

University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.

Data: Author (Year) Title. URI [dataset]

University of Southampton

Faculty of Humanities

School of Philosophy

**“Going Through the Motions of Weaving”: Wittgenstein on a Characteristic
Syndrome of Modern Philosophical Thought**

by

Kristen Jeffs

Thesis for the degree of Doctor of Philosophy

February 2018

University of Southampton

Abstract

Faculty of Humanities

School of Philosophy

Thesis for the degree of Doctor of Philosophy

“Going Though the Motions of Weaving”: Wittgenstein on a Characteristic Syndrome of
Modern Philosophical Thought

by

Kristen Jeffs

The principal thesis for which I will argue is that: not only is Wittgenstein, as is too often thought, not putting forth some naturalistic explanation, or scientific theory of meaning, language or of intentional mental concepts. He wishes to persuade us of something much stronger: not only is he not trying to offer any such explanation in the face of philosophical problems of meaning and mind, but Wittgenstein’s claim is that *the desire for explanation in regard to understanding linguistic and mental concepts is itself the problem*. In order to defend this thesis, however, there are three principal impediments which, I believe, serve to stand in the way of our being able to grasp the import of Wittgenstein’s arguments, and I aim to clear these impediments away. The first is that the huge influence of Saul Kripke’s celebrated reading is both a blessing and a curse for understanding the import of Wittgenstein’s discussion. The second impediment which I attempt to clear away, is a characteristic tendency to receive Wittgenstein’s insights within the very explanatory framework of thought which he is trying to train us out of; and hence not to receive Wittgenstein’s insights at all. The third impediment is that, whilst John McDowell has grasped the import of Wittgenstein’s insights into the desire for explanation of linguistic and mental concepts, his understanding of these insights remains obscured from the view of the vast majority of published commentators, who are distracted by an early published paper by McDowell which betrays a form of the very misunderstanding Wittgenstein is trying to warn us against. What has gone unremarked is that McDowell later explicitly repudiates his early reading, once he has grasped the import of Wittgenstein’s discussion. By clearing these impediments away, my hope is that Wittgenstein’s profound insights can be made more widely available to contemporary thought.

Table of Contents

Table of Contents	i
Research Thesis: Declaration of Authorship	iv
Acknowledgements	v
Introduction	1
Chapter 1 Kripke on Wittgenstein’s Rule-following Considerations	13
Introduction	13
§1 The Constitutive Nature of Wittgenstein’s Discussion of Rules.....	21
§2 Kripke and Boghossian Contra Dispositionalism	35
§3 Generality of the Putative Problem	45
§4 Kripke and Wight on “Non-misinterpretable Interpretations”	53
Conclusion.....	71
Chapter 2 From Conceptual Nihilism to Superlative Mythology (and back again....)....	75
Introduction	75
§1 Kripke’s Community Appeal	87
§2a Metaphysical Branch of Wright’s Reading of Wittgenstein	99
§2b Epistemological Branch of Wright’s Reading of Wittgenstein	111
§2b(i) Wright Manoeuvres Against the Sceptic	113
§2b(ii) Putative Paradigms of the Mental	119
§2c “A Gangway with Rigid Walls”	143
Conclusion.....	159
Chapter 3 “Turning Our Examination Around...”	163
Introduction	163
§1 McDowell’s Early Reading (or: More of the Same...)	167
§2 McDowell’s Later Reading (or: A Radical New Departure...)	183
Conclusion.....	215
Conclusion	227
Bibliography	233

Dedicated to the loving memory of

Michael James Dally

&

Pamela June Jeffs

...but the absence of treasure,

Absence of whatever would return the world

To the strangeness that as children we embraced

And recognised as *life*. *Rave on*.

B.H. Fairchild

Research Thesis: Declaration of Authorship

Print name:	Kristen Jeffs
-------------	---------------

Title of thesis:	"Going Through the Motions of Weaving": Wittgenstein on a Characteristic Syndrome of Modern Philosophical Thought
------------------	---

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signature:		Date:	
------------	--	-------	--

Acknowledgements

I would like to thank the University of Southampton for the opportunity to undertake this research, and to the Faculty of Humanities for their generous funding. Gratitude for forbearance and all the many years of support and guidance must go to my thesis supervisor Prof. Genia Schönbaumsfeld (no more talk of witches!)

I could not have completed this journey without the boundless love and support from my parents (not to mention “comma-control” from Dad – much obliged!) Sadly, my dear Mum is not here with us to see the journey to completion, but she has still very much been with me in spirit, offering her encouragement right until the very end. You will never leave me.

Thanks to my sisters Steph and Trace, to Dean, Leon, Stefan, and Luca, and all the rest of my family—both Jeffs and Stephen. I must also give acknowledgements to Justina, Alice, the boys back home (and Egg), everyone past and present at Wandsworth Council who were always asking how things are going (and I can’t not mention Carol by name, nor Andy and Kat), Lloyd and Ludo, and Ad. Ac i Omar a Ryan – *mae gen i fwy o amser i chwarae yn Llundain nawr, bechgyn!*

But when they sought him, he could
not be found. So they inquired again
of the LORD, 'Where is the man?'
And the LORD answered, 'Look, he has
hidden himself among the baggage.'

