game at all, then it is something other than chess.

This seems absurd: ignorance of this rule is not sufficient for failing to play chess, and manifesting one's ignorance by performing an illegal move is more readily characterised as an error, even if one does not recognise it, than as a revelation that one was not playing chess after all. Perhaps instead it should be contended that to engage in a practice is to follow some minimal number of its constitutive rules — or even, that the bishop-rule is not one of the relevant kinds of player-constituting rules of chess. Perhaps we want to say that there is a very small subset of very general rules that must be followed; that failure to do so is failure to engage in the practice (in playing chess).

This seems more plausible, though it is not obvious in the case of chess which rules should be considered necessary in this way, or whether merely following 'enough' rules is sufficient for playing chess rather than having one's actions coincide with those of a chess player. In the 'enough' case, it is also quite plausible that playing chess is a vague concept; that there is no sharp cut-off for the number of rules that must be followed by S in order that S plays chess.

Still, the question at hand is whether something like this applies to the case of meaning. Granted that meaning-rules could exist irrespective of anyone's mental content, would meaning itself be possible if people only conformed to meaning-rules as opposed to following them? In §3.5 I argued that it would. I argued that it is sufficient to mean *green* by 'green' if one conforms to the rule that determines whether one means *green* by 'green' or not, and so on for all other meaning-determining rules. The implicit supposition was that this would work for meaning as a whole: that for any instance of meaning, where meaning is something like the familiar notion of reference, it is sufficient for meaning such-and-such by one's words that one conforms to the rules for meaning such-and-such by them, and that these rules make no mention of intention. To use a familiar metaphor, if meaning something is like pointing at the referent, then I am arguing that to point at the referent does not involve intention on the part of the pointer, but is a function of the position and angle of his digits.

This position may be tenable within limits, but its downfall is that for certain cases we will want to permit that the speaker's intentions do modify the meaning.

Perhaps we can countenance that there is some sort of standard meaning, a pre-intentional meaning of one's words that can be subsequently modified — but even if this were so for all possible utterances, it should still be the case that the meaning can be modified by the speaker's private intentions. I countered this could be accommodated by the non-intentionalist by making it a condition on meaning-modifying intentions that they be *voiced* by the speaker or otherwise made public; then the meaning is still a question of what is *done* rather than what is privately *thought*, where 'doings' do not essentially involve intention.

However, this still violates a certain intuition that even if the speaker's intentions are private, they still ought to have some determining effect on the meaning of his words (or lack thereof). The idea was to play on the intuitively shared nature of language, and on the Fregean concern that the mere *Vorstellungen* I might entertain are not meaning determining. But the fact is, if meaning is determined by my *following* a rule, then intentions are involved. It may be that I conform to the rule for meaning *green* by 'green', and that to my associates it would be impossible to determine that I did not in addition *follow* this rule without my attesting to having lacked the relevant intention. But the fact remains that I did not follow that rule, even if I gave the impression of having done so, and it is not totally implausible that my associates could subsequently discover, after having decided that I meant *green* by 'green', that I in fact did not, for I failed to follow that rule. This does raise some important questions however: in the case where a whole text is produced by the random character output of a computer, this view commits us to the text not being meaningful, even if it might appear so; and to the idea that a human producing the same artefact by following the right semantic rules would thereby render it meaningful.

As before, I would want to challenge this idea. The difference between the non-intentionally generated text and the intentionally generated version is not a difference in the text itself. If Fermat's marginal notes were not intended to convey a theorem but as mere handwriting practice, then it seems they would have conveyed his theorem nonetheless. The difference in these cases is not in the product, but in its source, yet the question of meaning is about the product, not about its source. Rather, the question about the source is a question about *intended meaning*.

In this respect, I am committed to the idea that meaning is not so much like following a rule after all, since it is a property of things in the world like texts and utterances that is not essentially determined by any connected intentions. Still, to *intend* to mean something surely is; and so, perhaps, is to *understand* something in one way rather than another. In this way we have arrived in a more or less circumspect fashion at an important thought: how, if meaning is not an essentially intentional phenomenon, is it to be conceived as essentially normative? We are not in a position to ascribe obligations to our ASCII character generating computer, based solely on its having output some string of characters rather than another. Nor does the output itself incur any obligations. In fact, I would suggest that the idea that meaning as here conceived has very little to do with normativity, since the traditional normative element relates to intentions and rule-following.

Rather, the normativity of meaning would seem better described as the normativity of *intending to mean*. It derives properly, as Kripke says, from "the relation of meaning *and* intention to future action" (1982, p. 37, my emphasis). Trivially,

insofar as one holds an intention to mean one thing rather than another, one is bound to certain courses of action. This can be weakly characterised as a hypothetical ought, on the hypothesis that one aims to satisfy one's intentions — but this seems unnecessary, given that it is part of an intention to wish it fulfilled, and perhaps uninformative, given that 'to aim that one's intentions be satisfied' just sounds like 'to intend that one's intentions be satisfied'. So perhaps it is better thought of as a categorical obligation: one ought to act as would best satisfy one's intentions. But as mentioned, this also appears trivial. What kind of an obligation would it be to act as one wills? Or even, to act as necessary to satisfy one's will?

There may be genuine questions about the normative force of a prescription to act so as to satisfy one's intentions. However, it is clear that the normative dimension of Kripke's problem derives precisely from this kind of prescription. The challenge, so it seems, lies with the determination (in at least the epistemological sense of *discovering*) of what exactly it is that you ought to do, if you are to remain in keeping with your previous intentions. That is: you intended to mean something or other by 'plus' in the past, so just on the hypothesis that you intended to always mean the same thing by it, should you say that '68 plus 57 equals 5', or 'equals 125'?

As suggested (§7.3), the epistemological question here is no different from any other regarding cognitive access to 'oughts'. The normative dimension of rule-following presents exactly the same epistemological challenges that the normative dimension of ethics does, or indeed, of the normative dimension of many other aspects of human life. The idea that there are norms of belief and norms of assertion, for example, has gone to seed in more than a few quiet corners of philosophy — an illustrative case being Terence Cuneo's project of legitimising moral obligation by drawing a direct analogy with the comparative legitimacy of epistemic obligation in his *The Normative Web* (Cuneo 2007). Problems associated with ascribing knowledge of these norms to ordinary epistemic and assertion-making agents, and questions about the constitution of the 'oughts' in question, are just the same questions we should expect to face in the broader case of rule-following.

That is, unless we consider the possibility that rule-following may be fundamental to these lesser activities; that underpinning the normativity of moral facts, or of epistemic and assertoric warrant, is the normativity of intention that underpins the 'oughts' of rule-following. If a case could be made for thinking this, then it may turn out that pinning down the 'oughts' of rule-following is a more pressing, and perhaps also a more difficult challenge. It may be argued that without a foundation of rule-following, the notion of normativity does not make sense; in which case, what hope would there be for the possibility of rule-following, which is itself a normative activity? We would be caught in an impossible circle.

As noted, the challenge derives from the assumption that rule-following is the source of normativity, which is far from an established fact. For a start, we should be willing to accept that normativity instead derives from some inner impulse (perhaps intention) to fulfil some course of action, to which rule-following is but one possible enabling mechanism; we must beware of mistaking one possible enabler of normativity for its source. And if, as may be the case, we think that intention is the source of normativity in the rule-following case, then I would also dispute that this is sufficient to create a problem of the kind Kripke describes. I pursue that line of enquiry in chapter 7.

Chapter 5

Primitive Normativity

It seems that we need, one way or another, to involve the mental lives of agents in the normativity of meaning and in rule-following. As such, this chapter focuses on Hannah Ginsborg's development of a positive account of 'primitive normativity' in response to Kripke's sceptical challenge. The account can be understood as an attempt to reconcile a kind of dispositionalism with Kripke's challenge by locating normativity within an agent's primitive disposition to, in a sense, *feel* that things are right. Development of this account has taken place over several years, but I take the centrally relevant works to be Ginsborg 2011a,b, 2012.

I argue that while primitive normativity has its appeal, that appeal must be shown to clearly outweigh certain less-appealing characteristics. The chief among these are the seemingly anti-realist trappings of the account, and the question mark that hangs over its alignment to Kripke's own expectations for the normativity of meaning.

5.1 Dispositions Reconsidered

Ginsborg (2011b) sees the upshot of Kripke's discussion as "not merely the epistemological conclusion that you do not know what you meant, but the metaphysical conclusion that there is no fact about what you meant" (2011b, pp. 227–228). Of the putative solutions to the challenge that Kripke considers, Ginsborg focuses on two: the 'dispositional' solution, which holds that if you mean *plus* rather than *quus* by 'plus', this is because you are to be disposed to use 'plus' to denote the plus function, and so to answer '125' to '68 + 57'; and the 'primitive state' solution, where meaning *plus* by 'plus' is an irreducible *sui generis* (mental) state. She notes that Kripke focuses far more on the former, being relatively dismissive of the latter. Her hope is to articulate and defend a 'middle way', with an account of 'primitive normativity'.

Ginsborg's (2011b) response to Kripke's sceptic introduces and develops this concept. Aside from the need for a middle path between naked dispositionalism and anti-reductionism, Ginsborg is also significantly motivated by her disagreement with an assumption she attributes to Kripke's sceptic. She argues that the Kripke rule-following problem arises when we assume that the appropriateness of a given utterance — why it is what *ought* to be done — depends on appeal to an antecedently applicable rule (Ginsborg 2011b, p. 234–5). Her view, by contrast, is that the 'oughts' in question are primitive; the fact that I ought to answer "What

is $68 + 57?$ ” with “125” is not dependent upon the ‘plus’ rule, nor upon any rule I was previously following, nor upon anything more than my disposition to take my action as appropriate. That I *ought* to answer “125” consists in its being in this way primitively appropriate to the context to answer “125”.

She begins with a run-through of Kripke’s argument, identifying that he considers two alternatives prior to embarking on his own ‘sceptical solution’, of which the former is, famously, a kind of dispositionalism, and the latter is, less famously, a kind of *sui generis* account of the state that constitutes our meaning something by an expression, which we might label ‘anti-reductionism’.

Kripke’s arguments against the possibility of a dispositional account of meaning derive from a) the ‘finiteness of our dispositions’; b) the fact that we are sometimes disposed to make mistakes — the dispositionalist must provide a means for distinguishing between divergence and error; and finally of course, c) from the fact that meaning is normative – the source of the now-(in)famous dictum that “whatever in fact I (am disposed to) do, there is a unique thing that I *should* do” (Kripke 1982, p. 24). These points have all been extensively documented, disputed, accepted and rejected within the enormous literature on Kripke’s paradox. I note in particular that Hattiangadi’s 2007 commentary on the matter in *Oughts and Thoughts* denies the applicability of (a), accepts (b) as decisive against dispositionalism, but considers the wider force of Kripke’s problem (i.e. extending beyond dispositionalism) to derive from (c).¹

Ginsborg, however, suggests that both (a) and (b) can be addressed by an ‘ideal conditions’ account of dispositions, having examined this possibility previously (Ginsborg 2011a). However, she doubts that citing what we would say under ideal conditions can properly account for (c), the normativity of meaning, given that the ideal operation of, say, a thermometer, is not a normative matter:

“The idea that someone is responding or not responding as she would respond under ideal circumstances does not on its own license the idea that she is responding, or failing to respond, as she *ought*. We do not say of a sample of salt that dissolves or fails to dissolve in water that it is doing or not doing as it ought, but only that it is or is not manifesting its disposition to dissolve.” (Ginsborg 2011b, p. 229)

This is one of several reasons why writers on the topic have thought appealing to how we would use a word under ideal conditions would fail to capture the normativity of meaning. The depth of the problem becomes apparent if we follow Zalabardo in considering Kripke’s normativity of meaning thesis to reflect not only the thought that if I mean *plus* by ‘plus’ then ought to say “125”, but also that this meaning-fact both presents a conscious guide to my saying “125” rather than “5” and also serves to justify my doing so. Facts about what I would do under ideal conditions just don’t seem accessible in the right way (Zalabardo 1997, §IV).²

Still, Ginsborg is no more optimistic about the anti-reductionist alternative Kripke considers, i.e. presenting a view of meaning or understanding (or, presumably, intention) as a *sui generis* state. Kripke’s treatment of this option is brief, but he expresses a concern that such a meaning-state or understanding-state would

¹I consider Hattiangadi’s case in more detail in chapter 6.

²Cf. also §1.6 of the present work.

not play the *justificatory* role required of it in our picture of rule-following, and similarly, that it would not play the corresponding *explanatory* role in that picture. A move in this direction, he says, “leaves the nature of this postulated primitive state – the primitive state of ‘meaning addition by “plus” – completely mysterious” (1982, p. 51). This is far from presenting a knock-down argument, but Ginsborg concedes that where the anti-reductionist position surpasses dispositionalism in its ability to preserve the normative dimension of meaning, it does seem to fail in at least its explanatory task. As such, she proposes her own account of primitive normativity as a ‘middle way’.

This account takes the shape of an augmented dispositionalism. Since Ginsborg considers the normativity of meaning to be the dispositionalist’s most significant obstacle, her account is presented as a remedy. However, the most significant aspect of her position lies in a refusal to derive the normative impositions of meaning from our conscious recognition of meaning-rules — or indeed *from* anything. The normativity in Ginsborg’s account derives from a *sui generis* mental attitude, that of taking something to be ‘appropriate’, where this ‘taking’ is not parasitic upon any suspect mental content like representations of patterns, rules or intentions. The idea is developed in the context of childhood language-learning, where it seems a child can be trained to take sorting green things together to be appropriate without having mastery of the concept ‘green’, or to treat ‘10’ as next in the sequence ‘2, 4, 6, 8’ without any mastery of addition or ‘going up in twos’. In continuing on in the same way, we take ourselves to be doing what is simply ‘appropriate,’ even if no explanation is forthcoming as to *why* we think this, or by virtue of *what* our actions count as appropriate.

Ginsborg thus holds that we ought to answer the question “What is $68 + 57$?” with “125” simply because we are disposed to take this to be primitively appropriate, and the relevant ought derives from recognition of primitive appropriateness. Similarly, she takes the fact that we mean *plus* rather than *quus* by ‘plus’ to consist in our being disposed to give the sum rather than the quum, and in our taking this to be primitively appropriate. The involvement of rules in this picture is evidently tertiary; it is emphatically inessential to Ginsborg’s view that what is appropriate accords with the rules of addition, or more generally, any rules at all. The sense of appropriateness the subject feels is not derived from the demands of rules or mental content, e.g. linguistic intentions, previous meaning, ‘instructions I previously gave myself’; nor from the subject’s recognition or knowledge thereof. In summary:

I reject the assumption – which is of a piece with Kripke’s assumption about how the skeptic’s two challenges are related – that the “ought” has to be conditional on your past meaning or past intentions, or on a rule which you previously had in mind for the use of the term. (Ginsborg 2011b, p. 232)

In light of this, Ginsborg (2011b, p. 242) offers an interesting distinction between two kinds of normativity. On the one hand, that one “ought” to ϕ is taken to imply that one “has reason” to ϕ . On the other hand, she describes a thinner sense of normativity couched in terms, for at least the linguistic case, of how some expression “ought to be used” — the idea being, it seems, that the correctness

of some type of usage comes bundled with some related ‘ought’. While unfortunately she does not go into enormous detail here, it is at least clear that Ginsborg’s primitive normativity is not of the former, intellectual, ‘reason-giving’ variety.

She accordingly rejects two further alternatives. The first view is: that the subject’s sense that her actions are appropriate derives from her implicit knowledge of a rule to which they conform. Since the subject’s sense of appropriateness is located at the personal rather than subpersonal level, Ginsborg argues that we then face the impossible task of explaining how implicit (i.e. subpersonal) knowledge of the rule in question could inform or “underwrite” such an attitude, when the whole point is that “the rule is viewed as a representation to which the subject’s cognitive system, but not the subject herself, has access” (Ginsborg 2011b, p. 239). Ginsborg considers this objection decisive, but I suggest that there is something of a tangle here: her challenge seems justificatory in nature, a worry that implicit knowledge could not justify our feeling that “125” is an appropriate thing to say since we do not personally have access to it. But presumably the reason anyone would involve implicit knowledge in this picture is that they consider *this* to be the reason why we ought to say “125”, not the the fact we take it to be appropriate. That view would render the attitude of taking-to-be-appropriate an idle cog in the machine, but it is not immediately clear why this should bother us.

The second idea Ginsborg rejects is that our sense of appropriateness derives from a sense of primitive ‘similarity’, such that we consider (e.g.) some uses of ‘plus’ to be ‘just similar’ and others not. However, she argues that if we take “125” to be primitively similar to other uses of ‘plus’, then the question remains whether we ought therefore to consider it appropriate — or more broadly, whether the primitively similar use of ‘plus’ is the one we ought to make. Even if we concede that we always take those uses of ‘plus’ that we take to be primitively similar to be primitively appropriate, it cannot be our contentful *recognition* that these uses are primitively similar that causes us to take them to be primitively appropriate, or we will suddenly owe an explanation of why *this* — the primitive similarity — is appropriate. It seems we must either appeal to a rule, raising the familiar problems, or to primitive appropriateness, proving that this is explanatorily prior to primitive similarity.

However, whether we take a given response to be primitively appropriate will depend, as a matter of our hard-wiring and receptivity, on our previous usage of the word ‘plus’; if I am disposed to give the quum rather than the sum, *and* disposed to take the answer “5” to be primitively appropriate, then it would seem both that I mean *quus* by ‘plus’, and also that I *ought* to say “5”, since it is this response that I am disposed to take to be primitively appropriate. Ginsborg argues that this, intuitively a case of a malfunctioning disposition, will be ‘trumped’ by the prevailing *communal* sense that to answer “5” is *not* primitively appropriate, and that “125” is primitively appropriate. This, it seems, is the wider (and perhaps also ‘higher’) sense in which “125” is the answer anyone *ought* to give.³ Note also

³I am extrapolating from Ginsborg’s claims here, but it seems we should not consider this wider intuition to make it the case that someone who is disposed to answer “5” *ought* instead to say “125” in the higher level sense of ‘providing a reason’ why that person should do otherwise. Rather, it would seem that what she has in mind is that this person’s disposition to take certain actions as primitively appropriate would be causally shaped by the rejection of the community, such that a severe reprimand would alter her disposition to instead take “125” to be primitively appropriate — resulting in its being the case that *this* is what she ought to say.

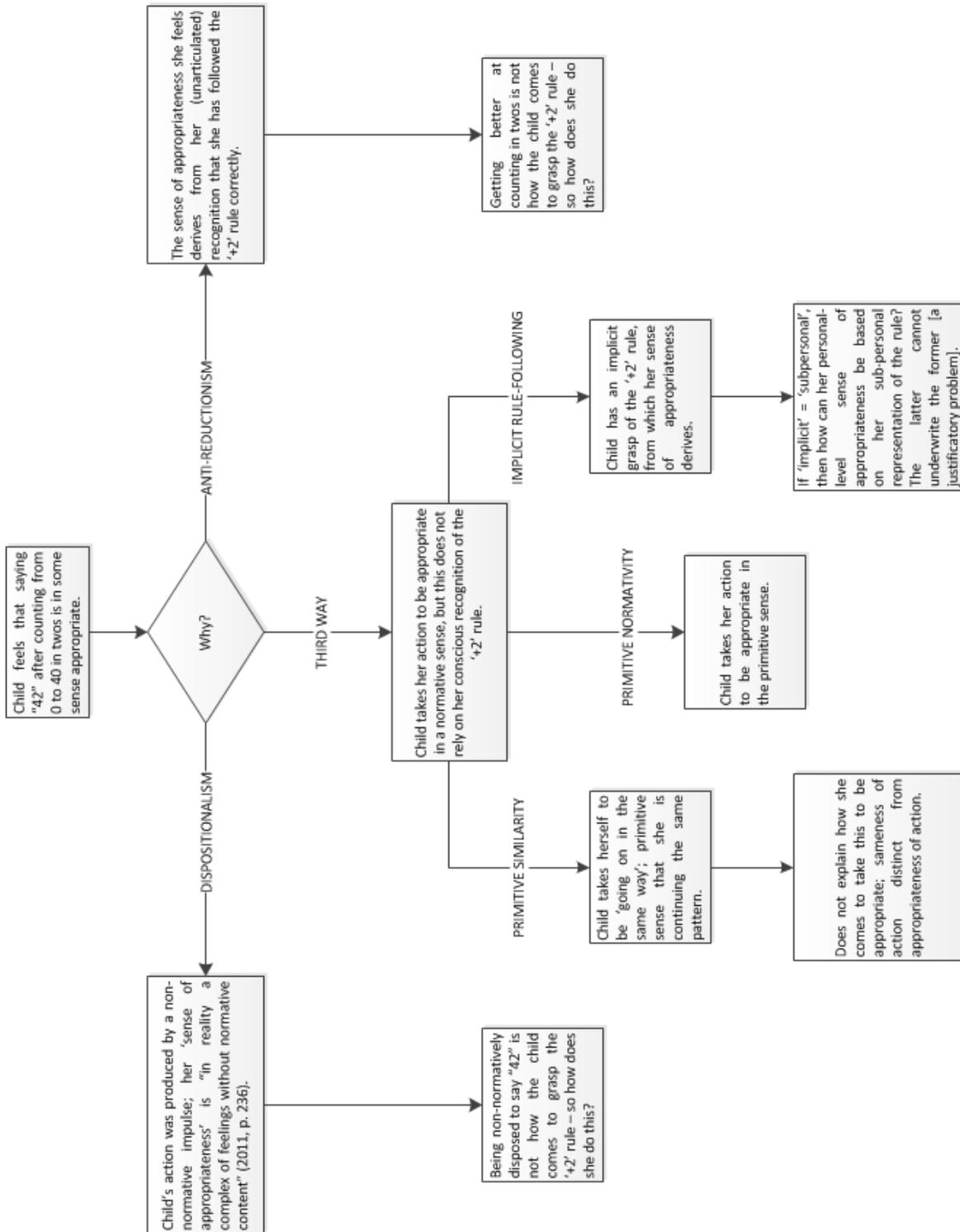


Figure 5.1: Ginsborg's discussion of appropriateness

that since a primitive ought is the product of our holding the ‘appropriateness’ attitude towards certain actions, our lacking that attitude towards say “5” does not entail that we ought *not* to say “5”, but only that it is not the case that we ought *to*. Ginsborg does not explore whether there might also be primitive ‘ought nots’ — that is, primitive inappropriateness — presumably because discussions of the normativity of meaning are couched in purely positive terms.⁴ This is one of many perhaps surprising commitments of her account.

5.2 Some Key Commitments

What might strike us as Ginsborg’s most troubling commitment is to the priority of normativity to meaning; that is, that the facts about what I mean are a product of the facts about what I am disposed to say and whether I take this to be primitively appropriate. This runs counter to the more traditional interpretation of meaning’s normativity as a product of the dictates of rules, intentions, or simply semantic facts. Compared with the traditional view, Ginsborg does not seem to be painting a picture of the normativity of meaning so much as denying that meaning is normative; what we learn seems to be that meaning itself is *not* normative. However, it is unfair to assume so simply because the normativity here does not derive from meaning itself; after all, normativists have often argued that the normativity of meaning derives wholly from its being necessarily bound up with conditions of correct use. Correctness is normative, so meaning is normative — or so the argument goes.⁵ If we should be concerned by this commitment, I suggest it should be with respect to its further implications downstream.

Ginsborg acknowledges that her account “makes no mention of the idea that meaning guides, justifies, or instructs us in the use of the word” (Ginsborg 2011b, p. 244n22). The normativity in question, if it is to be considered the normativity of *meaning*, is thus not of the familiar intellectual variety, the guiding hand of the rules of meaning that leads me to say *this* because *this* is what I ought to say. It is not normativity that *shows* me *what* to do, but acts rather more as a prop to what it is that I *will* do. If I am disposed to answer “What is $68 + 57$?” with “125”, then the claim that “125” is what I *ought* to say does not derive from facts about addition, nor from facts about my relationship to the concept of addition, nor from my intention (in Kripke’s parlance) to ‘denote’ addition; but solely from my disposition to take this answer to be appropriate in the primitive sense. Of course, this diverges not only from what might be called the ‘traditional’ conception of the normativity of meaning, but also from Kripke’s — as Ginsborg acknowledges (2011b, p. 244n22). She argues in her defence that, all the same, aspect (c) of Kripke’s argument against dispositionalism and anti-reductionism is not weakened by interpreting the normativity of meaning this way.⁶

I cannot discuss that particular issue here, but even if Ginsborg is right, I suggest this is rather swift. At the very least, Kripke takes himself to be presenting a

⁴The attempt to couch the normativity of meaning in purely negative terms is ill-advised, simply because specifying exactly how a term ought *not* to be used is not sufficient to establish that what uses remain correspond to how it ought *to* be used. All we learn is that these uses are permissible; the positive oughts Kripke seems to envisage cannot be derived in this way.

⁵Cf. chapter 6 below.

⁶Cf. Bridges (2014, §2) for more detail on this.

problem with rule-following — a problem with the idea that my actions are justified by some rule or other, whether arithmetical or semantic — so to reject that idea out of hand is to risk failing to connect with the debate. This is not to say that Ginsborg has no reason for presenting an account of primitive normativity in this context: on the contrary, she is careful to motivate the idea that normativity could be separable from facts of past meaning, by challenging Kripke's apparent assumption that facts about (past) usage dictate the facts of what I should presently say. However, she does not argue that this assumption is mistaken, only that it is plausibly false — and this hangs, in the dialectic she seems to present, on the plausibility of 'primitive normativity' as an alternative view. An alternative, I should emphasise, to the idea that what I ought to say is a derivative of facts about *meaning* at all. For on Ginsborg's account, what I ought to say is entirely a fact about *me*.

5.3 Ginsborg's Kripke

Ginsborg begins her response to Kripke by considering the passage where he clarifies the sceptical challenge:

"First, [the skeptic] questions whether there is any fact that I meant plus, not quus, that will answer his sceptical challenge [that I really meant quus]. Second, he questions whether I have any reason to be so confident that now I should answer '125' rather than '5'. The two forms of the challenge are related. I am confident that I should answer '125' because I am confident that this answer also accords with what I *meant*." (Kripke 1982, p. 11, my brackets)

Note that at this early point in Kripke's discussion the sceptic still focuses on the possibility of adherence to one's previous intentions, on doing now what is appropriate with respect to what I did before, and that he does not question the possibility of the protagonist's present meaning, usage or intention.

As Ginsborg understands them, the two challenges presented are as follows:

- Challenge 1** Show that there is some fact that I always meant *plus* rather than *quus* by my past usage of 'plus' ('+').
- Challenge 2** Show that my present confidence in answering '125' to the question 'What is $68 + 57$?' is justified.

She takes Kripke to be mainly interested in Challenge 1, and that Challenge 2 exists only to impose a constraint on answers to Challenge 1: a fact of what I meant previously must be such as to justify my confidence in what I now say — or more plainly, to justify my claim that I now say what I ought. She accepts this thought, but objects to the further implicit assumption she attributes to Kripke that Challenge 1 must be answered *first*; that "in order to claim legitimately that you ought to say "125," you need first to establish that you previously meant addition" (Ginsborg 2011b, p. 231). I paraphrase this as follows:

- Past Meaning Justification** Any claim about what you ought to say now must rest on a claim about what you meant previously.

Ginsborg must certainly deny this, as her strategy is to provide an answer to Challenge 2 independently of and prior to Challenge 1; to ground the ought-claim prior to the meaning-claim, and ultimately to show that her solution to Challenge 2 forms the basis of a solution to Challenge 1.⁷

We noted in §1.4 that the ‘should’ in Challenge 2 is a conspicuous piece of normative vocabulary. Ginsborg certainly views it as such, going so far as to reduce Challenge 2 to the question why one *ought* to reply ‘125’. As such, if Challenge 2 involves any question of my justification for saying ‘125’, for Ginsborg, the justification either lies with whatever makes it the case that I ought to say this, or else, the fact that I ought to say ‘125’ just *is* my justification for saying it. Let’s give her take on this a label:

Challenge 2_G Show that I ought to use some expression in a particular way (if I am to accord with my previous usage of it).

I bracket the latter clause in acknowledgement that it may prove cumbersome for Ginsborg’s purposes.

Ginsborg notes that Kripke slides between talk of accordance with your past usage, with your past meaning, and with your past intentions, and claims that he treats these as interchangeable notions — whereas a dispositionalist (or primitive-normativist) would do well to separate a dependency on ‘what I did’ from a dependency on ‘what I meant’:

“Kripke treats interchangeably the idea that “125” is what you ought to say if you are to accord with your past *usage*, and the idea that this is what you ought to say if you are to accord with your past *meaning* or with your past *intentions* . . . But I am denying that the idea of conformity to past usage depends on the idea of conformity to past meaning.” (Ginsborg 2011b, p. 232)

Kripke’s sceptic challenges me to show that I previously *meant* something in particular when I *used* a particular expression, entailing that my *intentions* were such as to mean I *ought* now to use it in some particular way in order to accord with my past intentions. Put in this way, it seems that my past intentions are entailed by some combination of what I meant and how I used the expression. So accordance with those intentions would entail accordance with that combination of meaning and usage.

Ginsborg’s objection is that I can accord with my past usage without having to accord with any past meaning or intentions — she accuses Kripke of assuming “that the “ought” has to be conditional on your past meaning or past intentions, or on a rule which you previously had in mind for the use of the term” (2011b, p. 232). This is crucial to her aim of denying *Past Meaning Justification*, and moving from an answer to Challenge 2_G to an answer to Challenge 1.

⁷I am not entirely convinced that Kripke really does assume *Past Meaning Justification*, hence the priority of Challenge 1, at all. He does indeed say that no fact will count as satisfying Challenge 1 that does not also satisfy Challenge 2. But this is not to forbid attempts to meet Challenge 2, only to exclude attempts to answer Challenge 1 without accounting for Challenge 2 in one’s answer. Kripke does subsequently focus on Challenge 1 as if it took logical priority — but this rings more of an analytical strategy than of a theoretical commitment.

This seems plausible enough, so long as by ‘usage’ we are sure we mean something that is not parasitic on the possibility of meaning; to meet the challenge as Ginsborg intends, the possibility must remain available that my usage of ‘plus’ was just a purely meaningless ‘sounding off.’ Moreover, we must ensure that no facts about this usage bring in anything that would depend on its having been meaningful. So, for example, if the ‘ought’ of Challenge 2_G turned out to be dependent, as per *Past Meaning Justification*, on the existence of some fact about what I meant, then Kripke would be right to treat usage as inseparable from meaning (even if likely wrong to equivocate).

5.4 Meaning in the Slab Game

In her (2012), which builds on the above, Ginsborg argues from a starting point I find myself agreeing with: that sceptics of the normativity of meaning have been too quick to assume that correctness is a stand-in for truth in formulations like “F applies correctly to x”. She agrees with David Papineau that truth is a ‘descriptive property, like car-speed or celibacy’ (Papineau 1999, p. 20): and that descriptions are not normative in and of themselves. But she argues that correctness is normative. Here I will attempt to expound and critique her case.

Does a fact like ‘w ought to be applied to x’ give anyone a conclusive reason to do so? The question here is whether the ought is prescriptive. So far, I have answered in the affirmative; Ginsborg differs here, though subtly. She is willing to accept that the ought is not normative, despite maintaining that normativity is a precondition of meaning. She is also willing to accept that ‘w ought to be applied to x’ can be used interchangeably with ‘w applies correctly to x’. This might seem to be in tension with her thesis that correctness is in some important sense normative for its distinctness from truth (a mere descriptive property), since if correctness is normative, and truly interchangeable with the ‘ought’ phrase here, then the ought phrase should be normative too. But the point here is simply that the properly relevant kinds of oughts and correctness *are* normative. More directly, she will argue that the relevant kind of correctness is not tantamount to truth. Meaning, she claims, arises through the existence of conditions of correct application: but this fact does not amount, *pace* Blackburn, to the fact of ‘there being truth and falsity’ (Blackburn 1984, p. 281) – the correctness in question is a normative matter, where truth and falsity are not.

An important part of Ginsborg’s argument arrives in the form of a discussion of an intuition to the effect that a situation like Wittgenstein’s ‘language game (2)’ could not be described as involving meaningful utterances. To recap on that scenario:

Let us imagine a language for which the description given by Augustine is right: the language is meant to serve for communication between a builder A and an assistant B. A is building with building stones: there are blocks, pillars, slabs and beams. B has to pass him the stones and to do so in the order in which A needs them. For this purpose they make use of a language consisting of the words “block”, “pillar”, “slab”, “beam”. A calls them out; B brings the stone which he has learnt to

bring at such-and-such a call. — Conceive of this as a complete primitive language. (Wittgenstein 1953, §2)

A key point of this, we might say,⁸ is that nothing in this scenario seems to depend on the mental lives of A or B. We can read plenty of that kind of thing into the scenario, but it strikes me — and apparently, Ginsborg — as deliberately minimalistic.

Ginsborg imagines a situation like Wittgenstein's, where A shouts out things like "Slab!" and B gets a building stone of a particular shape and brings it to A, but also where B first has to identify the stone and shout out a corresponding sound ('Pillar!', and so on). Ginsborg entertains the intuition that if and when B unveils a slab and shouts 'Slab!', this is not akin to his *meaning* 'slab' with that sound. She supports this intuition by imagining that A and B are replaced with automata — though she is keen to stress that this is merely an intuition.