1 Sam. 10: 21-22

Introduction

§138 of the *Philosophical Investigations* begins a series of remarks, culminating at §242, which have since come to bear the title “the rule-following considerations.” This series of remarks constitute only one of ‘a number of sketches,’ as Wittgenstein says in the *Preface*, of the vast terrain which mark his ‘long and involved’ philosophical travels across ‘a wide field of thought’ —but they are a central district of that landscape nevertheless. As the label suggests, this series of remarks contain the fruit of Wittgenstein’s sustained reflections, ‘criss-cross in every direction,’ on the notion of responding to, or behaving in a way which manifests understanding of the meaning or significance of a word or a rule. As Crispin Wright has observed, ‘the principal issues to do with rule-following impinge on every normatively constrained area of human thought and activity: on every institution where there is right and wrong opinion, correct and incorrect practice’ (Wright 2001: 1). And that means the principal issues impinge on every area of human thought and action, and on every institution.

A quarter of a century ago, John McDowell lamented the fact that ‘the thrust of Wittgenstein’s reflections [on following according to a rule] are often misconceived. The result is that...this part of his legacy is still not widely available to contemporary philosophy’ (McDowell 1998b: 263). Unfortunately, this is still more than ever the case to this day. The burden of the present work is to attempt to rectify this sorry state of affairs. The principal thesis for which I will argue is that: not only is Wittgenstein, as is too often thought, not putting forth some naturalistic explanation, or scientific theory of meaning, language or of intentional mental concepts. He wishes to persuade us of something much stronger: not only is he not trying to offer any such explanation in the face of philosophical problems of meaning and mind, but Wittgenstein’s claim is that *the desire for explanation in regard to understanding linguistic and mental concepts is itself the problem*.

In order to defend this thesis, however, there are three principal impediments which, I believe, serve to stand in the way of our being able to grasp the import of Wittgenstein’s

arguments. Thus I will endeavour to clear these impediments away. The first impediment, to be discussed in Chapter 1, is the influence of Saul Kripke's celebrated reading of Wittgenstein on rules and of rule-following. I will argue that the huge influence of Kripke's treatment is a double edged-sword. Over and above the fact that Kripke was responsible for bringing the previously neglected, but central, rule-following considerations within earshot of the community of academic philosophers (at least within the Anglo-American sphere of influence), there are two blessings of Kripke's reading. The first blessing is that Kripke aims to bring out the decidedly normative, or constitutive nature of the issues raised by Wittgenstein. The second blessing is that Kripke, most notoriously, gives vivid and forceful expression to the devastating and quite general nature of the "sceptical paradox" which he famously unearthed from the rule-following considerations. On the other hand, however, Kripke's reading serves as a curse on the reception of Wittgenstein as Kripke's treatment of the issues, I will argue, serve to veil, at the very same time as it seeks to unveil, these very same two crucial features.

In regard to the first blessing, Kripke aims to bring out the decidedly normative, or constitutive nature of the issues discussed by Wittgenstein. As he recognises, Wittgenstein's principal question is 'How can we show *any language* at all...to be *possible*? (Kripke 1982: 62). In Wittgenstein's own words, 'our investigation...is directed...towards the "*possibilities*" of phenomena' (*PI* §90).¹ In other words, the principal issue is not that of classical epistemological scepticism—how can we *know*, say, whether what the sense of an assertion determines is true or false—but, rather, the *constitutive* question of how an utterance can so much have a determinate sense; or of how a word can so much be used according to a determinate meaning; or how a thought, or an intentional mental state more generally, can have a specific content. As I hope will become clearer throughout the present work, the fundamental issue, then, concerns the very *possibility* of thought or deed undertaken according to an understanding of a correctness condition, or of a constitutive connection.

¹ Throughout the present work, I will make reference to the numbered remarks of *Philosophical Investigations* in this manner. All other references will continue in the Harvard style, unless otherwise indicated.

The second blessing of Kripke's reading is that he brings out very forcefully, and most notoriously, the devastating and general implications of the "sceptical paradox," which seemingly shows that '*all* language, *all* concept formation, to be impossible, indeed unintelligible' (Kripke 1982: 62). The paradox seemingly threatens more than the possibility of using a word according to a determinate meaning, however, and seemingly impugns the possibility of formulating a thought, or a belief, desire, wish, or command—any intentional mental state whatsoever—with a specific content. This is because, as Kripke says, for any 'mental item', according to the "sceptical paradox", can be 'interpreted so as to accord with anything we may choose to do': in other words, 'no matter what is in my mind at a given time, I am free to interpret it in different ways' (ibid: 55, 107). The implications are, then, completely general: the "sceptical paradox" apparently renders all and any behaviour undertaken according to a standard of correctness as incoherent. In Kripke's terse summation: 'there can be neither accord nor conflict' (ibid: 55).