The question is then put: what would it take for A or B to mean *slab* by 'slab'? Ginsborg leaps for the answer: the missing aspect here is the speakers' *understanding* of the expression. She wishes to say that in the imagined scenario, neither speaker has a qualifying understanding of the word 'slab'.⁹ What would make the difference would be, at least *prima facie*, their having some *grasp* of the meaning of the word. And so she tells us,

In order, say, for the assistant to mean *slab* by 'slab', there must be a sense in which he understands 'slab' to mean *slab*, which is to say, grasps or recognizes that it means *slab*. (Ginsborg 2012, p. 135)

The challenge she puts to the dispositionalist from this is as follows. If this condition obtains, then the activity of meaning something by an expression is conditional upon having some grasp of the meaning of that expression. But this is difficult for a dispositionalist, who is required to say that 'slab' in fact means *slab* in B's language (perhaps just his idiolect) just in case B is disposed to apply 'slab' to slabs. The difficulty is that in order to respect the former condition, B must be able to recognise that he is disposed to apply 'slab' to slabs simply by virtue of his in fact being so disposed. Yet there is no obvious way to infer, from the fact that 'S is disposed to F', that 'S recognises that S is disposed to F' - much less for S himself.

In anticipation of the dispositionalist's denial of the intuition that frames all this, Ginsborg makes the further point that a meaning-constituting fact is supposed to account for two things: here, that a) 'slab' means *slab* and b) that 'slab' is recognised to mean *slab* (by someone). On this basis, she argues that the very existence of the former problem – of B's recognising that he is disposed to apply 'slab' to slabs – opens up the further problem that if this disposition is itself a meaning-constituting fact then *someone* must recognise that in B's language, 'slab' means *slab*. But if not B, who else? And of course, the argument has the potential to be universalised beyond the (somewhat contentious) idiolectic case, since no matter how many agents we add, their disposition to mean *f* by 'f' is not sufficient for them to know that they have that same disposition.

⁸Though I do take it as practically contentious by definition to provide a blunt interpretation of any part of the *Philosophical Investigations*.

⁹Note here that the focus is on understanding expressions - not the meanings of expressions, which are grasped.

Ginsborg notes (2012, p. 136) that perhaps the best response for the dispositionalist at this point is to say that having the disposition to apply ‘slab’ to slabs *just is* to have an implicit grasp of the meaning of ‘slab’, an approach she identifies with Horwich (1998a, pp. 16–18). However, Ginsborg believes this counter-intuitive response is unwarranted, since there is a better alternative.

Ginsborg’s alternative seems to get at something which underlies all this talk of meaning things by expressions. Before that, she seems to say, comes the notion of something’s being *appropriate* – in a very thin sense, the sense of satisfaction one feels for order of a certain kind, that one has been exposed to in training (that green things can be grouped with other green things better than with blue things, say). This notion does not seem to replace the ‘implicit grasping of meaning’ that was so troublesome for the dispositionalist, so much as to say: well, there is something *appropriate* about saying ‘slab’ in the context where there is a slab in front of me, and that appropriateness is what makes B’s activity different from a machine that beeps in a particular way in proximity to slabs. The machine does not feel this *primitive normativity*, nor does it need to, in order to fulfil its fundamentally non-semantic task. We might object that the machine performs a meaningful activity – but it could be countered that if we look again, the meaning comes from our sense that what the machine does is appropriate.

This account shares many features with dispositionalism, but stops just short of the mark by refusing to identify the relevant kind of correctness — the kind invoked as “primitive normativity” — with truth.¹⁰ Rather, the normativity in question is identified with the possession of a certain attitude towards e.g. pairings of expressions and objects (here, the word ‘slab’ and some slab), such that one feels the expression ‘fits’ the object. Whatever this amounts to in more detail, the dialectically striking thing about it, as Ginsborg hints at herself (2012, p. 139), is its positive presentation. I have often been inclined to define normativity as it pertains to rules and meaning in negative terms: specifically, in terms of the ‘possibility of error’. One ought to F, because it would be *wrong* to not-F. But Ginsborg’s account operates on the back foot, such that one either has this positive normative attitude or one does not; thus the opposite, contra Kripke, of ‘S ought to apply w to x’ is not ‘S ought not to apply. . .’ but rather ‘S “not-ought” to apply. . .’ — which is to say, that there is here the *absence* of normativity, rather than normativity that points us in the opposite direction.

The risk with this, I suggest, is that it lets in too much dispositionalism to claim victory. In what sense, after all, is an attitude really normative? An attitude is something one has or has not — that is where the explanatory appeal of Ginsborg’s position originates — but consistently feeling a certain way about something does not sound like normativity, even if that feeling is wont to guide one’s actions. That it feels right to me to say ‘slab’ here is not normativity in action, no matter how

¹⁰Ginsborg has an argument contra an imagined objector for why this is necessary (2012, pp. 142–143). The gist of it is that if we can stipulate some of the content of B’s disposition such that he implicitly grasps the meaning of ‘slab’, and that (since it is an assertion in this language) that it is correct to assert ‘slab’ iff ‘slab’ is true, then we can paint a picture of meaning that involves correctness etc. as identified with truth. Ginsborg’s reply is effectively that cramming so much implicit information into B’s disposition makes it impossible to use the various parts in the kind of inference we would expect from ‘I use ‘slab’ in these contexts’ to ‘because that states something true’. In other words: inference to the identity of truth and correctness seems important, but cannot be possible with premisses held in merely implicit knowledge.

acute my taste for ‘correct’ circumstances. And that is what it is: a sense of taste. There is no ‘ought’ involved. What we learn from Ginsborg, if she is correct, is that meaning is not normative, and that it does not presuppose anything more normative than what might charitably be called judiciousness, and uncharitably called preference, neither of which is necessarily norm-governed.

Again, if what I have called McDowell’s (1984) ‘Illusion Argument’ ever held water then it does so here also (Cross 2013, pp. 39-40).¹¹ To recap, the point there was that we should like to distinguish between the shared impression of normativity and *actual* normativity — a distinction that Ginsborg’s account does not seem able to accommodate, since the one is effectively identified with the other. Applied to ‘primitive appropriateness’ the question becomes: how do we distinguish between what most people take as ‘appropriate’ and what is really appropriate? It does seem objectionable on some level to say that, at base, what one ought to do is driven by what seems little more than the gut-instinct one would have cultivated under ideal conditions.

We thus get a thin kind of normativity from Ginsborg, but there is a long way to go to get from this idea to something broadly acceptable — if simply because, as in Millikan’s case,¹² the detail we’d want to hear about complex and apparently intellectually-underwritten scenarios is presented as more or less just a matter of ‘scaling up’ from the base case of a primitive feeling of appropriateness (or in Millikan’s case, certain fundamental evolutionary imperatives). And this is before we begin to investigate the inevitable onslaught of responses from the dispositionalist camp to the effect that what Ginsborg has said can be reduced to what they themselves were saying all along: that the most relevant facts of the matter are non-normative, and that all the activity required for meaning to arise can be accounted for in descriptive terms.

5.5 Primitive Normativity and Anti-Realism

Another perhaps worrying issue is that Ginsborg’s position seems to entail a kind of anti-realism about (for example) mathematical truth. For if the answer to Kripke’s sceptic, “125”, is correct on her view, it is at root correct *simpliciter*, rather than by appeal to mathematical rules. The rules seem to serve a clarificatory purpose: rules of addition might help communicate to the sceptic our reasoning to the conclusion we offer. But when pressed, our best *justification* for the answer is simply that it is correct. The *explanation* of such an answer is that it is how we were trained, a primitive association in which we have been inculcated. The result of this is that the correctness of our answer seems to be unavoidably relativised to our training — a different training might yield a different answer, but it too could, on this reasoning, be equally correct.¹³ Modal logic can help us here: for Ginsborg, it seems that when we say, “‘125’ is the correct answer”, this means more completely that “‘125’ is the correct answer at the actual world”. That said, the favourite ‘possible worlds’ metaphor might mislead here, since we are really talking at present about

¹¹See also my chapter 2 above.

¹²Cf. chapter 3 above.

¹³Of course it could also be incorrect. If my training led me to perceive “5” as the fitting answer, and yet in a moment of fuzziness I answered “125”, then that answer would be wrong.

the perspectives of different individuals in the same world - hence something like ‘possible perspectives’ would be the relevant currency.¹⁴

This raises a familiar set of difficult ‘who is right?’ questions. For it seems, from my actual perspective, that:

□(S is asked “What is $68 + 57$?” \rightarrow S ought to answer “125”)

Yet from some other possible perspective *within the same world*, the consequent here would be false. Thus we must either admit, at great theoretical expense, the actual perspective-relativity of mathematical truth, or else establish that some perspective is correct. Ginsborg’s position seems to preclude the latter, though, simply because it aspires to be the be-all and end-all of normative matters like correctness — including the correctness of some perspectives over others. Thus, relative to my perspective it may be that □(S should adopt my perspective). But relative to some different perspective, the same could be true — the matter thus cannot be settled by appeal to primitive normativity relative to anyone’s perspective. But primitive normativity is relativised in this way by definition, so it seems we must conclude that mathematics is perspective-relative, or that primitive normativity is not the ultimate source of all relevant notions of correctness.

In this context, a communitarian view like Wright’s (1980) might spring to mind. While I may not be able to defend the correctness of adopting my perspective on its own terms, I could bring the support of my fellows into the equation. The effect of bringing more like-minded agents into the mix (i.e. those who share my perspective) is to lend non-cognitive support to my claims: there are more of us and we bolster each other’s confidence, so the detractor may be moved to accept that his retraining is necessary. But this seems to mean accepting that it is not primitive normativity that accounts for our commonly held notion of the objectivity of mathematical truth, but rather, our inclination — as described by Wright (1980)— to follow a crowd, and the crowd’s decision is what underwrites my actions.

On the other hand, this is a little uncharitable to Ginsborg. It seems she can quite legitimately respond that this inclination, this susceptibility to peer pressure, is exactly the kind of thing that primitive normativity describes. Thus, the scope of the most relevant kind of perspective covers all functioning human beings, who share these dispositions at a genetic level. The training that any given human may have serves to differentiate his perspective from that of another with different training, but there are features of their responsiveness to the world that they hold in common, and this is what is able to account for the correctness of some perspective over another.

How would this work in the mathematical case? Let us say that A has the standard perspective on what answer we ought to give to the sceptic, viz. “125”, and that B has a deviant perspective according to which the correct answer is “5”. How is A to assert the correctness of his perspective? Well, he brings in his friends, and (in not so many words) makes a demonstrative appeal to B’s feeling that it is right to follow the views of his peers in such circumstances. B has this understanding in common with A; and thus, relative to their shared perspective,

¹⁴For more on ‘perspectives’ see Cross 2013, chapter 4.

he is able to understand something to the effect that:

$$\begin{aligned} \square(S \text{ is asked "What is } 68 + 57?" \wedge \text{ most of } S\text{'s peers ought to answer "125"} \\ \rightarrow S \text{ ought to answer "125"}) \end{aligned}$$

Here, the first ‘ought’ would be relativised to the perspective of S ’s peers (i.e. that which he does not share with them), while the second would be relativised to the perspective S shares with them. The strategy is to appeal to the common ground, and establish the primitively normative requirement to adopt certain kinds of training, such as one that results in the primitive correctness of the answer “125”, from the ground up.

So Ginsborg seems to be able to account for alternative possible perspectives within a given possible world, by building the idea that human beings share the right kind of authority-deferring character traits into the notion of primitive normativity. The next question is whether she can account for the bigger question of the objectivity of mathematics.

Possibly, one of Kripke’s reasons for considering one’s ability to justify answering “125” against the catalogue of one’s previous actions as explanatorily prior to that of justifying your conviction that you ought to say “125” on this occasion per Challenge 2, might be a desire to preserve some of the more intuitively platonistic trappings of mathematics. If the explanatory arrow goes from the former to the latter issue, then we are on the more traditional ground of thinking that your understanding of “plus” relates to your understanding of the rule for the concept’s application — a rule which is something indisputable, universal, and (we tend to hope) absolutely necessary. How is it that you are confident that you are applying this rule? By appeal to the fact that you mean *plus* by ‘plus’. And how is it that such is the case...? Well, that is the question, but at least we are appealing to *facts* now, of which it seems at least some must be objective in the required sense. Kripke ends up concluding the opposite, of course, but at least this way of looking at the relationship of the tasks permits of a realist answer for both.

It is hard to see how Ginsborg’s view on the matter can achieve this. For if there are mathematical facts of an intuitively objective kind then we are still owed an explanation as to how our training comes to align our inclinations about what is primitively ‘correct’ with those facts, if indeed it does do this. The alternative is to accept that the alignment in question is unverifiable, leading (or so it seems) straight into a kind of anti-realism about mathematical facts.¹⁵ We thus seem led back onto ground familiar from engagement with Wright’s anti-realism, and as such, back into the same concerns McDowell raises about the ability to preserve an intuitive conception of objectivity.

5.6 Haddock’s Kripke

Ginsborg (2012) represents one half of a two-paper symposium, of which the latter half is Adrian Haddock’s response (Haddock 2012). Haddock presents his doubts

¹⁵This is not necessarily the case. Unverifiability does not entail non-existence, or even mere quasi-existence as a ‘construct’ or some such. The more relevant point here is, in fact, that we don’t tend to think that mathematical truths are unverifiable. Historically they have been treated as self-evident and *a priori*.

about Ginsborg's ability to a) respond effectively to Kripke's sceptic, and b) to account sufficiently for the notion of meaning's normativity as it appears in Kripke's discussion (1982).

His response plays off an idea that appears most notably in Zalabardo 1997: that the source of the trouble in Kripke's paradox is the weight placed on the idea that justified action means conscious engagement with the facts, such that if my meaning something consists in a certain neurological configuration in my brain, then justification consists in my being suitably 'tuned in' to the neurological facts. For Haddock's Kripke, this means that I should be instructed, unambiguously, in what it is that I must do if I am to mean such-and-such. This is the thought that Haddock ultimately wishes to reject, arguing instead that I can answer 'Kripke's challenge' — which Haddock reads as 'one challenge with two aspects,' rather than two separate challenges — by finding a fact that anaemically justifies my actions (shows that they accord with my past intentions). He claims that this kind of fact can be found without appeal to Ginsborg's account of primitive normativity, so on this basis he rejects her account as both unwieldy and unwarranted, and suggests we should hold back from primitive normativity as a solution to the trouble. That is, he picks up on the major weakness of Ginsborg's position: the question of motivating primitive normativity in the face of viable alternatives, especially given unanswered questions regarding its alignment to Kripke's own starting point.

As Haddock's response turns on motivation with respect to the Kripkean challenge, his analysis of that challenge merits our attention. He begins his account of Kripke on the same footing as Ginsborg, taking note that the two challenges are supposed to be connected, or are even two aspects of the same requirement on an answer, and that any fact that would meet Challenge 1 must also meet Challenge 2. However, he presents Challenge 2 (the 'second aspect') in a slightly different way, focusing on Kripke's tendency to say that a fact as to what I mean must 'contain' the 'directions' for future use, that it must "*tell* me what I ought to do in each new instance" (Kripke 1982, p. 24). As Haddock puts it:

"It is not enough for the putatively meaning-constituting fact to justify what I do in some 'externalist' sense, in which its justificatory capacity is independent of its availability to me. Only something which is available to me can be said to *tell* me what to do."

(Haddock 2012, pp. 149–150)

The importance of this 'internalist' requirement is an important observation about Kripke's sceptical challenge — and indeed, this is the major downfall of 'external' accounts of my justification as discussed in chapter 3. Whatever fact answers Challenge 2, it must be available to me in the right kind of way so as to provide direction (and justification) for my usage of the expression in question. Haddock concludes that this fact must be therefore be "something that I have in mind," such that its obtaining is not "for me a matter of conjecture" (2012, p. 150) — the idea being that one can cite one's infallible access to such mental objects to prevent a further sceptical objection. Yet part of the problem is that precisely this manoeuvre is not permitted.

Haddock follows Stroud (2002, p. 185) in isolating the problematic assumption as:

“... the idea that what constitutes my meaning must be something I have in mind, on a conception according to which it cannot be intrinsic to anything I have in mind that it means one thing rather than another.” (Haddock 2012, p. 152)

This is a neat way of expressing the source of the sceptical regress in Kripke’s paradox. The regress itself is caused by the fact that anything held ‘in mind’ that would satisfy Kripke’s internalistic requirements on answers to Challenges 1 and 2 would *not* be something that stands, as McDowell puts it, without need of ‘interpretation.’¹⁶

Haddock also follows Zalabardo (1997) in emphasising the relation of this to the requirement that it be no matter of conjecture for me that I mean one thing rather than another.¹⁷ The difficulty of the challenge is that certainty about my own mental content is required of me in order to ground certainty about my own mental content — but we are not permitted a starting point of certainty, which would have to be just such a mental item, and thus are unable to prevent a regress. As Kripke paints the picture, nothing I have ‘in mind’ can ‘tell me what to do’ unless I am certain it tells me one thing rather than another — but we assume from the outset that it cannot provide this direction.