Despite these blessings, however, the influence of Kripke's reading serves, I will argue, as a curse over the reception of Wittgenstein's rule-following considerations by obscuring, at the very same time as it seeks to unveil, these very same two features—the constitutive nature of the issues, and the general and devastating implications of the paradox—both of which are indeed integral to a proper understanding of the import of Wittgenstein's arguments. Kripke's treatment confuses matters by almost continually assuming that the source of the justification for one's present use of a word must be found, not in the word's meaning, but in one's *previous intentions* regarding the meaning of the word. Kripke presents the problem in this "transtemporal" form in order to avoid the obviously absurd and self-refuting implications of the paradox ("all language is meaningless!") for long enough that his sceptic can be seen to be in a position to even formulate the problem at all. But, in so doing, Kripke thereby obscures the decidedly constitutive nature of the issues, which he otherwise wishes to accentuate.

The further curse of Kripke's treatment of the rule-following considerations, for which I will argue, is his presentation of the "private language dialectic" within which there is

enacted a subtle shift from the general and devastating implications of the constitutive paradox, to Wittgenstein's supposed acceptance of the watered-down "sceptical conclusion". According to the "sceptical conclusion", talk of determinate meaning is only incoherent *when we consider an individual in isolation*, and, so, its implications are that attributions of meaning to particular speakers ("Jones means addition by '+'") lack factual content. This move to a watered-down version of the implications of the paradox is mirrored in an analogous move, made contemporaneously by the prominent Wittgenstein commentator Crispin Wright, who concludes from the paradoxical reasoning that unilateral recognition of a ratification-independent standard of correctness should be rejected as incoherent. In both cases, this opens the way for each commentator to make appeals to the wider linguistic community as putatively enabling us to live with, or to avoid, the implications of the paradox. As we shall discuss in Chapter 2, Kripke's "sceptical solution" is based on the claim that, although attributions of meaning to speakers lack factual content, they nevertheless have a function—"conditions of assertability"—which serve to place an individual speaker within the context of her wider linguistic community.

More or less concurrently, although independently of Kripke, Crispin Wright also began to develop a theory in response to Wittgenstein's paradox which relied directly on some form of "community appeal" (although Wright's *constitutive* brand of Communitarianism—the dispositions of the linguistic crowd stand judgement over the linguistic inclinations of the solitary speaker—differed from Kripke's decidedly *non-constitutive* Communitarianism). Wright soon came to abandon his earlier crude Communitarianism, however, together with any pretensions that Wittgenstein shared sympathy with any such theory. Subsequently, he began to develop a more subtle and sophisticated response to the paradox, a response with two interlinking branches: on the one hand, a metaphysical branch, dealing with the reality of rules, and, on the other, an epistemological branch, dealing with our abilities to follow according to rules. The "third-personal" (metaphysical) branch of Wright's two-pronged response aims to respect Wittgenstein's explicit rejection of the paradox as based on a misunderstanding: 'an inclination to say: any action according to the rule is an interpretation' (*PI* §201). This forms a contrast with Kripke's reading, which notoriously mis-attributed acceptance of

the implications of the “sceptical paradox” (or, more accurately, the “sceptical conclusion”) to Wittgenstein.

As we shall see in Chapter 2 (but will not become clearer until Chapter 3), however, Wright makes a fateful error in this branch of his reading of Wittgenstein. Rather than give full weight to Wittgenstein’s diagnosis of the misunderstanding—‘an inclination to say: any action according to the rule is an interpretation’—Wright runs straight away to give explanatory substance to the obverse of the misunderstanding: ‘a way of grasping a rule which is *not* an *interpretation*’ (*PI* §201). This failure to give due weight to Wittgenstein’s misunderstanding leads Wright to make a fateful assumption in his reading of *PI* §201, an assumption which embodies a form of the very misunderstanding Wittgenstein is concerned to warn us against: Wright reads Wittgenstein as ‘allowing, apparently, that there is indeed an ineliminable multiplicity of conflicting possible interpretations’ of the expression of a rule (Wright 2001d: 124). The implication of this fateful assumption is that Wright thus identifies rationality itself to be a species of interpretation. Thus, when Wright turns to consider a positive account of the obverse of the misunderstanding diagnosed by Wittgenstein, the substance Wright reads into this is that ‘something other than the exercise of interpretative (rational) faculties enters into the capacity to “read” another’s behaviour’ (Wright 2001e: 124). And this something is, according to Wright, “subrational propensities” or “classificatory dispositions”, which supposedly guide our linguistic and mental behaviour.

The “first-personal” (epistemological) prong of Wright’s reading of Wittgenstein starts off in similar fashion to the first, in that we are led once more into a particular version of Wittgenstein’s misunderstanding: the “inclination to say: any action according to the rule is an interpretation”. Wright begins the epistemological branch of his reading by reaching a very important insight: that the challenge put forth by Kripke’s sceptic is founded on an unsupported reductionism: that is to say, when the sceptic challenges one to produce a meaning-constituting fact which shows that one meant addition, rather than quaddition, by one’s use of the word “plus”, it is presumed that one cannot simply cite the fact that: “I meant, or mean addition by ‘plus’” (or some such). That is to say, the sceptic assumes that

any response must be essentially inferential. As I will argue in Chapter 3, this insight is fundamental to an understanding of the import of Wittgenstein's diagnosis of the paradox as based on misunderstanding.

Wright mars this crucial insight, however, by assuming that any response to the sceptic must, nevertheless, be some species of *interpretation*: that is to say, 'the substitution of one expression of the rule for another' (*PI* §201). Thus, Wright is convinced that 'the challenge is to recall some *independently characterised fact*' that one meant addition rather than quaddition by "plus" (Wright 2001g: 176). Thus, despite his objections against Kripke's sceptic, Wright falls into the sceptic's snare by still assuming that one cannot directly respond to the challenge by simply citing the fact that one meant, or means addition by "plus". Rather, Wright assumes that the response must be in terms of something else, and the "something else" with which we need to rebut the sceptic is, Wright suggests, facts about intentional mental states, particularly intention ("I intended to use "plus" to denote addition"). What makes appeal to intentional mental states seem promising is that they share crucial features in common with meaning: avowals of intention are, like self-knowledge of the meaning of one's own words, *immediate or non-inferential*. Moreover, no less than the meaning of a word, which one may call upon to apply in an indefinite number of situations, intentions are "infinitely fecund", as Wright puts it. Yet, although such parallels between meaning and intention provide hope for some non-question-begging direct response to the sceptic, as I will argue in Chapter 2, the very same similarity casts a dark shadow over any such promise—for the same paradoxical chain of reasoning can be brought to bear against the content of intentional mental states, no less than against the determinacy of meaning.

The point of Chapter 2 is entirely negative. I will argue that, despite the impressive intellectual dexterity and ingenuity on display, neither Kripke's nor Wright's respective responses to the paradox enable us to avoid its devastating implications. This is because, despite both commentators explicitly rejecting dispositionalist theories of meaning as an adequate response to the paradox, each of their respective pictures of linguistic life are as one in conceiving linguistic behaviour as, at bottom, brutally dispositional. Neither

Kripke's "sceptical solution" nor Wright's two-pronged "straight solution" is therefore, as I will argue in Chapter 2, a solution at all. Building on Kripke's own, and Paul Boghossian's further, arguments against dispositionalist theories (discussed in Chapter 1), I will show that neither Kripke's nor Wright's response can avoid the dilemma which Wittgenstein delineates at *PI* §§185-188. *Either*, putatively linguistic behaviour is conceived as merely brute, in which case it cannot avoid the problem of underdetermination, falls prey to the regress of interpretations argument, and then determinate meaning seemingly vanishes into thin air. *Or*: the theories end up presupposing the meaning that is meant to be being explained to us, to the detriment of any explanatory ambitions. Moreover, as I aim to make clear in Chapter 2, for so long as putatively linguistic behaviour is conceived in merely brute, dispositional terms, then determinate meaning can only show up in an ultra-objective, superlative manner. In a mathematical context, the most potent imagery is of the principle of a series extending like a super-hard rail into infinity, and which, through some occult process is able to guide and direct our merely brute behaviour in response to finite expressions of the infinite, mythological rule. But as we shall see in Chapter 3, building on arguments from the first two chapters, a version of the mythology can also show up in various forms, including in the guise of "quasi-magical" mental states and processes. Given that this is the very dilemma which Kripke's and Wright's respective responses were supposed to enable us to avoid, neither provide any workable solution to the paradox.

As I mentioned, the thrust of Chapter 2 is entirely negative. Thus it may seem that my dogged objections to Wright's response to the paradox is something of an attempted hatchet-job. However, this is meant as a compliment to Wright in the sense that I am taking him as paradigmatic of the very style of thinking that Wittgenstein is trying to warn us against, even as Wright attempts to absorb Wittgenstein's thought. And this brings us to the second impediment to the reception of Wittgenstein's reflections on rules and of rule-following, of the three which I mentioned earlier. And this second impediment is the tendency for commentators to receive Wittgenstein's thought within a mode of reflection and habits of thinking which, in actual fact, Wittgenstein is trying to train us out of. The result is to warp and deform the import of Wittgenstein's arguments, and to fail to hear

the message he is trying to impart—that trying to provide theoretical explanations of concepts such as meaning and understanding is futile.

There are two principal reasons why I have chosen Wright as paradigmatic of this impediment to the reception of Wittgenstein's thought. The first is that Wright is candid enough to acknowledge that, as he himself puts it: 'I have persistently attempted to understand and appropriate [Wittgenstein's] thinking within the resources and idiom of a philosophical tradition from which he deliberately stood aside' (Wright 2001: viii). The second reason for which I have chosen Wright as emblematic of such thinking, however, is that Wright has engaged with important Wittgensteinian themes over an extended period and, in my opinion, takes the traditional framework of thought to its very limits in face of Wittgenstein's discussion, and with a clarity of expression which is only to be admired. If I am almost entirely critical of Wright in Chapter 2, it is because I am taking him to be paradigmatic of a widespread style of thought which, as I will argue, Wittgenstein reveals as hopeless. Thus, I hope the trenchant criticisms of Wright will be mitigated by my paying him the complement of taking his particular response to Wittgenstein's paradox to hold a significance of more universal import. It is only proper of me also to register that my own understanding of the later Wittgenstein, such as it is, has been helped immeasurably through critical engagement with Wright's reading.