5.7 Haddock’s Critique

Haddock begins (2012, p. 148) by noting the concern, voiced similarly by Ginsborg,¹⁸ about the thin notion of meaning-normativity as arising simply because correctness is an issue for any use of an expression. The concern is, simply, that ‘correctness’ for an utterance will often appear simply to amount to ‘truth’; and this is apparently the case for most if not all assertoric utterances. It seems correct to assert that P is true iff P is true. But the issue here is that truth is not an essentially normative notion, no matter how enamoured we are as a society with the value of honesty, factual accuracy and so on. It is a descriptive notion. Something is true, or it is not. What follows from that fact depends only on *further* contextual information, which will introduce any norms relevant to the case. Thus ‘S ought to call x green iff x is green’ is a fact beyond the simple claim that the utterance ‘x is green’ is correct (true) iff x is green.

He observes (2012, p. 149) that Ginsborg’s solution here is to argue that ‘primitive appropriateness’ is the relevant form of correctness, not truth or anything else, and that this is “distinct, non-semantic, and irreducibly normative. . . the mark of ‘primitive normativity’”. He himself prefers to avoid this definition of normativity altogether, and ultimately argues that the normativity actually in question points to the availability of facts to me that will provide justification for my usage of an expression relative to my intentions:

“To say that the relation between what I mean by an expression and how I use it is ‘normative’, and not merely descriptive, is just to say that the facts as to what I mean must be available to me, because they must

¹⁶Cf. McDowell 1984, §§3–4.

¹⁷For more on this, cf. also my §1.6.

¹⁸Cf. my §5.4 above.

be such as to show how I go on to be anaemically justified, and therefore rationally—rather than merely causally—intelligible.”(Haddock 2012, p. 161)

By ‘anaemic justification’ here is meant that the fact’s availability to me reveals that my usage of an expression — *viz.* my response ‘125’ to the question ‘What is $68 + 57$?’ — realises one of my intentions (Haddock 2012, p. 153). This is a departure from the ‘instructional’ view whereby I, the agent, must in some sense be told what to do, or have it made plain to me *that* I am justified in going on as I do. Rather, what matters is that I am in such a state as to *be* justified by my actions — the effect being that what I do is, as Haddock puts it, “rationally intelligible”. Thus we return to the ‘reason-giving’ kind of justification that Ginsborg rejects, but it is in a sense smuggled in, with the key caveat that the underwriting reason need not be plain to the agent.

From Haddock’s perspective, he and Ginsborg diverge most notably at the point where Ginsborg accepts that (contentful) justification for how I ought to go on is unavailable (Haddock 2012, §IV). Instead, he thinks we should reject the operative conception of mind that is in force for Kripke’s challenge, leaving both the possibility of facts of my meaning something in particular, and the possibility of being justified in going on, intact. The aforementioned notion of facts providing ‘anaemic’ justification depends upon this rejection. As such, he and Ginsborg part ways at the point of acceptance or rejection of the conception of mind “according to which it cannot be intrinsic to anything I have in mind that it means one thing rather than another” (Haddock 2012, p. 152).

I will not pursue an analysis of Haddock’s own positive alternative here simply because it does not profess to be complete — he presents a strategy based on the points considered above, and gestures towards the suitability of an account of implicit knowledge as a candidate for providing the mooted anaemic justification. The implicit knowledge literature itself is vast, and has some important champions in the rule-following debate, none of which is explicitly presented by Haddock as his preferred alternative — but, as I will argue in my analysis of Kripke (chapter 7 below), we learn enough from Haddock’s discussion of *how* we should answer Kripke to make Ginsborg’s account — or indeed, any *other* account of the normativity of meaning — seem unwarranted in that context.

5.8 Conclusion

Ginsborg’s major challenge, as suggested in §5.2, is to motivate primitive normativity as a more appealing alternative than competing theories in spite of the above, and the looming question of whether it connects sufficiently well with the idea of meaning-normativity central to Kripke’s challenge. Of course, if Ginsborg can demonstrate that his conception was deficient while hers is not, then that would be a major step in the right direction. But the larger concern, at least from where I sit, has to be about the anti-realist leanings of the account. For as with Wright’s anti-realism, the difference between primitive normativity and Kripke’s supposedly unpalatable ‘sceptical solution’, whereby there are no semantic facts but only ‘justification facts’, seems a little thin on the ground — the contrast needs

to be strongly emphasised and clarified if primitive normativity is to stand as a realistic contender.

Meanwhile, Haddock's suggestion that the key here is not to reject rational justification as Ginsborg does, but instead to reject assumptions about how the agent relates to that justification, strikes me as truly noteworthy. If we can say more about how this would work, then we have what looks like a viable response strategy for Kripke's paradox — and at the same time, one that would come at lower up-front cost than Ginsborg. To this end, I develop the idea in chapter 7 below.

Chapter 6

Normativity, Correctness and Truth

In this chapter I examine some attempts to deny the normativity of meaning that are united by a common theme: the idea that semantic meaning-determining rules themselves are normatively prescriptive, in something like the same way that a 'moral rule' (e.g. 'don't murder the innocent') can be understood as prescriptive.

I begin with an examination of Anandi Hattiangadi's (2007) 'ought implies can' argument against the normativity of meaning, which reasons from the universal scope demands of semantic rules to the conclusion that these rules cannot be normative without imposing infeasibly huge normative demands on agents. She argues that since this is the only way for meaning to be prescriptively normative, the meaning-normativity thesis is intuitively false.

From here I move into a discussion of the relationship between correctness and normativity. Correctness is often perceived to be a normative matter; it is not controversial to say that the levying of certain prescriptions upon us, or otherwise indicating what ought to be the case, is an essential part of what it is for something to be correct (or indeed, incorrect). Moreover, normativists may be inspired to defend the thesis that meaning is normative on the grounds that meaning is in some important sense dependent upon correctness. Such an appeal might play out in terms of *correctness conditions*: so for example, to specify a condition of correctness for the application of some expression is to specify the circumstances under which one ought (or ought not) to use that expression. If we accept that such specifications are part of what makes meaning possible, then meaning is clearly a normative matter.

On this topic, I examine and critique a short discussion featured in Glüer and Wikforss' 2009 paper 'Against Content Normativity'. In that paper, Glüer and Wikforss present an interpretation of the reasoning normativists go through to establish their core thesis, which they call the 'Simple Argument' for the normativity of meaning. Glüer and Wikforss claim that this argument depends on it being a *conceptual truth* that correctness is normative, but that what they call *semantic correctness* is non-normative, and so a counter-example to that premiss. They conclude that the 'Simple Argument' is, therefore, unsound. I challenge the relevance of their counter-example, and conclude that their objection ultimately fails.

6.1 Defining the Normativity of Meaning

Arguing effectively against the thesis that meaning is normative means defining that thesis in sufficient detail, so as to have a clear target for one's objections. This means providing a definition that would be appealing to the proponents of that thesis. However, the involvement of *truth* and *correctness* in the normativity thesis is a matter of contention, and as I will argue here, the formulation does matter.

To say that an expression is used correctly is not to say that it is used to express the truth, nor indeed to say anything more than that its use is correct as measured against some standard. The concept of correctness applies more broadly than that of truth. There are plenty of circumstances where the purpose of some utterance is not to assert the truth, and where perhaps it is even to assert a falsehood, but in such circumstances the proper execution of the action is a matter of whether the speaker acts correctly or not.

It might be suggested also that whether such utterances are *meaningful* or not is also a function of whether they are used correctly. That would amount to the claim that one succeeds in using an expression meaningfully only by virtue of applying it correctly with respect to the purpose of the action. Someone who says something like this presumably thinks that the 'purpose' of my action is defined by what I intended to be doing. So then, whether I succeed in meaning *f* by *x* is a matter of what I intend to do with *x*, and whether what I in fact do aligns with that intention; is correct with respect to my intention. But that would be quite a limited perspective, as actions — even linguistic ones — could plausibly have a purpose external to that defined in the agent's intentions.

When Kripke initially mentions the normativity of meaning, the point he makes is that the meaning of a term, and my intention to honour that meaning, jointly engender some fact about what, in some given scenario, I ought to say. It is not difficult to see that 'what I ought to say' here is a consequence of what it is *correct* for me to say under the circumstances in question. Hence, the normativity of meaning is inherently bound up with the idea of correct use of language — however that correctness is defined.

And yet, there is a tendency in the anti-normativist literature to represent the normativity of meaning as inherently bound up with *truth*. Of course, there is a strong tradition in the philosophy of language of thinking of language as a vessel for conveying truth, but it would be bold to ascribe this view to normativists as it is by no means trivial. With this in mind, in this chapter I'd like to examine Hattiangadi's powerful-sounding 'ought implies can' objection to the normativity of meaning (Hattiangadi 2007, pp. 179-181). I will also return to the theme of truth v. correctness in §6.3 below.

6.2 Hattiangadi's 'Ought Implies Can' Objection

Hattiangadi (2007, p. 180; cf. also p. 57) proposes that the essence of the thesis that meaning is normative can be represented as follows:

$$\text{H-Normativity} \quad S \text{ means } F \text{ by } x \rightarrow (a)(S \text{ ought to (apply } x \text{ to } a)) \leftrightarrow a \text{ is } f$$

Take this to be a formulation aimed at cases of property attribution, so that the x in question will be a term like 'green' or 'square' that means a property like *green* or *square*. It seems that *H-Normativity* is designed to gel with certain truth-conditional intuitions about meaning, and so it aims to provide an obligation to apply a term in all and only truth-yielding ways of doing so. Thus if 'green' means *green*, then one ought to apply 'green' to some object if and only if the object is green.

But Hattiangadi quickly points out that the reverse half of the biconditional is false. In the normal sense of 'ought', it is not the case that if something is green then one ought to apply 'green' to it — and certainly not simply by virtue of the merely *semantic* fact that 'green' means *green*. *H-Normativity* is presented as a genuine prescription to action, but the effect of this here would be to demand that we go around calling all the green things in the universe 'green', on pain of failing to meet our semantic obligations and hence failing to mean *green* by 'green'.

This is clearly too demanding. Instead, she suggests, the most we should commit to is the first half:

H-Normativity* S means F by $x \rightarrow (a)(S$ ought to (apply x to a)) \rightarrow
 a is f

But stripping half of the biconditional away to yield *H-Normativity** leaves this characterisation of the meaning-normativity thesis unable to fulfil the meaning-determining role it was supposed to play. *H-Normativity** tells us that whatever one ought to apply 'green' to is green, but not of any particular green thing that one ought to apply 'green' to it. To determine the meaning of 'green' in anything like the idiom of a truth-conditional semantics, we would need a principle capable of specifying, for any meaningful property-term, all and only those things to which it *ought* to be applied. Applied to our example, it would need to reach out to all and only the things to which 'green' ought to be applied — namely, to *all* the green things — as well as limiting the scope of application to *only* green things. But of course, the problem Hattiangadi has identified is that while the truth-condition for application of a term like 'green' requires this kind of huge scope, an 'ought-condition' that states cases where an agent has obligations *needs* to be more limited,¹ i.e. relativised to the personal operating domain of that agent, or we violate the 'ought implies can' principle.²

So it looks like the needs of truth-conditional semantics and meaning-normativism are fundamentally at odds. There does not seem to be a way to capture the universal scope of a truth-conditional theory of meaning with the necessarily limited scope of ought statements. Hattiangadi's conclusion is thus that we cannot formulate the normativity thesis as a genuine prescription while also expecting it to determine the meaning of words.

¹Indeed, the positive imposition to use an expression w in some set of contexts requires for plausibility that we limit the scope of such contexts. So one example for S 's usage of the word 'green' could be: ' S wants to correctly describe the colour of an emerald under normal conditions'; the possibility of creative use of language means that the satisfaction conditions for any 'ought' of meaning need to be at least as narrow as this, lest they become prescriptively definitive (in the overzealous grammarian's sense of 'prescriptive').

²'Ought implies can' is not, I note, a truth universally accepted. However, it seems sufficiently widespread an intuition as to seriously harm the normativist's case, if it needs to be sacrificed in order to preserve the meaning-normativity thesis.

The big assumption here is that the meaning of a term must be generated by the fact of its normativity, and also in a way that is as reminiscent as possible of a conventional truth-conditional theory of meaning. In other words, a theory that takes meaning to be determined by the following sort of construction:

$$\text{TC-Schema} \quad \forall x('Fx' \leftrightarrow Fx)$$

I suppose that the source of this expectation is the supposition that a truth-conditional theory of meaning would have to be superseded by the normativist's account of meaning. However, this supposition is under-motivated. In fact, the more fundamental assumption here, that the normativity of a term's meaning has to *determine* that meaning, is also crucially under-motivated. In fact, it is hard to justify the assumption that truth-conditional theories of meaning are important here at all, and surprising that Hattiangadi takes this point as the crux of her objection to Kripke's sceptic 'on his own terms' given the patently anti-realist commitments of Kripke's own preferred solution.

We can go some way towards explaining this by investigating potential interactions between a justification-based alternative semantics and the idea that meaning is normative. A typical anti-realist theory of meaning maintains that there are no facts beyond non-semantic facts about human behavioural propensities that could determine meaning. Thus, whether 'green' means *green* on a given occasion is determined by whether we, as a product of our practices and communal inclinations, will consider ourselves justified in using 'green' on such an occasion. Such self-regulation is possible as a product of something normative — or so the thought might go. Hence, meaning here is an essentially normative matter: there simply could not be the phenomenon we describe as 'meaning' without the norms governing the relevant practices.

However, as crude as this picture is, it already tells us more than we need to know to rule out Hattiangadi's motivations. What is described here are meaning-determining norms — plausibly non-evaluative facts about us by virtue of which the phenomenon of meaning occurs — and these are to be distinguished from normative prescriptivity *from* meaning *to* action. Indeed, this distinction is a central idea in Hattiangadi's discussion of the nature of correctness conditions (Hattiangadi 2007, p. 51ff).³ There may be prescriptions at the root of this picture, certainly, at least enough to enable the degree of self-regulation required for the more complex regulatory norms of language practice to become available. But there is not on the face of it any reason to believe that what is required to underwrite Kripke's argument is a full blown prescriptively normative theory of meaning.

Indeed, whether Kripke turns out to be an 'ME' or an 'MD' normativist (or perhaps, neither), I'd suggest the minimal commitment of meaning-normativists is simply that you don't get meaning without some kind of normative prescription — that we wouldn't succeed in meaning *green* by 'green' but for *some kind* of prescriptivity with respect to its use by us. The interpretation of this commitment as a demand for a normative 'theory of meaning' in the traditional ground-up sense is controversial — especially given the roots of the discussion in Kripke's (1982, p. 37) injunction that the relation of meaning *and intention* to future action is

³For more on this distinction see §1.5 above. For more on the relationship between correctness conditions and normativity, see §6.3ff. below.

normative, which seem strongly to suggest that meaning-normativism is a thesis that applies at the level of individual agents (the bearers of ‘intention’) rather than at a universal level. Moreover, given that Hattiangadi’s objection is packaged as the killing stroke in an onslaught against Kripke’s paradox, it would seem to miss its target — even if it were to threaten other kinds of normativism — since the point is that there is in some sense a correct answer to the question ‘What is $68 + 57$?’ that is determined by the semantic rules being *followed* by the speaker and his interlocutor.

6.3 Normativity from Correctness

Entertaining that point, it would seem that the involvement of correctness in the discussion of meaning-normativity is necessary, and Hattiangadi does in fact clearly recognise this. However, it does not play a big enough part in her characterisation of the normativity of meaning as described by Kripke.

What she points out is that a traditional picture of truth-conditional semantics of the kind epitomised by *TC-Schema* above can be captured in terms of ‘correctness conditions’ just as well as it can in terms of ‘truth-conditions’. The resulting position, which Hattiangadi calls ‘semantic realism’, seems just to be the familiar view that

$$‘\phi’ \leftrightarrow \phi$$

with the addition of some Kripkean views on what it is to understand any given expression. So she says,

According to the semantic realists, the meaning of a word can be given by its correctness conditions. We can specify what ‘blue’ means by saying that it applies correctly to all and only blue things. But this tells us nothing about understanding — for that, we need the assumption that understanding a word is akin to following a rule. This rule can either be a prescriptive rule, or not. (Hattiangadi 2007, p. 179)

So assuming that there are rules ‘somewhere’ that determine meaning, the question is whether they themselves prescribe action or not. If they do, we are told, then we should expect something like *H-Normativity* to follow — but then we run into the ‘ought implies can’ problem, where a semantic rule which by its very nature has universal scope is required to have very local prescriptive scope or else be over-demanding. But then, if these rules do *not* prescribe action themselves, so says Hattiangadi, then “meaning is merely norm-relative, but not also normative” (2007, pp. 179-180). ‘Norm-relativity’ she defines as follows:

$$\text{H-Norm-Relativity} \quad S \text{ means } F \text{ by } t \rightarrow (a)(S \text{ applies } x \text{ ‘correctly’ to } a) \leftrightarrow a \text{ is } f$$

The idea, then, is that someone who understands the meaning of ‘green’ will only apply ‘green’ correctly to any given thing where that thing is green, but that they will not, without the involvement of some further fact, be under any obligation *thereby* to make correct use of ‘green’ on any particular occasion. Tying back

to Kripke, the result is that if the rules of meaning are non-prescriptive in themselves, then while it follows that there is a correct answer to ‘What is $68 + 57$?’ by virtue of the meaning of the plus sign in that expression (and in any answer given), it does not thereby follow that there is an answer that *should* be given. Hattiangadi’s thought is that something more than *H-Norm-Relativity* is required to sustain the conclusion that, on an intuitive ‘semantic realist’ view, the meaning of the terms involved compels us to say one thing over another. The key issue here is, of course, whether we think ‘correctness’ is itself a normative concept, hence a suitable source of such obligation. Hattiangadi is confident that it is *not* inherently normative, and indeed this fits with her apparent identification of truth-conditional semantics with ‘correctness-conditional’ semantics; truth is, it seems, a kind of correctness, but not one that entails normativity unless we assume, in a way that I would suggest is out of fashion, that all language aims at the truth.