Failure to heed Wittgenstein's message, as I hope will become clearer as we proceed through the present work, only condemns us to trace ever decreasing circles of thought around the very misunderstandings Wittgenstein is trying to warn us against. As we shall see, thought captured by explanatory ambitions is doomed to reverberate hopelessly between the two horns of the dilemma delineated at *PI* §§185-188: threat from the regress of interpretations, on the one side; and various forms of superlative, Platonistic mythology on the other. Remarking on the movements of thought which are characteristic of thinking which has been captured by the misunderstanding he aims to clear away (including that of his earlier self), Wittgenstein observes: 'You think that after all you must be weaving a piece of cloth: because you are sitting at a loom—even if it is empty—and going through the motions of weaving' (*PI* §414).

John McDowell, at least in his later reading, is one commentator who does heed Wittgenstein's warnings, and who does aim to quell the explanatory ambitions which, as I hope to show as the work proceeds, necessarily implicate us in "going through the motions of weaving". However—and this brings us to the third impediment to receiving this insight, and which must be cleared away—McDowell's early misreading of Wittgenstein has served to obscure, amongst commentators, his later grasp of Wittgenstein's insights. At the beginning of Chapter 3, we will observe that McDowell's early reading of Wittgenstein remains captured by a form of the same hopeless dilemma: either putatively linguistic behaviour is viewed as merely brute dispositions to react, and is thereby inevitably threatened by a regress of interpretations; or meaning can only be conceived as superlative, Platonistic fantasy. Not unlike Wright, (early) McDowell fails to reflect deeply on Wittgenstein's misunderstanding—'an inclination to say: any action according to the rule is an interpretation' (*PI* §201)—and, instead, moves immediately to import substance into the obverse of the misunderstanding—'there is a way of grasping a rule which is *not* an *interpretation*' (*PI* §201)—in the hope that it will *somehow* open up a "third way" between the two horns of the familiar dilemma. Yet this proposed "third way" (an appeal to communal institutions, or culture as that which will *somehow* bridge the perceived gap between the expression of a rule and appropriate performance) still retains an aura of mystery, and therefore seemingly necessitates further explanatory research. Thus McDowell's early reflections on the issues are insufficiently distanced from Kripke's and Wright's respective readings, and, like theirs, his picture of meaning and of rule-following remains haunted by vertiginous terror at the sheer contingency of agreement in individual linguistic reactions. Below "bedrock", according to (early) McDowell, 'there is nothing but contingency' and 'the ground seems to have been removed from our feet' (McDowell 1998a: 251; 1998e: 207).

Despite attracting justified criticism, what has gone unremarked within the literature is the fact that McDowell comes later to explicitly repudiate his early reading as based on a form of the very misunderstanding Wittgenstein is trying to warn us against. McDowell's later reading begins, in contrast to the earlier, not by immediately looking for the substance of the obverse of the misunderstanding, but with greater reflection on the nature and implications of the misunderstanding itself. The fruit of this greater reflection

reveals what I will call “the three-step insight” into the misunderstanding of *PI* §201. And this is that, in an attempt to explain the language-game from “outside”—that is, without presupposing meaning and understanding—we are led to conceive the expression of a rule as merely “dead signs” and behaviour made in response to the rule as merely brute (Step 1). Once this step has been made then it seems that appropriate behaviour can be connected with the expression of the rule *only under an interpretation* (Step 2). But, given our initial steps, then anything we care to bring forth as a candidate mediating interpretation, and which is supposed to bridge the gulf between the expression of a rule (merely “dead signs”), on the one hand, and (merely brute) behaviour which is in accord with that rule, on the other, will almost inevitably be conceived in such a way that it, too, requires interpretation. And so this is why we come to face the familiar interpretational regress (Step 3).

As we shall consider in Chapter 3, there are two principal lessons of Wittgenstein’s regress of interpretations argument. The first is that we must not allow any such conceptual gap to open up between the expression of a rule and behaviour which accords or not with the rule. Once we allow such a gap to open ‘it will,’ McDowell warns, ‘be unbridgeable’ (McDowell 2009a: 101). The second principal lesson of Wittgenstein’s arguments brings us eventually to the obverse of the misunderstanding: ‘there is a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call “obeying the rule” and “going against it” in actual cases’ (*PI* §201). At *On Certainty* §28, Wittgenstein remarks: ‘What is “learning a rule”—*This*. What is “making a mistake in applying it”?—*This*.’ For those who are inductees into a particular normative practice, in response to the expression of a rule, they can act immediately on their *understanding* of that rule. Of course, there can be misunderstanding, and there can be failure to understand. But such notions are clearly parasitic on the concept of *understanding*. A deviant rule is just that—*deviant*. It is not “the same rule under a different interpretation”, but, at best, a different rule, or a different concept.

But such direct appeals to unmediated understanding of a word or rule are not problematic for Wittgenstein, for their purpose is not connected with any explanatory ambitions.

Rather, their function in Wittgenstein's dialectic are as "reminders" (*PI* §127). And the purpose of such reminders is precisely to ensure that we do not take the fateful first step (propelled by our explanatory ambitions) into a picture of normative inertness. Once behaviour is viewed as merely brute, and expressions of rules as merely "dead signs", then eventual regress is inevitable. And any attempt to squeeze normativity back into the picture inevitably deforms such normativity, such that the demandingness of rules ("the hardness of the logical must": *PI* §437) can only show up as extraordinarily queer, spooky, and mysterious—as ultra-objective superlative rules (the rule as a "despotic demand": *PI* §437).

Consequently, and as (later) McDowell comes to recognise, Wittgenstein's appeal to *custom, forms of life*, and so forth, are *not*, as they are often conceived amongst commentators—including (early) McDowell—programmatic gestures towards a naturalistic *explanation* of our linguistic and mental lives. On the contrary, Wittgenstein asserts that: '*Our mistake is to look for an explanation...*' (*PI* §654). As I hope will become clear as we proceed, trying to step "outside" the language-game in order to explain it merely leads us to become 'entangled in our rules' (*PI* §125). So: 'The question is not one of explaining a language-game...but of noting a language-game' (*PI* §655). And the function of the reminders of instances of unmediated understanding are, then, for 'clearing misunderstandings away' (*PI* §125). To think that we require further constructive philosophical explanations is still to be in the grip of the misunderstanding. As, again, I aim to make clear, what we require, according to Wittgenstein, is to pay close attention to 'the spatial and temporal phenomenon of language'—unmediated understanding of a word or rule (by those, at any rate, who have been successfully inducted into linguistic and rule-governed practices) together with the circumstances which surround such norm-governed behaviour (*PI* §108).

Once misunderstandings have been cleared away, then our linguistic practices need not show up as deeply problematic, such that they really do then seem to require explanation. Rather, we are then free to contemplate our linguistic life as it really is, free from the deformations which our explanatory ambitions impose on such ordinary, everyday

phenomena—ordinary and everyday, at least, for extraordinary linguistic animals, such as we ourselves are.

Chapter 1 Kripke on Wittgenstein's Rule-following Considerations

But how can a rule shew me what I have to do at this point? Whatever I do is, on some interpretation, in accord with the rule.

Philosophical Investigations §198

Introduction

The import of Wittgenstein's investigation of rules and of rule-following, as well as its relation to the so-called "private language argument" (which forms the principal focus of the remarks from §243 of the *Investigations* up to about the end of the 420s) is a contested matter. What is certain, however, is that it is nigh on impossible to discuss the significance of Wittgenstein's discussion of rules and of private language without giving due consideration to Saul Kripke's celebrated, and hugely influential, treatment of these Wittgensteinian themes. As is by now notorious, Kripke unearths a deep "sceptical paradox" at the heart of Wittgenstein's remarks, one which seemingly renders incoherent the very possibility that we can mean anything determinate by our words. As I hope will become clear throughout this and later chapters, the threat of paradox is undeniably crucial to a proper understanding of the central problems which lie at the heart of the *Philosophical Investigations*, and to the discussion of rules, and of rule-following, in particular. Yet, as I will argue in Chapter 3, Wittgenstein's paradoxical chain of reasoning is meant as a *reductio ad absurdum* of certain philosophical (mis-)conceptions of meaning and of understanding, and is not, as Kripke contends, meant to impugn the very possibility of meaning and of understanding. In fact, as we shall see in Chapter 3, Kripke's discussion embodies a form of the very misunderstanding Wittgenstein is trying to warn us against.

Unfortunately, the term “*sceptical paradox*” is misleading inasmuch as it might lead one to think that the problem raised by Kripke is primarily epistemological, when the American logician goes to great lengths, as Paul Boghossian has pointed out, to insist that the paradox is a constitutive problem: the fundamental issue is, he rightly states, ‘the possibility of correctness’ (Boghossian 1989: 517). In other words, the issue is not that of classical epistemological scepticism: how can we *know*, say, whether what the sense of an assertion determines is true or false, but, rather, the *constitutive* question of how an utterance can so much as have a determinate sense, or of how a word can be used according to a determinate meaning—the issue concerns, that is to say, the very *possibility* of thought or deed undertaken according to an understanding of conditions of correctness. As Wittgenstein says at PI§90, ‘our investigation...is directed not towards phenomena, but, as one might say, towards the “*possibilities*” of phenomena’. This is a point recognised by Kripke when he remarks that the question considered by Wittgenstein is ‘How can we show *any language* at all...to be possible?’ (Kripke 1982: 62).

Leaving the proviso over the appellation of “scepticism” to the paradox aside for the time being, Kripke confesses an inclination to regard the paradox he unearthed from Wittgenstein’s text ‘as the most radical and original sceptical problem that philosophy has seen to date’ (ibid: 60). The seemingly unavoidable implications of the paradox appear to show that ‘*all language, all concept formation, to be impossible, indeed unintelligible*’ (ibid: 62). Whatever judgement posterity might bestow upon the merits of Wittgenstein’s own solution to ‘this new form of philosophical scepticism’, Kripke is clear that discovery of the problem on its own constituted ‘an important contribution to philosophy’ (ibid: 7). If his book has a ‘main thesis,’ Kripke confides, it is ‘that Wittgenstein’s sceptical problem and argument are important, deserving of serious attention’ from amongst philosophers (ibid: ix).