Indeed, it is on this basis that Glüer and Wikforss (2009a) reject the essential (ME) normativity of meaning. As Glüer and Wikforss pick up in this direction where Hattiangadi leaves off, I will now pursue the issue in reference to their work on the topic.

6.4 The Simple Argument

In ‘Against Content Normativity’ (2009),⁴ Glüer and Wikforss attribute a line of reasoning they call the ‘Simple Argument’ to normativists⁵ who think the essential normativity of meaning can be derived directly from meaning’s connection to correctness. The argument itself is a formulation of Glüer and Wikforss’ own, and we have it mainly on their word that a) normativists would endorse the Simple Argument, and that b) normativists intend the argument’s premisses to be as humble as possible, in order to establish the conclusion as a basic, common intuition. I will assume for present purposes that both are true. The Simple Argument is as follows:

SA(1) There is an essential connection between meaning and correctness conditions;

SA(2) Correctness is an (essentially)⁶ normative matter;

SA(3) Therefore, meaning is essentially normative.

The idea here is that, since SA(2) is a conceptual truth, SA(1) is the argument’s only substantive premiss. Moreover, SA(1) is not controversial. So on the face of it, the Simple Argument is impressive: the only contestable premiss happens to be a commonly held intuition, and the conclusion is a universal claim about meaning. So if the argument is valid then it seems likely sound as well.

⁴Cf. also Wikforss (2001), Glüer (2001), Glüer and Pagin (1999) for the development of this discussion.

⁵They find it most notably in Boghossian (1989) and Blackburn (1984).

⁶The Simple Argument appears in the paper’s body text with the word ‘essentially’ omitted from what I have labelled SA(2), though it is clearly implied by their subsequent discussion. That said, perhaps “Correctness is a normative matter as a matter of conceptual truth” would be a fair substitute for SA(2). Cf. e.g. p. 35: “the simple argument depends on its being a conceptual truth that correctness... is normative”.

On that point, we might have some concerns about whether the argument's conclusion really follows. The reasoning seems to depend on the idea that if A has an essential connection to B, and B has some essential property, then A has that property too. But that is clearly false: human beings are 'essentially connected' to oxygen, but they are not a gas. So if the argument is valid at all, then it must be fair to assume that whatever the nature of the essential connection posited in SA(1), that connection must be such that meaning would be rightly called 'normative' if the correctness conditions in question were essentially normative. I will return to this point: it may be significant that such a substantive detail is not explicit in the argument as Glüer and Wikforss present it. For now, the conclusion to draw is that the Simple Argument's effectiveness does depend importantly upon our interpretation of SA(1).

This all turns on what is meant by a 'correctness condition'. The idea is frequently alluded to in discussions of the normativity of meaning, and also in the literature surrounding rule-following, though often without much elucidation. Glüer and Wikforss (2009a, p. 35) suggest the relationship between meaning and such correctness conditions, where w is any expression, should be:

$$(C) \quad w \text{ means } F \rightarrow \forall x(w \text{ applies correctly to } x \leftrightarrow x \text{ is } F)$$

Hopefully it is plain that Hattiangadi's *H-Norm-Relativity* is simply an agent-involving invocation of (C) — and as it happens, (C) is for all practical purposes identical to the formal definition of 'semantic realism' that Hattiangadi provides (Hattiangadi 2007, p. 52).⁷ The correctness-condition proper appears as the consequent of (C)'s major connective, while the antecedent links it to meaning. (C) is not perfect though: ideally we'd like a condition that would establish whether full declarative sentences apply correctly to contexts of utterance, but (C) makes a mess of this — the word "is" makes things awkward. Nonetheless, it seems fair to say that (C), or at least something very like (C), must be true.

Glüer and Wikforss claim that the purpose of the Simple Argument is to derive the normativity of meaning from the truth of (C). If so, that would mean that (C) expresses precisely the essential connection between meaning and correctness conditions described in SA(1). That is, I suggest, a bold assumption: Glüer and Wikforss are already attributing the Simple Argument to normativists without their express approval, and adding more precision can only weaken the plausibility of the attribution. But as previously indicated (§6.4), I'm going to overlook problems of this kind in favour of directly engaging with the objection.

So if we play along, SA(1) appeals to the fact that (C) tells us whether any given application of w to x is correct, and that this is a condition upon w 's meaning anything. Since by SA(2) correctness is normative, it follows that a condition upon w 's meaning anything is that normative consequences follow. Hence, meaning is normative. But Glüer and Wikforss deny that the second step here is legitimate, since a plausible reading of (C) yields only the *semantic* fact that some applications of w are categorised as 'correct' - or whatever other basic semantic concept we substitute for "applies correctly to" in (C). They take truth as an example:

⁷Versions of (C) have been invoked in several papers on this topic by Glüer and Wikforss (Glüer and Pagin 1999, Wikforss 2001), as well as elsewhere by Hattiangadi (2006) and, in defence of the normativist, by Whiting (2007).

Granted, if I mean *green* by ‘green’, then ‘green’ is true only of *green* things, and if I say ‘That is green’ while pointing at a red object, I have said something false. But it does not immediately follow that I have failed to do what I ought to do—not even from a merely semantic point of view. There are non-normative uses of ‘correct’, and this is one of them. The relevant notion of correctness in this context is that of *semantic correctness*.(2009a, p. 36)

So “semantic correctness” is offered as a counter-example to the normativist’s pivotally important claim that all kinds of correctness entail prescriptions (‘oughts’) as a matter of conceptual truth. For the concept Glüer and Wikforss describe has exactly the conditions of application described by (C): ‘green’ applies correctly (or we might say, truthfully) to some object iff that object is ‘green’, and yet while any application of ‘green’ to a non-green object is, by that standard, incorrect (false), this does not tell us anything about how we *ought* to apply the word ‘green’.

We might be tempted to interject here that *of course* the incorrectness of an assertion proscribes its utterance—for an assertion to be incorrect is for it to be an assertion we ought not to make. But Glüer and Wikforss have a reply for this: while it may be the case that, for us, falsehood in assertion is to be avoided (or is otherwise relevantly proscribed), for this to be the case there must be some norm in force to that effect—say, the norm that we ought only to assert the truth. But (C) *itself* does not put any norms in force. In other words, (C) is perfectly intelligible as a condition of correctness where “correct” is not considered an evaluative or otherwise norm-laden term. It is, therefore, *not* a conceptual necessity that falsehood entails proscription. And of course, the point Glüer and Wikforss are pressing is that the exact same reasoning can be applied to any categorisations that(C)produces:

... what (C) gives us is nothing more than the conditions for the application of the basic semantic concept to applications of *w*. Nothing in (C) shows that this has to amount to anything over and above the possibility of categorizing, or sorting, applications of *w* into two basic semantic kinds; for instance, the true and the false.(2009a, p. 36)

So (C) operates as a kind of semantic sorting function that merely classifies all applications of *w* as (semantically) correct or not without attaching any prescriptions to the classifications it makes. The comparison is made with a function that sorts objects into tables and non-tables: nothing directly follows about what one ‘ought’ to do with something as a result of its merely being *classified* as a table or as a non-table. Everything that we might *think* follows from something’s being a table (having legs connected to a horizontal surface, having mass, being subject to physical laws etc.) is a consequence of other norms and conditions which, for us, are already in force—not, Glüer and Wikforss maintain, a consequence of the mere categorisation of something as a table. And this seems a fair claim — hence the convincingness of the argument that semantic correctness, a precise analogue of the table-sorting function, has no direct normative consequences. If something is correct, anything normative that follows is a result of further facts about the norms that govern our language practices.

6.5 Analysis

There are a few important things to note about the position this objection leaves us in if we are to defend the Simple Argument. As already mentioned (§6.4), the argument's proponent aims to establish that meaning is normative from the most intuitive premisses possible. Thus SA(2) is taken to be a given conceptual truth, and SA(1) is intended to be the only substantive premiss, albeit one which is uncontroversial. The point of all this is to establish SA(3) as a basic intuition that ought to be widely held; adding more precise substantive details would detract from the purpose of the argument.⁸ This means that any defence of the argument against Glüer and Wikforss' objection should add as little as possible in the way of substance to the premisses already admitted. That leaves us in a difficult position. How should the normativist respond?

Ideally we'd hope to find some clear logical error in Glüer and Wikforss' reasoning, demonstration of which would require no substantive additions to the materials we already have. But in reality, there are two main contenders for response strategies. The first is to deny that 'semantic correctness' qualifies as an example of the kind of "correctness" SA(2) mentions, making it inadmissible as a counter-example. However, this would almost certainly involve making some substantive qualifications to the Simple Argument, undermining the normativist's motivation for making it, and putting us at risk of dogmatism and, worse, circularity. The second option is to deny that 'semantic correctness' is non-normative. But Glüer and Wikforss voice their scepticism about this strategy in a footnote:

This might look like nothing but a basic clash of intuitions. According to us, semantic correctness is not a normative concept, according to the normativist, it is. Thus, the normativist might feel safe in just digging his heels in in one of two ways: he might be tempted to conclude either that we do not talk about the same topic, or that his opponent has a poor grasp of the relevant concept and that the opponent's intuitions therefore can safely be ignored. Less dogmatically, he could move on and provide what we claim is required: further argument. And that would be wise, for there is no safety in dogmatism here. Since (C) is not disputed, any non-normative concept of semantic correctness will be co-extensional with its normative sibling. The latter provides nothing more than a normative gloss on the basic semantic categorization both equally effect. The need for an argument to the effect that this gloss is needed at all is, thus, very real. Unless such further argument is provided, the only correct conclusion to draw from these considerations is that normativity does not seem to be anything but an idle wheel in the theory of meaning. (Glüer and Wikforss 2009a, p. 36, n. 8)

They thus admit that we may not share their intuitions, but argue that the onus is on us to demonstrate that 'semantic correctness' is normative. But again, it seems we would be forced to offer substantive argument for this positive claim,

⁸This is why the nature of the "essential connection" referenced in SA(1) is left open; whatever you think that connection is, the only thing you must agree with for the argument to work is that it would make meaning an essentially normative matter if correctness were an essentially normative matter.

so we undermine our position on this approach as well. Realistically, it seems that any response on behalf of the normativist will detract at least slightly from the Simple Argument's breadth of appeal. Both these options have this consequence, though the extent of the sacrifice necessary in each case is a further question. The best approach will clearly be the one which defends the Simple Argument with minimal collateral damage.

Before we proceed, I should make an observation about SA(2). Since the Simple Argument is something being attributed to Glüer and Wikforss' opponents, we would be forgiven for thinking that the universal claim here that *all* kinds of correctness are normative is just unnecessary — that the Simple Argument is a straw man. After all, meaning does not have an essential connection to *all* kinds of correctness: there are plenty of kinds of correctness that seem to have nothing at all to do with meaning. But in fact, the normativist's (alleged) motivation for making the argument at all necessitates this universal claim. If SA(2) specified precisely the kinds of correctness the normativist is interested in then it would become another substantive premiss.

That said, it is one thing to specify which kinds of correctness SA(1) refers to, and another to eliminate those which are clearly irrelevant — provided that the process of elimination is suitably incontestable. This opens up the possibility of a subtle variant on the second response strategy: if we can demonstrate that 'semantic correctness' cannot be a kind of correctness condition to which meaning is essentially connected, then we can preserve the reasoning and spirit of the Simple Argument with the simple caveat that 'semantic correctness' is an (inconsequential) exception to SA(2), i.e.:

SA(1) There is an essential connection between meaning and correctness conditions;

SA(2)* Correctness is an essentially normative matter, except for 'semantic correctness';

SA(3) Therefore, meaning is essentially normative.

This would work because the point of SA(2)'s universality was to ensure that whatever kind of correctness we think meaning *is* essentially connected to will be normative — but if we can all agree that we *cannot* be talking about 'semantic correctness' in SA(1) then the Simple Argument will still function pretty much as intended. As such, this is the kind of strategy I will adopt.

6.6 In Defence of the Simple Argument

Based on these considerations, the success condition for Glüer and Wikforss' objection is simply that the following two statements are true:

- a) Meaning is essentially connected to 'semantic correctness';
- b) 'Semantic correctness' is not essentially normative.

My precise response strategy here is thus to deny that condition (a) obtains if condition (b) obtains. That is, I will argue that if (b) is true, then (a) is false. As a secondary concern, in §6.6 I will also question whether (b) actually obtains at all.

To begin: the plausibility of SA(1) rests on its implication that meaning would not be possible were it not for the correctness conditions in question. So whatever form the relationship between meaning and correctness conditions takes, it is such that the relevant correctness conditions perform some kind of enabling role for meaning. As such, any correctness conditions that do not play some kind of enabling role are not such that meaning is essentially connected to them.

Recall that Glüer and Wikforss assert the following:

... what (C) gives us is nothing more than the conditions for the application of the basic semantic concept to applications of w . Nothing in (C) shows that this has to amount to anything over and above the possibility of categorizing, or sorting, applications of w into two basic semantic kinds; for instance, the true and the false. (2009a, p. 36)

This amounts to saying that it is in principle possible for (C) to categorise applications of w without those categorisations having any normative consequences. They are not denying that, as a matter of contingent fact, norms are in force for us such that (C)-categorisations have normative consequences: but they *are* asserting that the semantic correctness of an application is possibly separable from such norms, and (b) entails this.

I would like to put that claim to the test in the context of the objection—remember, my hypothesis was that $(b) \rightarrow \neg(a)$. So let us imagine some world W_1 such that (b) is true at W_1 , and test whether (a) can be true at W_1 . We should also stipulate that meaning is possible at W_1 , otherwise (C) will be at most trivially true, and would cease to be a fair characterisation of (a). More importantly, (a) is a *conceptual* claim: it says that *wherever meaning is possible*, it is essentially connected to semantic correctness.

So at W_1 we have meaning, we have the truth of (C), and for any application of w , the semantic correctness or non-correctness of that application has no normative consequences. Can there be an essential connection here between meaning and semantic correctness? Well, let's examine what it means for something to have no normative consequences. Note first that what we mean here is not that there are no norms in force *at all* at W_1 , which would make it hard to imagine meaning being possible, but just that nothing normative follows from (C)-categorisations. But what does that mean in practice? Let's assume, since meaning *is* possible here, that there must be some language-users around at W_1 . If one of them, S, makes an application of w , and it is semantically correct (or not), then nothing normative follows. How is S to respond to this categorisation?

The short answer is: arbitrarily. By stipulation, S is in no way guided, obliged or directed to do anything in particular with semantic correctness values. Since there are no norms governing the practical deployment of (C)-categorisations, it is simply a matter of brute decision or disposition whether S reacts in any given way to them. He may, for instance, treat semantically correct applications of w in precisely the same way he treats semantically non-correct applications of w , treating them as synonymous. That is of course to presume that he even chooses to acknowledge them at all, which is but a matter of choice or disposition here.

But then it is hard to imagine what kind of contribution such categorisations, which are interpreted in no consistent way by language-users (if at all), could make to language-practices at W_1 . Indeed, it has been stipulated that no norms *at all* are to govern what language-users must do with (C)-categorisations, which actively precludes them from being given a role in any such norm-governed practices. But surely meaning itself is just such a language practice — after all, we are not attempting to deny that language practices are normatively regulated, only that (C)-categorisations provide the regulation. If so, then semantic correctness clearly makes no contribution to meaning here, since we stipulated that meaning is possible at W_1 but semantic correctness cannot play any role in meaning. If it cannot play any role in meaning, then *a fortiori* it cannot play an enabling role in meaning. Thus, meaning is not essentially connected to semantic correctness; (a) is false at W_1 . But (a) is a *conceptual* claim. So the falsehood of (a) at W_1 entails the unqualified falsehood of (a).