Certainly, Kripke’s own justly celebrated presentation of the problem, published in 1982 as *Wittgenstein on Rules and Private Language*, brought to fulfilment his hope that the problem should receive serious attention from amongst academic philosophers and has, for better or ill, become something of a *locus classicus* for commentary on Wittgenstein

on rules and rule-following. Kripke's book has had a noticeable impact within mainstream Anglo-American philosophy over the past few decades, an influence whose reverberations have spread beyond the narrow confines of Wittgensteinian scholarship. This influence was not only due to 'the accessibility of Kripke's style,' and partly to his 'personal reputation,' Michael Morris considers, but also because his work 'presented a striking challenge to everyone who believed that words really mean something', and thereby 'provoked a minor industry of work designed to avoid the scepticism which it proposed' (Morris 2007: 271).

In this chapter, I will argue that the influence of Kripke's reading of Wittgenstein over recent decades is a double-edged sword. Kripke's book casts both a blessing and a curse over a proper understanding of the import of Wittgenstein's arguments. Amongst the blessings of Kripke's reading is, as I hope to show, that it does serve to highlight two very crucial features of Wittgenstein's discussion. But, as I will also make clear in this chapter, Kripke's influential treatment casts a curse on those investigations by obscuring, at the very same time it seeks to unveil, those very same two features —both of which are indeed integral to an understanding of the import of Wittgenstein's arguments. The principal aim of this chapter, therefore, is to clear away the misunderstandings in Kripke's text which serve to obscure the two essential features of Wittgenstein's investigations.

The first feature I aim to bring into sharper relief is the decidedly *normative*, or constitutive nature of the issues given consideration in the remarks on rule-following. That is to say, Wittgenstein reflects on behaviour undertaken in accordance with a standard, or standards, of correctness —for instance, using a word according to its meaning, or of responding to the expression of a rule in accordance with the significance of that rule. The second important feature which requires consideration is a form of *reductio ad absurdum* argument, repeated, in a variety of contexts, throughout the text of the *Investigations*, which, at least on the face of it, seems to render the very notion of behaviour undertaken according to a standard of correctness, or of the possibility of understanding a word according to a determinate meaning, as incoherent. We will turn to consider the true import of Wittgenstein's discussions, and the actual dialectical role

played by this variety of *reductio* argument in Chapter 3 but, in this chapter, I will only suggest that Kripke's claim that Wittgenstein accepts any "sceptical paradox" is misguided. As almost all commentators have pointed out, Wittgenstein rejects the paradoxical chain of reasoning as based on a misunderstanding. Of course, this still leaves open the question of what such a rejection amounts to but, for the purposes of the present chapter, it will be enough to make clear that, *pace* Kripke's reading, Wittgenstein does reject the paradoxical reasoning as based on a misunderstanding.

This present chapter is divided into four sections. In the first section, I aim to bring out the *normative*, or *constitutive* nature of the problems discussed by Wittgenstein and Kripke. The first thing I hope to make clear is that Wittgenstein's invocation of the notion of manifesting understanding, in one's speech or behaviour, of the meaning or significance of a word or rule, is not meant to be freighted with any particular load over and above the basic fact that acting upon an understanding of a rule necessitates operating according to a standard, or condition, of correctness. The following declaration of Simon Blackburn, on the normative nature of his investigations, could thus apply equally well to Wittgenstein's reflections:

I intend no particular theoretical implications by talking of rules here. The topic is that there is such a thing as the correct and incorrect application of a term, and to say that there is such a thing is no more than to say that there is truth and falsity. I shall talk indifferently of their being correctness and incorrectness, of words being rule-governed, and of their obeying principles of application. Whatever this is, it is the fact that distinguishes the production of a term from mere noise, and turns utterance into assertion – into the making of judgement (Blackburn 2002: 28-29).

As Paul Boghossian has argued, Kripke is keen to highlight the normative nature of the problem which he finds in Wittgenstein. Nevertheless, as Boghossian justly points out, Kripke confuses the issue somewhat by choosing 'to present the constitutive question in

an epistemological guise' (Boghossian 1989: 515). Whilst I agree with this assessment, in the first section of this chapter, I will argue nonetheless that Boghossian, in his criticisms of Colin McGinn, misleadingly underplays, or ignores, some of the ways that Kripke's treatment serves to obscure the constitutive, or normative nature of the issues. Specifically, as I will show, Kripke confuses matters by almost continually assuming that the source of the justification for my present use of a word according to a determinate meaning must be found in my *previous intentions* regarding the meaning of the word. Thus, a principal aim of the first section is to explicate Kripke's influential discussion, but in a manner which clears away some of these confusions, in order that the normative nature of the problem - which Kripke anyway, at other moments, wishes to highlight - may stand in greater relief.