It might be objected in a sort of ‘naturalistic’ spirit that while S’s treatment of (C)-categorisations may be determined only by his brute dispositions and inclinations, it may well be that his peers are subject to the exact same dispositions and inclinations. That is, while nothing *obliges* S to treat (C)-categorisations in a particular way, it may be that he *does* treat them consistently in one particular way — perhaps he always ‘reacts positively’ to the semantically correct and ‘reacts negatively’ to the semantically non-correct. Moreover, perhaps all other language-users at W_1 react in the same way. So what is to stop this being the basis for meaning? After all, what we call ‘normativity’ may be a product of just such brute dispositions and inclinations.

But this is to miss the point: we have established that at W_1 there are *no norms* in force that govern what ought to be done with semantically correct and semantically non-correct applications of w (even if there are other unrelated norms in force). So while the scenario imagined in this objection is certainly possible, the adoption of (C)-categorisations into a language-practice involves subjecting them to norms, which has been disallowed.

We established that for any world where (b) is true, it must be possible to separate semantic correctness from the norms that tell us what to do with it, but denying semantic correctness any normative consequences falsifies (a). So either (a) or (b) must be false. But both (a) and (b) are jointly necessary for the success of Glüer and Wikforss’ objection to the Simple Argument, so it seems the objection fails.

We return now to the issue of collateral damage. If the response offered is effective, how badly does it detract from the Simple Argument’s impact? It is hard to deny that substantial points are made, and even harder to imagine how we might turn them into new premisses for the Simple Argument.

However, there is one consideration that may tip the balance in the normativist’s favour here. It could be argued that Glüer and Wikforss’ objection depends on the hidden substantive premiss that semantic categorisations are both possible and essential to meaning in a world where they have no normative consequences, and that the substantive parts of the response are intended to balance that assumption. Thus the Simple Argument is merely shown to be incompatible with what turns out to be a false additional claim, and one which is not made by the normativist anyway.

Whether that kind of argument will balance out the dialectic is a complicated matter, and one which I cannot pursue here. But I can offer one further thought that may slightly improve the normativist's position. In §6.6 our thought-experiment assumed that meaning would be possible in a world where no norms attached to semantic (non-)correctness, for the sake of testing (b)'s compatibility with (a). But I harbour doubts about the coherence of that idea, since (a) seems intuitively quite plausible, but we learned that this scenario entails the falsehood of (a). In fact, the conceptual truth of (a) would help limit our collateral damage because it falsifies (b). For if (b) is true then we will still need to adapt the Simple Argument by replacing SA(2) with SA(2)* as suggested in §6.5. But if (a) is true then (b) is false, meaning semantic correctness is no counter-example to SA(2) and the Simple Argument succeeds in its original form. This would at least marginally reduce the dialectical cost of mounting our defence.

6.7 Conclusion

Lying and other deliberate 'misuse' of meaningful expressions might be thought to cause a problem for the normativist. If so, this is because the position involves some close association between truth and correctness, and a corresponding close association between correctness and what one ought to do. This position might entail that if 'green' means *green* then 'green' applies correctly to anything iff it is green, and additionally, that one ought to apply 'green' to something iff it is correct to apply 'green' to it. A consequence of this position is that one ought not to apply 'green' to non-green things — hence lying about whether is something is green, or using the word 'green' in some creative, non-literal way, are both things that one ought not to do. But this is dubious: that 'green' means *green* does not obviously entail anything about only using it to speak the truth, or in a literal sense — or any of the other 'ought not's' that this condition entails.

However, the source of the trouble is not the idea that meaning is normative, but the questionable version of that idea under consideration. As characterisations like (C) demonstrate, anti-normativists frequently represent normativism as involving a tie between truth and correctness, and indeed, this in itself is not the problem. But to represent the *whole idea* of normativism in these terms seems disingenuous, since the thought itself has far closer ties to a school of thought that rejects such a dominant role for truth in language practices. Similarly, as I have previously suggested, Kripke's concern in the rule-following case is not obviously about the normativity of the rules themselves so much as the normativity of the 'following' relationship of individuals *to* those rules.

To illustrate a little, let's assume a kind of normativism that holds oughts to follow from correctness. This position subscribes to the 'correctness iff ought' principle:

CIO $(x)(a)S$ ought to apply x to $a \leftrightarrow$ it is correct to apply x to a

Hence if an application is correct, S ought to make it; and if S ought to make an application, then it is correct. Suspending concerns about the demandingness of the first conditional, it is plain that CIO requires some notion of correctness. In the straw man kind of normativism we discussed before, the only condition under

which ‘green’ applies to something is if that thing is green. The correctness of any given application of ‘green’ is thus determined entirely by the truth of the property ascription. For simplicity, let’s stick with canonically property-ascribing words like ‘green’ for our discussion.

For a large class of cases, to say of something that it is green is to assert that it is, as a matter of fact, green. In such circumstances it is surely plausible that the correct use of the word depends on whether the thing described really is green, although it is not entirely wrong to say that the word is correctly used so long as the speaker believes that what they are describing is green, and that they understand what ‘green’ means.

Fred 1 So, forgetting that he’s wearing green-tinted sunglasses, Fred points to a red tomato and blurts out: ‘It’s green, that tomato!’. What Fred said is false. Insofar as his intention was to say something true, what Fred said is also *incorrect*. But: does it follow purely from *linguistic rules* that what Fred said is incorrect?

The sticking point here is that Fred successfully means *green* by ‘green’ in this instance, and he applies it in a situation where arguably he ought to have done: one where he intends to say that some tomato is green. It is quite plausible that what determines correct use in some cases is not truth but intention. With respect to Fred’s intention to mean *green* by ‘green’, what he has said is correct. And as the objection goes, surely questions about meaning have nothing to do with whether one ought to speak the truth or not.

Fred 2 Now imagine Fred takes his tinted glasses off and, in a vain attempt at trickery, or perhaps out of a combination of stubbornness and embarrassment, he points to a red tomato and says: ‘Yes, I’m sure of it. That tomato is definitely green.’

Fred is lying. Regardless of whether Fred ought to lie on this occasion, whether he has done so correctly depends on the success conditions of lying. Assume that successfully lying is just saying something you know to be false in an assertoric manner. Then it seems Fred has succeeded, despite applying ‘green’ to something non-green, and which he knows is not green, without even having the slightest intention of applying it to something which is green. Hence this is a correct application of ‘green’ in the context of a lie.

But did Fred mean *green* by ‘green’ on this occasion? Perhaps his actions are explained by his meaning something other than *green* here: perhaps he meant *red* by ‘green’, hence his applying it to a red tomato.

No. This kind of analysis doesn’t work — principally because it assumes that Fred’s *intention* is to say something true. The thought is that if his describing something non-green as green counts as correct, then he must mean something non-standard by ‘green’, because, after all, ‘green’ only applies correctly to green things.

But he doesn’t: he is lying. He claims, with deliberate misanthropy, that the property *green* attaches to something when he knows it doesn’t. He does this by indicating the property *green* with the word ‘green’, and through a standard assertoric mechanism he asserts that the property he means by ‘green’ attaches to a red tomato.

Thus, Fred does mean *green* by ‘green’. He also correctly applies it to a non-green object: in the context of the lie Fred intends, it would be *incorrect* to apply

‘green’ to a green tomato. Neither meaning nor correctness is thus tied to truth. Moreover, CIO is sufficient to entail that Fred *ought* to apply ‘green’ to the red tomato in question.

This leaves a final explanatory gap. The points just established jointly entail that *H-Normativity* (chapter 6.3 above) is false, since they imply that there are conditions where the antecedent of the main conditional is true and the consequent is false. This also means that the weaker (non-biconditional) variant *H-Normativity** is false, for exactly the same reason. As such, we are owed a better explanation of the connection between meaning and the ‘oughts’ described — and I think that the involvement of agential intentions in that explanation is paramount.

Chapter 7

Defusing Kripke's Paradox

If we could sustain the notion of meaning-normativism in some intelligible form, would there be any hope of escape from Kripke's paradox? My final line of enquiry focuses on the terms of the challenge. Kripke envisages his problem as metaphysical: how could it be that you mean one thing rather than another, given that your actions are compatible with your having meant just anything, or nothing at all? However, I argue that Kripke is not entitled to conclude from the resources available to him that there is no way you could mean anything, and that the bite of this question is not genuinely metaphysical.

Rather, his challenge is an epistemological one: how could you know that you mean one thing rather than another? I trace this point from the origins in rule-following, through the concepts of intention and understanding, returning finally to the concern that whatever intentional facts about you might obtain irrespective of your knowledge thereof, you still do not know what you ought to do in light of your (past) intentions. Yet this is no trouble in itself – the domain of moral philosophy exists precisely because it is permissible not to know what ought to be done, so to discuss and commune on the matter, and to reach an agreement that will do in the place of what might be inaccessible knowledge, but about which we have some plausibly well-founded beliefs.

I argue that Kripke's sceptic must ultimately defer to a wholly familiar (but not uncontroversially fatal) kind of scepticism about internalistic justification to show that our intentions are indeterminate, or else fail to draw the desired *metaphysical* conclusions. Whether the sceptic succeeds or not, to premiss the problem on the idea that meaning is determined by intention has the effect of making Kripke's subsequent call for the rejection of truth-conditional semantics look question-begging, as opposed to merely unmotivated. In §7.2 I argue that the involvement of the notions of correctness and interpretation does not present any further problems.

I conclude that Kripke's sceptic does not present a problem that is essential to the normativity of meaning or, indeed, to rule-following, as all the force behind the challenge derives from scepticism about the possibility of knowledge. Any adequate response to this standard scepticism is therefore sufficient to resolve Kripke's paradox.

7.1 Intention without Knowledge Thereof

In §3.5 we evaded the main sceptical conclusion by denying that meaning depends upon intending to mean one thing rather than another. We granted that the sceptic may have uncovered a problem about the latter, but consequently denied that this tells us anything about the former. We reasoned this way by adopting the premiss that the meaning of any application of a word is determined by semantic rules, but so long as it is *not* dependent on intention, the conclusion that language has no determinate meaning is unwarranted.

As discussed, the success of this counter-argument depends on how willing we are to jettison the idea that meaning is determined by intention. I have suggested that the association of the two might be traceable to a confusion between the two senses of 'meaning' (i.e. semantic and intentional) we distinguished, and that in fact it is highly plausible that meaning and intention to mean do not determine one another. But I permit that the reader may not be convinced. In this section, therefore, I will take it as a premiss that meaning *is* determined by intention. I will argue that, even so, the sceptical line of reasoning does not lead to the conclusion that language is meaningless, because it does not establish that there are no facts about what I intend.

First, though, I note that perhaps the reader will be concerned that I reject some well-motivated internalist premiss of Kripke's debate. As has been widely observed, Kripke holds that not just any old fact can serve to establish what I mean by 'plus', and there is some kind of internalistic condition in place. A suitable fact must be 'accessible', perhaps such that I could 'consciously engage' with it as Zalabardo (1997, p. 480) suggests. Contrary to our thoughts about acting merely 'in accordance with' semantic rules, Kripke seems to think it is a condition on my words meaning anything that I *follow* such rules — which requires that I *understand* the rule, that I *know* that I follow one rule rather than another, that the rule *justifies* my usage, and that I *regulate* my behaviour as the rule prescribes. But, so the argument goes, I cannot do any of these things unless I can access the fact as to which rule I *intend* to adhere to, which will tell me what I *ought* to do. Since the sceptic proves that my intentions are indeterminate, the conclusion that I actually cannot do any of these things follows, and on Kripke's premisses, so does the conclusion that the words I use never mean (semantically) anything.

Even if we accept Kripke's premisses here, this analysis makes it clear that the argument still depends on the indeterminacy of intention. Which rule did I *intend* to follow in the past: the rule for *plus* or the rule for *quus*? Looking at the facts about me, it seems I could have been intending to follow either. In fact, to follow either would be to do the exact same thing. So if I want to 'carry on in the same way', it is entirely indeterminate what that 'same way' is.

Let's clarify what it is to intend to follow a rule. If there is some semantic rule that tells me for any case how to use the word 'plus' in a way that will mean *plus*, then to intend to follow that rule is not only to act as it demands. My actions could 'conform to' a rule without my intending them to: perhaps someone playing football doesn't know the offside rule, but as he is never offside, his actions accord with it nonetheless. Perhaps in continuing the sequence '5, 10, 15, 20, ...' I am stumped and randomly select the number '25': if the rule of the sequence is ' $n + 5$ ' then this action conforms to it, but it would be wrong to infer that I therefore

followed this rule.¹ Rather, as is often observed, to *follow* a rule is surely also to *intend* that one's actions conform to the rule.²

But in the case where I guess the answer, surely then I *intended* to conform to whatever rule it was that determined the sequence. I did not know what that rule was, but I intended that whatever the rule was, my actions should conform to it. Does this create a problem for our definition?

No: this is the intensional fallacy. Assume that I intend to buy whatever Fred wants, and that Fred wants a car; it does not follow that I intend to buy a car, since I may not know what Fred wants. It is not possible to substitute intentional content like this. Hence I may intend to conform to (the rule that generated the sequence '5, 10, 15, 20, ...') and yet not intend to conform to (the rule ' $n + 5$ ').

Still, this does raise an important question: is it the case that I only follow a rule (conform to it intentionally) if I *know* which rule it is that I intend to conform to? That is: do I intend to follow the *plus* rule only if I know that I intend to follow the *plus* rule? The supposition is the following:

KI S intends to $\phi \rightarrow$ S knows that (S intends to ϕ)

To assume KI is possibly central to Kripke's challenge. Note also that KI also entails a conditional in the reverse direction on the intuitive assumption that S only knows that P if P is true. Hence the assumption in its full glory is:

KI* S intends to $\phi \leftrightarrow$ S knows that (S intends to ϕ)

However, this point is not a given. There are plausibly cases where intention might be ascribed without ascribing simultaneous knowledge thereof: we might cite animal examples. Perhaps more appropriate would be cases where a human acts immediately on instinct, such as to catch a projectile heading for his or her face, where the intention (to avoid injury) is apparent but where claims that the subject knows this at that very instant are contestable.³ Moreover, we might question KI* on the contention that I cannot know that P unless P is the case at least some period of time before I putatively know that P, since this claim would necessitate periods of time where S intends to ϕ but where S does not know this.

Such a position is not impossible, but it is controversial. So let us grant the assumption that KI* obtains: that intention to ϕ is biconditional on knowledge thereof. Is this sufficient for the conclusion that S cannot intend one thing rather than another? Clearly the sceptic must demonstrate a problem with S's knowledge

¹Those in doubt might consider, as Kripke does (1982, p. 18), that this kind of case could be presented by an intelligence tester. In the result analysis methodology, it does not matter that I guessed the answer, only that it was correct — and it is correct if and only if it accords with the application of the rule $n + 5$ to the next n of the existing sequence.

²Cf. for example Pettit 1993, p. 388.

³Oliver Sacks' famous 1973 account, *Awakenings*, of the behaviour of catatonic victims of an encephalitis lethargica epidemic might present a case of intention without knowledge thereof. Sacks mentions (note 14) a case where the patient remained catatonic unless thrown a ball, at which moment he would suddenly leap to his feet, dribble the ball around the ward, toss it from one foot to the other and so on before suddenly stopping and resuming his previous catatonia. Of course, given the assumption that this constitutes intentional action, it is still no proof that the patient did not *know* what he intended. I merely suggest that it might strike us as unlikely given the circumstances.

of what S intends. Suppose that S has an intention, and that the question of this intention's determinate 'aboutness' thus hangs in the balance.⁴ Suppose also that S believes that S intends to conform to rule R1. S's belief may be true — that cannot be denied unless we have a further reason to doubt that S knows that (S intends to conform to R1). The only source of sceptical doubt here is thus the question of whether S's belief is justified.

Justificatory internalism seems to be a premiss of Kripke's discussion, so we must either accept this premiss or enter into a debate about whether we should be internalists or externalists. Neither the assumption that meaning is determined by intention nor KI* (nor their conjunction) entails one or the other without additional supporting premisses. Externalism clearly invalidates the sceptical reasoning even on both assumptions, since denying that I must have any kind of access to what makes my beliefs justified for them to constitute knowledge means that I can know what I intend without having to back up that belief. S's belief that (S intends to follow R1) can likewise be justified without S possibly knowing that it is. Thus the sceptic's proposed difficulty in S's finding justification for any belief that S intends to ϕ is no argument that this belief does not constitute knowledge.

But of course, on this account the proposed difficulty need not even arise. If S cites what S believes is S's intention to conform to R2, the doing of which would entail that S intends to conform to R1, then the trouble here is supposed to be that S would not know that S is justified in the latter belief, hence the situation does not improve. But this also is no argument for the externalist, since S may in fact be justified 'all the way down' for as many rule-following intentions as S can cite, since S may in fact know that S has all these specific intentions.

Still, simply citing the externalist answer is not particularly fair to Kripke, and will be most unsatisfactory to internalists. And this is clearly not the place to settle the internalism/externalism debate. The appropriate question here is just whether adopting internalism is sufficient for Kripke's conclusion that one cannot intend one thing rather than another.

The hypothesis was that S's past actions would be identical whether S's actions conformed to the *plus* rule or to the *quus* rule. The difference can only consist in which rule S intended for S's actions to conform to, but any hypothesis about what S intended is based on evidence that is consistent with S's having intended something else. If it is solely in virtue of the *intention* to conform to R1 that S in fact follows R1 rather than the deviant R1*, then KI* entails that S cannot have this intention unless S knows that S does. On a weaker variety of internalism, S must at least possibly know that S is justified in believing that S intends follow R1. But if the only possible justification of this belief is some further determinate intention, then the latter can only be determinate if S knows its content.

Any justification S can present to this effect is vulnerable to a familiar epistemological scepticism. If S were to claim that S's intention to conform to R2 justifies S's belief that S intends to conform to R1, then a sceptic can challenge not only the claim that it does justify this, but also the knowledge claim entailed (on KI*) by the assumption that S has a determinate intention to conform to R2. This operation can be repeated for any new claim, whence the justificatory regress that sits

⁴Kripke's conclusion is not that we do not have intentions at all, which would be more radical by an order of magnitude than what he does propose, which is that although we have intentions, our intentions do not point one way rather than another.

at the heart of Kripke's challenge.

However, as indicated, there is nothing new about this kind of scepticism. It presents us as usual with Agrippa's trilemma of foundationalism, coherentism or accepting the justificatory regress.⁵ We might assert that some intentions to conform are foundational such that they determine precisely what I will do in any given case;⁶ or in coherentist fashion, that at some point S's beliefs in what S intends justify one another; or else we adopt Kripke's view that neither is satisfactory and just dispense with the whole idea that meaning (hence intention) is determinate. Kripke's sceptic simply applies ancient questions about justificatory internalism to the comparatively new idea that meaning is in some way determined by rule-following. Notable also is Kripke's implicit-here-made-explicit assumption that no answer to this kind of scepticism can be found for the rule-following case. Yet there is nothing special about knowing one's intentions that seems to force this conclusion if we think justificatory internalism does not otherwise necessarily doom us to scepticism.

Moreover, it is not clear where we source the idea that the only possible justification of S's belief that S intends to conform to R1 has to be an intention to conform to some further rule, or even an intention at all. Perhaps S's justification is just the further belief that S always intends just what S believes S intends, which again, may or may not constitute knowledge. Further questions can always be asked, but if we admit this presents an insoluble problem then it is hard to see why the problem cannot be generalised to *all* knowledge claims.

I therefore conclude that Kripke's sceptic presents no essentially new problems by doubting that I know what I intend, and that if Kripke thinks that the problems that *are* presented are insurmountable, then he should probably be a sceptic about knowledge in general.

7.2 Understanding without Knowledge Thereof

Perhaps the problem is not that I do not know what I intend, but only that there is nothing to establish whether I *interpret* it correctly.

Assume I intend to F. Perhaps in order to act on my intention, I must interpret its content correctly. Let's say that to correctly interpret an intention is to *understand* it. Understanding might, for instance, entail knowing what my interpretation of my intention *means*. If to F is to buy a pint of milk, then to interpret my intention correctly is for me to interpret my intention as *to buy a pint of milk*, and also to know what buying a pint of milk *is*. But, the sceptic interjects, perhaps the correct interpretation of this intention is *to quuy a pint of milk*, where to quuy *x* is just to buy *x* except at *t* where it is to sell *x* — and *t* is the present time. How could I know that I have interpreted my intention correctly? Any intentions I have about how to correctly interpret my intentions could also be misinterpreted, and so on into regress.

This *does* seem to present a new twist on scepticism: it is scepticism about the possibility of correctness in cases where whether something is correct is wholly

⁵I do not hereby endorse the trilemma as the exhaustive analysis of *all* the ways out of this kind of scepticism, and present it only as indicative of the highly familiar nature of the problem.

⁶I think McDowell would call such intentions 'super-rigid' — see McDowell 1984, p. 332, and also Cross 2013, §3.2

determined by my intentions, which must always be interpreted. There does not seem to be any *fact* about what would be the correct interpretation of my intention to buy a pint of milk, since how to interpret what I intend is surely determined by some rule I give myself, and yet following a rule is just to *intend* to conform to it, which entails correctly interpreting *that* intention, and so on. No standard of correctness can be appealed to at any point in the regress without claiming that it is interpreted correctly, which is also subject to doubt. Since any standard of correctness can be interpreted as what I intended, “every course of action can be brought into accord with the rule” (Wittgenstein 1953, §201), and the standard of correctness we thought was rigidly determined is in fact as flexible as my freedom to interpret it however it strikes me.

But there is a mistake in this reasoning. Notice that in order to get started on this regress we first have to reject KI* (cf. §7.1): it cannot be the case that I intend something but misinterpret it, since by KI* if I intend something then I also know that I intend it, and knowing what I intend seems to entail understanding (correctly interpreting) what I intend.⁷ So we reject the assumption that I must know what I intend. However, this opens up the possibility that KI* was supposed to shut out: that what I intend is a fact independent of what I believe I intend. Now of course I may interpret my intention incorrectly, since I may falsely believe that I intend something, but this does not preclude the alternative possibility that I interpret my intention correctly, even if I do not know it. That is, perhaps it is a fact that I intend to buy a pint of milk and that the correct interpretation of this intention is as the intention *to buy a pint of milk*, as determined by my similarly determinate intention to conform to a certain rule for interpreting my intentions. There may be some rule that determines whether I interpret my intentions correctly, and it may be a determinate fact that I intend to conform to this rule rather than some compatible deviant, hence a determinate fact whether any interpretation I make of any of my intentions is correct or not, *hence* whether I in fact understand what I intend.

Problems only arise here if we introduce the following sort of constraint on understanding an intention:

KU S understands (S's intention to ϕ) \leftrightarrow S knows that (S understands (S's intention to ϕ))

This is simply the transposition of KI* to the case of correct interpretation (understanding) of intentions. The biconditional formulation is again a consequence of having the truth of P as conditional on knowledge that P, which is the core of what we want, since to know that P is intuitively conditional on the truth of P.

It should be clear at this point that any problems here are generated simply by the same means as they were in §7.1. Externalists face no problem with cases where my belief that I understand my intentions must be justified by appeal to some rule to which I intend to conform, since I do not need to have any kind of access to that justification, nor access to the justification of my interpretation of the rule. Correspondingly, internalists face familiar sceptical concerns which may

⁷I leave open the possibility that interpretation, like knowledge, may be tacit. Still, to know that P seems to entail at least a correct tacit interpretation of P as P, else it is hard to ascribe the content 'P' to what the putative knower knows.

or may not be terminal, but this fact entails the conclusion that understanding intentions is impossible only on the assumption that justificatory internalism results in the impossibility of knowledge in general, which is as unwarranted by Kripke's discussion as it is unacceptable to most internalists.

As such, and contrary to appearances, introducing questions about interpretations of internally determined standards of correctness does not present us with a new sceptical problem. It is either overcome by the assumption that facts about intentions are not determined by knowledge of those facts, or else by justificatory externalism. It generates problems for internalists, but only the problems with which they will already be well acquainted.

7.3 The Normativity Question

In §3.5 I argued that the sceptic is unable to reason from a fairly intuitive conception of semantic facts as non-intentionally-determined to the conclusion that there are no facts about what I mean. Given that the paradox is supposed to ensue from intuitive beginnings, it seems we are owed some argument for thinking meaning is determined by intention. Despite the absence of such an argument in Kripke's discussion, in §§7.1–7.2 we investigated whether an intention-determined conception of meaning would be sufficient to generate the sceptical conclusion. I argued that if intention and correct interpretation thereof are conceived as separable from the possibility of our knowing them, then there is no reason to think they are not determinate, and thus no reason to think that they could not determine meaning. Conversely, if they are conceived as existing only if we know them, then this only presents us with certain familiar sceptical challenges to the possibility of knowledge. Externalists seem to be able to deny these issues, while whether internalists can overcome them is a matter of continuing debate. Either way, the problem generated is nothing in particular to do with the notion of following a rule, since resolving these issues is sufficient to resolve the paradox.

Might we have misunderstood some crucial aspect of the reasoning? After all, despite mentioning it at the beginning, we have not said a tremendous amount specifically about Kripke's most famous commitment: that "the relation of meaning and intention to future action is *normative*, not *descriptive*" (1982, p. 37).

Kripke suggests that we think it is my past understanding of 'plus' that serves to justify or otherwise underwrite my certainty that I *ought* to answer '125' when asked 'What is $68 + 57$?'. The thought here is that understanding or 'grasping' a rule entails normative consequences, at the very least on the supposition that I intend to continue to follow the same rule as I previously did. But of course, "the sceptic doubts whether any instructions I gave myself in the past compel (or justify) the answer '125' rather than '5'" (1982, p. 13). If I told myself to follow R1 rather than R1*, that 'telling' still permits of innumerable interpretations, according to which in this new, *ex hypothesi* unconsidered case, what I ought to answer to continue following the rule could be anything from '5' to 'That's a nice knock-down argument'.

In §3.4 we supposed that meaning could be determined by rules, and that the meaning of an utterance would thus be determined by the semantic rule(s) with which it *accorded*. However, Kripke seems to be committed to the stronger view that the meaning of what I say is wholly determined by the fact that I *follow* some

particular rule. On such a view, to mean *plus* by an utterance of 'plus' is to intend for the utterance to conform to the *plus* rule and for it to actually do so. So to mean the same thing by 'plus' now as I did previously is to intend now for the utterance to conform to the same rule as I intended my previous utterances of 'plus' to conform to, and for both my present and previous utterances of 'plus' to actually conform to this rule.

Here the distinction between meaning and intention to mean is notably important. Following a rule is comprised of intention to conform and actual conformity. To *intend to mean* something by an utterance is best interpreted on this account as merely to intend for the utterance to conform to a particular rule. To *mean* something by it is for the utterance to *actually* conform to the particular rule one intended. But I do not mean *quus* by '+' just by saying ' $68 + 57 = 5$ '. Nothing in that utterance determines that I meant anything at all by '+', since the meaning is determined by which if any rule I intended my utterance to conform to. If I intended the *quus* rule then my utterance did mean *quus* by '+'; but if I intended the *plus* rule, things get a bit more complicated. I obviously made a mistake, and not only an arithmetical one. If I intended to mean *plus* by '+' then I apparently *ought* not to have said what I did. But even if I did not do what was prescribed, the question remains: do I succeed in meaning *plus* by 'plus' here or not?

Clearly I may well have meant *plus* here. There was plausibly not what Kripke calls a 'metalinguistic' error on my part, only an arithmetical one.⁸ This might seem troubling, but it can be explained on reflection that the *plus* rule need not simply mandate the use of 'plus' or '+' in cases of arithmetical correctness. It would likely also mandate the use of 'plus' (utterances of which would thereby still mean *plus*) in cases where for example the speaker believes his utterance to be an arithmetically correct addition. Perhaps it prescribes that if S believes that $68 + 57 = 5$, then in the right context, S ought to say ' $68 + 57 = 5$ '; perhaps the context is a kindergarten class.

More importantly, the meaning/intending distinction helps explain why Kripke sometimes describes meaning what I now do as dependent on 'according with my past intentions'. Whether I have successfully meant *plus* by 'plus' in the past is irrelevant to whether I now do; but we can still make sense of the idea that I mean* the same thing by 'plus' as I always did if we understand this as 'I intend to conform to the *plus* rule and always have'. Perhaps I never meant (semantically) *plus* by 'plus' in the past because I never succeeded in conforming to the *plus* rule — but perhaps it was always my intention to conform to it, and perhaps in this new instance I succeed.

The big question here is just as follows. If intending to conform to a rule and succeeding is necessary for meaning something by one's words, then one must act as the rule prescribes. That is what it is to conform. Apparently, then, there are certain things that I *ought* to say in certain contexts, and following a rule seems to mean guiding one's actions to conform with what they ought to be. But how am I to guide my actions to accord with the rule, if I cannot myself determine which rule I intend to follow, and hence, be sure of what it is that I ought to do?

The answer is surprisingly unproblematic. People do as they merely believe they ought all the time, aiming to do as they really ought, and with varied degrees of success. That S does not know that what S *believes* S ought to do is *in fact* what

⁸For this distinction see Kripke 1982, p. 8.

S ought to do, is no obstacle to it turning out that S's belief is true, nor to S acting as S believes S ought, and happening to act as S really ought to act. The fact that I can be mistaken about what I ought to do does not prevent me from actually being correct, nor intending to act as I in fact ought, nor does it affect facts about what I ought to do. Even on a hard internalism, it is implausible to claim that S cannot guide S's actions by the rule's prescriptions unless S knows which rule S is following: that S ought to ϕ implies that S *can* ϕ , but not that S *knows* that S ought to ϕ .

Kripke might object at this point that I do act with certainty when I answer '125', and I "do not form tentative hypotheses" about what I mean, "wondering what I should do if one hypotheses or another were true" (1982, p. 40).⁹ Since we concede it is possible that I do not know which rule I intend to follow, how in that case is it that I reply so confidently?

The answer should be obvious to anyone who has ever acted confidently on what turned out to be a false belief. The sceptic causes me to doubt that I meant one thing rather than another, but prior to that I was certain because I was confident that I knew what I meant. Moreover, the assumption here is that I do not merely reply with equal certainty that the sceptic's hypothesis is wrong, even if rationally I cannot provide justification for my certainty — which, as has already been established, is neither necessarily the case nor a good cause for doubt unless we surrender to traditional scepticism about knowledge.

So if there is a problem here it really is nothing to do with the involvement of normativity: the problem lies with determining which rule S is following. But as we saw, this is simply a question of which rule S *intends* to conform to and whether S succeeds — and there is no problem with the possibility of intending something, nor of succeeding in something, that is particular to either of those notions. Rather, the problem, if there is one, is just the familiar epistemic scepticism described in §§7.1–7.2 above. Moreover, our earlier hypothesis (§3.4) that meaning is not intention-determined is also compatible with the normativity of meaning: it could be that intention-independent semantic facts prescribe action, since the possibility of my failing to know this does not entail the impossibility of their normative force.¹⁰

7.4 The Constitutive Question

The questionable inference from epistemological arguments to metaphysical conclusions appears several times in Kripke's sceptical dialogue. Despite this, he assures us that the challenge presented is not primarily epistemological in nature:

⁹For more on Kripke's worry about 'tentative hypotheses' here, see Zalabardo 1997, §IV and also my §1.6.

¹⁰Contrary to this, Hattiangadi (2007) argues that the essential prescriptivity of meaning is the sceptic's strongest tool for arguing that meaning is impossible. However, her reason for saying this is the claim that prescriptive normativity is a generally dubious notion for reasons familiar from meta-ethical debates, such that if the sceptic can prove that meaning depends on normativity, the possibility of determinate meaning becomes as troublesome as the possibility of determinate prescriptive moral facts. Yet the most evidence she can find for this interpretation in Kripke is his hinted concern that the relevant norms might be 'queer' (Hattiangadi 2007, p. 48), which is of course a primarily *epistemological* concern about knowing what is prescribed.

“Given, however, that everything in my mental history is compatible both with the *conclusion* that I meant plus and with the *conclusion* that I meant quus, it is clear that the sceptical challenge is not really an epistemological one. It purports to show that nothing in my mental history [or] past behaviour — not even what an omniscient God would *know* — could *establish* whether I meant plus or quus.” (Kripke 1982, p. 21, my emphasis and brackets)

Curiously, as the added emphasis shows, Kripke even couches the supporting reiteration of this supposedly non-epistemological challenge purely in terms of what can be known, deduced or established. Certain internalist commitments might explain this, if for example Kripke takes it that P is true if and only if it can be established that P is true. But as I have argued, permitting the requisite internalism is not sufficient to establish any paradoxical conclusions, since it is compatible with there being facts about what a word means, facts about what I intend, facts about what I thereby ought to do, facts about whether I know or correctly interpret any of this, and therefore, facts about whether I follow a given rule or not. I argued that to create any problems at all, Kripke's sceptic owes us a convincing sceptical argument against the possibility of knowledge, which would present as an entirely familiar puzzle about belief.

However, there is an aspect of the challenge that remains to be covered. I have suggested that the fact as to what I mean by ‘plus’ can easily ‘contain’ the ‘directions’ for what I am to do with that expression in any particular case as Kripke demands (Kripke 1982, p. 11), since this fact consist in my following the rule for meaning *plus* (say) by ‘plus’. If both my intention to conform to this rule and the rule itself are determinate, then there is no reason to question what it is that a rule demands of me in any given application of it. As Hattiangadi (2007) insightfully observes, the force of the sceptical reasoning derives from contrasting a rigid element with a fluid element. We take the rule to be rigid, and our interactions to be fluid, so the question remains: how could rigid meaning result from their combination? My argument so far has been that in fact both are rigid, and that the sceptic must do more to demonstrate otherwise. Yet it may be objected that I am not entitled to this claim.¹¹

I have presented as epistemological a problem Kripke takes to be metaphysical; that is, on my reading, the central question is of the form ‘How do we know that P?’, as opposed to ‘How is it the case that P?’. Moving from epistemological worries to metaphysical ones is no simple matter, and my response to Kripke has been that not enough information is presented to fuel the transition. We have, I argue, no reason to doubt the possibility of facts about intention, and whether we can know what those facts are is at worst an open question. Yet Kripke himself believes the sceptical challenge to be *constitutive*; “An answer to the sceptic... must give an account of what fact it is (about my mental state) that *constitutes* my meaning plus, not quus” (1982, p. 11, my emphasis and ellipsis). The answer I offered may seem bluntly dismissive of the worry, for all I contended was that the fact about what you intend is sufficient to metaphysically determine which meaning-rule you are following, and hence is partially constitutive of what you mean, the remaining

¹¹Thanks are due to the attendees of the UCL Philosophy Graduate Conference 2014 for bringing this to my attention.

constituting to be done by the rule itself and your successfully conforming to it. But perhaps what Kripke really wants is an account of what that *intention* consists in. This is certainly something I have not offered.

On this view of things, the sceptical challenge remains inadequately answered by me. What needs to be said is how it could be the case that I intend to conform to this rule rather than that. Kripke's argument against possible accounts of meaning thus becomes a demonstration that none of the putative facts about what I intend is actually a fact to that effect. Faced with the inadequacy of so many alternative accounts (well, a few at least), the burden is surely on someone in my position to present a satisfying account of intention — otherwise it would seem the sceptic has won.

Still, this is a curious piece of reasoning. It sits well within a certain conception of epistemic virtue: it would seem that the sceptic has done enough to warrant the *belief* that there are no facts about intention, pending evidence to the contrary. But that is no metaphysical proof. What needs to be shown to support the *metaphysical* conclusion is that there is some inherent problem with the very possibility of an intention-fact. That is, it must be shown that the putative fact that S intends to ϕ is either identical or consistent with the putative fact that S intends to $\neg\phi$, or that assuming it produces some other contradiction. Yet I cannot envisage how this could be achieved. Kripke's discussion focuses on our alleged inability to provide support for either the sceptic's hypothesis or my own about what I mean by 'plus'. The argument is thus entirely negative, to the effect that neither of us has the tools available, and does not transcend a strategy of *disconfirmation* to provide a *disproof*. Moreover, if the epistemological challenge could not be met, it would still change nothing about the *fact* as to whether intention is possible. The inability to build a case for either hypothesis does not amount to disproof of either. So at its very best the paradoxical conclusion that there are no intention-facts amounts only to our best supported theory, the paradox consisting in our inability to disprove it — assuming that we do not manage this. However, this falls short of the metaphysical point Kripke portends to have proven, unless we assume some inferential link between lack of proof and lack of facts. To do so would be to embrace anti-realism without motivation internal to the sceptical reasoning.

I do not aim to provide a sceptic-busting account of intention here, but it does seem important to analyse how the sceptical considerations are supposed to impede my doing so. Say we hypothesise that S intends to ϕ , and the sceptic proffers the conflicting hypothesis that S intends to ψ . My argument in this chapter has been that our inability to decide the matter does not result in the indeterminacy of S's intentions, since the fact of our inability to find determinacy in the situation is not identical with the facts of the situation's actual indeterminacy. The constitutive argument, however, is that *nothing* could decide this matter; there is just no way that anything could amount to S's intending to ϕ rather than intending to ψ . As discussed, this is something of a conjecture, an inductive inference drawn from the frustration of a small sample of our best efforts to account for what S intends. But suppose this worries us enough to serve as a *reason to believe* that S intends nothing in particular; the onus is then on the counter-sceptic to provide a better reason to the contrary, and the only thing that would seem to serve would be a watertight explanation of what S's intention to ψ would consist in.

Is this what Kripke's subsequent discussion — the famous anti-dispositional argument, the fairly off-hand rejection of anti-reductionism, and of course the regress of rules — is really about? It is hard to see how this supports the metaphysical claim, if so. In the latter case, the regress of rules, the claim was that I cannot be following some rule R by virtue of some rule-for-following-rules, since that does not explain what my following the second-order rule consists in, and so on. Transposed to the intentional case, this seems to be the point that I cannot intend to ϕ by virtue of my intending to intend to ϕ , as this likewise leaves intention unexplained. "Maybe you in fact intended to intend to ψ ," the sceptic will interject. And of course, the point is simply that nothing has been said about intention so far that disproves this or any number of conflicting hypotheses about the bottom-level intention. That is, S's intention to ϕ cannot consist in something 'of the same character as that which preceded it in the regress'. There must be more to intention to this.

Some have attempted to counter Kripke's worry by giving an anti-reductionist account of rule-following. Ginsborg, for example, takes it that there is some primitive mental state of feeling that something or other is appropriate, and that this 'primitive normativity' is sufficient to answer Kripke's sceptic (Ginsborg 2012). However, considering this strategy makes me wonder why taking intention itself as a primitive phenomenon is not sufficient to answer the issue. Could we not contend that to intend to ϕ is a neuro-psychological phenomenon? I confess general ignorance with regards to the empirical literature on intention, but Kripke professes no more expertise in this area than I in his discussion. Clearly the recursive definition of intentions in terms of intentions is deficient, but why would we expect cognitive phenomena to be ultimately explicable in terms of other cognitive phenomena anyway? Certainly we can imagine cases where the neurological difference between S's intending to ϕ and S's intending to ψ would defy measurement, but again, our contingent inability to determine a fact does not result in its indeterminacy.

In this chapter it was argued that the meaning of my words is plausibly determined irrespective of what I intend them to mean, and that if so, this is sufficient to invalidate Kripke's sceptical argument that all language is meaningless. However, I then argued that this conclusion does not follow even if meaning is assumed to be determined by intention, since either there are facts about what I intend whether I know them or not, or else we merely face some familiar epistemic questions which are by some accounts adequately solved, and by others, a work in progress, but which few would concede are fatal to the knowledge claims in question.

I then argued that Kripke's most famous claim, that in order to follow a rule I must deliberately accord my actions with what the rule prescribes, does not in fact present any further problems beyond those already considered. In the cases of intention, understanding and normativity, the only further question is a constitutive one: *what is it* for me to intend to ϕ , to interpret this intention *correctly*, or for it to be the case that I *ought* to ϕ ? And yet the ability to ask these questions in perpetuity does not underwrite the stronger claim that *nothing* is constitutive thereof.

I conclude, therefore, that there is nothing especially problematic or paradoxical about the notion of following a rule, since Kripke's sceptical challenge is novel only in its application of the traditional problem of knowledge to a new case. I

contend that the resolution of Kripke's rule-following paradox is simply a matter of facing up to standard epistemological scepticism.

Chapter 8

Conclusion

Kripke's problem derives from a legitimate but not insurmountable concern about the possibility of inner knowledge — specifically, of knowledge about what I intend. The secondary challenge derived from this is that of determining the course of action that ought to be followed — or on a different interpretation, what meaning-fact can suitably justify a choice of action — as derived from those very same intentions, and this too is an epistemological question. In the context of these challenges I have considered accounts that have the following in common: they attempt to explain how the second challenge can be met, for the case of meaning one thing rather than another. They ask: how is it that I ought to say something as determined by the rules of our language? Yet they also variously treat this as a *'how possible?'* question, asked in response to Kripke's *'not possible'* challenge, and thereby inherit Kripke's premiss that knowledge of one's own intentions is an unachievable prerequisite to following any rule. I have denied the latter on the basis that it is undermotivated, and so am optimistic that if a solution to the more familiar kinds of epistemic scepticism can be found, then as a result, so too can an answer for Kripke's sceptic.

I conclude, then, that contrary to widespread belief, Kripke's paradox does not turn on the normativity of meaning, so much as on something like the idea of 'conscious engagement' described by Zalabardo. And yet even this premiss, so it seems, is not the main obstacle: the main difficulty lies in knowing, when one believes oneself to be consciously engaged with the relevant meaning-determining facts, that this is genuinely the case.

Bibliography

- Blackburn, S. (1984). 'The Individual Strikes Back'. *Synthese*, 58:281–301.
- Boghossian, P. (1989). 'The Rule-Following Considerations'. *Mind*, 98(392):507–549.
- Boghossian, P. (2012). 'Blind Rule-Following'. In Coliva, A., editor, *Mind, Meaning, and Knowledge*, pages 27–48. Oxford: Oxford University Press.
- Bridges, J. (2014). 'Rule-Following Skepticism, Properly So Called'. In Conant, J. and Kern, A., editors, *Varieties of Skepticism: Essays After Kant, Wittgenstein, and Cavell*. Berlin: De Gruyter.
- Cross, J. A. (2013). Rule-Following, Normativity & Objectivity. Master's thesis, UCL, London.
- Cuneo, T. (2007). *The Normative Web: An Argument for Moral Realism*. Oxford: Oxford University Press.
- Dummett, M. (1959). 'Wittgenstein's Philosophy of Mathematics'. *The Philosophical Review*, 68:324–348.
- Field, H. (1972). 'Tarski's Theory of Truth'. *The Journal of Philosophy*, 69(13):347–375.
- Fodor, J. A. (1990). *A Theory of Content and Other Essays*, chapter 'A Theory of Content, I: The Problem'. Cambridge, MA and London, UK: The MIT Press.
- Frege, G. (1997 [1918]). 'Thought'. In Beaney, M., editor, *The Frege Reader*, pages 325–345. Blackwell.
- Ginsborg, H. (2010). 'Review of *Oughts and Thoughts: Rule-following and the Normativity of Content*, by Anandi Hattiangadi'. *Mind*, 119(476):1175—1186.
- Ginsborg, H. (2011a). 'Inside and Outside Language: Stroud's Nonreductionism About Meaning'. In Bridges, J., Kolodny, N., and Wong, W.-H., editors, *The Possibility of Philosophical Understanding: Essays for Barry Stroud*. Oxford: Oxford University Press.
- Ginsborg, H. (2011b). 'Primitive Normativity and Skepticism About Rules'. *The Journal of Philosophy*, CVIII(5):227–254.
- Ginsborg, H. (2012). 'Meaning, Understanding and Normativity'. *Proceedings of the Aristotelian Society Supplementary Volume*, 86(1):128–146.

- Glüer, K. (2001). 'Dreams and Nightmares: Conventions, Norms and Meaning in Davidson's Philosophy of Language'. In Kotako, P., Pagin, P., and Segal, G., editors, *Interpreting Davidson*, pages 53–74. Center for the Study of Language and Information.
- Glüer, K. and Pagin, P. (1999). 'Sense and Prescriptivity'. *Acta Analytica*, 14:111–128.
- Glüer, K. and Wikforss, A. (2009a). 'Against Content Normativity'. *Mind*, 118(469):31–70.
- Glüer, K. and Wikforss, A. (2009b). 'Against Normativity Again: Reply to Whiting'.
- Glüer, K. and Wikforss, A. (2010). 'The Normativity of Meaning and Content'. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Winter 2010 edition. Archived at <http://plato.stanford.edu/archives/win2010/entries/meaning-normativity/>.
- Glüer, K. and Wikforss, A. (2013). 'Against Belief Normativity'. In Chan, T., editor, *The Aim of Belief*, pages 80–99. Oxford: Oxford University Press.
- Haddock, A. (2012). 'Meaning, Justification, and 'Primitive Normativity''. *Proceedings of the Aristotelian Society Supplementary Volume*, 86(1):147–174.
- Hattiangadi, A. (2006). 'Is Meaning Normative?'. *Mind & Language*, 21(2):220–240.
- Hattiangadi, A. (2007). *Oughts and Thoughts: Rule-Following and the Normativity of Content*. Oxford: Oxford University Press.
- Hattiangadi, A. (2009). 'Some more thoughts on semantic oughts: a reply to Daniel Whiting'. *Analysis*, 69(1):54–63.
- Horwich, P. (1998a). *Meaning*. Oxford: Oxford University Press.
- Horwich, P. (1998b). *Truth*. Oxford: Oxford University Press, 2nd edition.
- Horwich, P. (2012). 'Regularities, Rules, Meanings, Truth-Conditions, and Epistemic Norms'. In Coliva, A., editor, *Mind, Meaning, and Knowledge*, pages 77–95. Oxford: Oxford University Press.
- Hume, D. (2007 [1739–40]). *A Treatise of Human Nature*. Eds. D. F. Norton and M. Norton. Oxford: Oxford University Press.
- Kripke, S. (1982). *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Oxford: Blackwell.
- Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*. London: Penguin.
- McDowell, J. (1984). 'Wittgenstein on Following a Rule'. *Synthese*, 58(3):325–363.
- McDowell, J. (1996). *Mind and World*. Cambridge, MA and London, UK: Harvard University Press, 2nd edition.

- McDowell, J. (1998). *Mind, Value, & Reality*. Cambridge, MA and London, UK: Harvard University Press.
- McDowell, J. (2009). *The Engaged Intellect: Philosophical Essays*. Cambridge, MA: Harvard University Press.
- Millar, A. (2002). 'The Normativity of Meaning'. *Royal Institute of Philosophy Supplement*, 51:57–73.
- Millar, A. (2004). *Understanding People: Normativity and Rationalizing Explanation*. Oxford: Oxford University Press.
- Miller, A. and Wright, C., editors (2002). *Rule-Following and Meaning*. Chesham, UK: Acumen.
- Millikan, R. G. (1984). *Language, Thought and Other Biological Categories*. Cambridge, MA and London, UK: The MIT Press.
- Millikan, R. G. (1990). 'Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox'. *Philosophical Review*, 99(3):323–53.
- Nagel, T. (1986). *The View from Nowhere*. Oxford: Oxford University Press.
- Papineau, D. (1999). 'Normativity and Judgment'. *Proceedings of the Aristotelian Society Supplementary Volume*, 73:17–43.
- Peacocke, C. (2012). 'Understanding and Rule-Following'. In Coliva, A., editor, *Mind, Meaning, and Knowledge*, pages 49–76. Oxford: Oxford University Press.
- Pettit, P. (1993). 'Problem of Rule-Following'. In Dancy, J. and Sosa, E., editors, *A Companion to Epistemology*, pages 386–391. London: Blackwell.
- Quine, W. V. O. (1976 [1936]). 'Truth by Convention'. In *The Ways of Paradox and Other Essays*, pages 77–106. Cambridge, MA: Harvard University Press.
- Rosen, G. (1994). 'Objectivity and Modern Idealism: What is the Question?'. In O'Leary-Hawthorne, J. and Michael, M., editors, *Philosophy in Mind*, pages 277–319. Kluwer.
- Rosen, G. (2001). 'Brandom on Modality, Normativity and Intentionality'. *Philosophy and Phenomenological Research*, 63(3):611–23.
- Searle, J. (1980). 'Minds, Brains, and Programs'. *Behavioral and Brain Sciences*, 3:417–457.
- Shah, N. and Velleman, J. D. (2005). 'Doxastic Deliberation'. *Philosophical Review*, 114(4):497–534.
- Stroud, B. (2002). *Meaning, Understanding, and Practice: Philosophical Essays*. Oxford: Oxford University Press.
- Stroud, B. (2011). *Engagement and Metaphysical Dissatisfaction*. Oxford: Oxford University Press.

- Whiting, D. (2007). 'The Normativity of Meaning Defended'. *Analysis*, 67.2(April):133–140.
- Wikforss, A. (2001). 'Semantic Normativity'. *Philosophical Studies*, 102(2):203–226.
- Wittgenstein, L. (2010 [1953]). *Philosophical Investigations*. Trans. G. E. M. Anscombe, P. M. S. Hacker and Joachim Schulte. John Wiley & Sons, 4th edition.
- Wright, C. (1980). *Wittgenstein on the Foundations of Mathematics*. London: Duckworth.
- Wright, C. (1984). 'Kripke's Account of the Argument against Private Language'. *The Journal of Philosophy*, 81:759–777.
- Wright, C. (1989a). 'Critical notice of Colin McGinn's *Wittgenstein on Meaning*'. *Mind*, 98(390):289–305.
- Wright, C. (1989b). 'Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics'. In George, A., editor, *Reflections on Chomsky*, pages 233–264. Oxford: Blackwell.
- Wright, C. (1992). *Truth and Objectivity*. Cambridge, MA and London, UK: Harvard University Press.
- Wright, C. (2001). *Rails to Infinity*. Cambridge, MA: Harvard University Press.
- Wright, C. (2007). 'Rule-Following without Reasons: Wittgenstein's Quietism and the Constitutive Question'. *Ratio*, XX(4):481–502.
- Zalabardo, J. (1997). 'Kripke's Normativity Argument'. *Canadian Journal of Philosophy*, 27(4):467–488.
- Zalabardo, J. (2000). 'Realism Detranscendentalized'. *European Journal of Philosophy*, 8:63–88.