Kripke claims that the "sceptical paradox" is devastating for dispositional theories of meaning because, given the unavoidably normative features of the meaning of words, any such theory must adequately account for the correctness-conditions involved in using a word according to a determinate meaning (using "+" or "plus" to denote the addition function, for instance). But this is what dispositional theories are unable to do for, as Kripke rightly argues, specifications of dispositions to use a word in a particular way can only ever be *descriptive*, not normative. Dispositions specify only what I will or would do, not what I *ought* to do. We will consider this objection in the second section of the present chapter where, together with the aid of Boghossian's further arguments, we will conclude that dispositional theories are indeed inadequate to the task of accounting for the normativity of meaning. This is because, as Boghossian argues, dispositional theories are unable to satisfy, simultaneously, the two requirements which any such theory of meaning must satisfy: on the one hand, and in the face of a seeming infinity of possible dispositions which can be brought into relation with one's use of a word, the theory must specify only those dispositions which are meaning-determining; on the other hand, and on pain of vicious circularity, that specification must be couched in non-semantic and non-intentional terms. But, as we shall see, dispositional theories are unable to meet these requirements simultaneously: they either end up pre-supposing the meaning which the theory was supposed to explain; or they fail to preserve the possibility of determinate meaning in face of the constitutive (or, alternatively, normative) paradox.

Arguably the most striking feature of Kripke's book is the way it serves to bring out very vividly the devastating and absurd consequences – all language is meaningless! – of the paradoxical train of reasoning to be found in Wittgenstein's text. Equally as notorious, however, is Kripke's claim that Wittgenstein accepts the implications of the paradox. Given the fact that the problem is constitutive —that it concerns the very possibility of behaviour manifesting understanding of correctness-conditions— then, as we will see in the third section, the paradoxical reasoning can be repeated on behalf of *any activity at all* which involves manifesting understanding of a condition of correctness. Both Wittgenstein (implicitly) and Kripke (explicitly) acknowledge this point. Thus, the examples discussed by Wittgenstein include, amongst others: using words with a determinate meaning; reciting the alphabet; developing an arithmetical series in response to the expression of a rule (say, "Add 2"); and following a sign-post. Furthermore, Wittgenstein acknowledges that the paradoxical reasoning seemingly threatens the very possibility of formulating a thought, or of making an assertion with a determinate meaning; not to mention appearing to threaten the very possibility of a belief, desire, wish, or command —or any intentional state whatsoever— with a definite content. For any 'mental item', according to Kripke's "sceptical paradox", can be 'interpreted so as to accord with anything we may choose to do' (Kripke 1982: 55). Thus, Wittgenstein's arguments have apparently shown, not only '*all* language [and] *all* concept formation to be impossible, indeed unintelligible' (ibid: 62); but also the same goes for the possibility of understanding *all* rules, signs and symbols, *all* mental representations, *all* intentional thought and action — in short, *all* behaviour undertaken according to conditions of correctness! Thus, the implication of Kripke's formulation of the conclusion of Wittgenstein's argument —that the "sceptical paradox" apparently renders all and any behaviour undertaken according to a standard of correctness as incoherent— is, quite simply, that "there can be neither accord nor conflict" (ibid: 55).

As I have already stated, far from accepting, as Kripke claims, such evidently absurd implications (whatever "acceptance" could possibly amount to here), Wittgenstein rejects the paradoxical reasoning as based on a misunderstanding. In Chapter 3, we will investigate the nature of this rejection in closer detail, but, in the fourth and final section of the present chapter, I will show that, despite Kripke's claims of acceptance of the

paradoxical reasoning on Wittgenstein's behalf, Kripke's reading of the "private language argument" merely serves to distract attention from the evidently absurd and self-refuting implications of the paradox, which Kripke brings out with such force elsewhere in his book. As we shall see in the fourth section, Wittgenstein's "private language argument", at least as Kripke reads it, enacts a subtle shift from acceptance of the "sceptical paradox" —all language is meaningless!— to the "sceptical conclusion" that talk of determinate meaning is only incoherent *when we consider an individual in isolation*. On this construal, the sceptical conclusion entails that ascriptions of meaning to speakers, such as "Jones means addition by '+'", lack factual content. But this does not entail, according to the "sceptical solution" attributed to Wittgenstein by Kripke, that the language game of ascribing determinate meaning to the utterances of another ("Jones means addition by '+'") lack a role or function in our lives. Thus, although we are supposed to accept that ascriptions of determinate meaning to another's utterances lack factual content (the "sceptical conclusion"), nevertheless, it is supposed, this does not entail that ascriptions of meaning lack "conditions of assertability". According to the "sceptical solution", such assertability conditions involve reference to the wider linguistic crowd whose communal dispositions somehow stand judgement upon the linguistic dispositions of individual speakers.

Only the most charitable commentator, I think, could find much textual evidence for Kripke's—admittedly ingenious—Communitarian construal of meaning in the *Investigations*. And, as I will argue in Chapter 2, Kripke's "community appeal" fails to save his Wittgenstein from the devastating implications of the paradox. But this is to get too ahead of ourselves at this time. Before beginning to make good on the promises made in the introduction to this chapter, however, it is sufficient to note here that, as shall be seen in the fourth section, Crispin Wright reads Wittgenstein as making essentially the same move as Kripke has him make (and this, intriguingly, despite both having developed their respective readings independently—if more or less concurrently). In face of the apparent paradox, Wright concludes on Wittgenstein's behalf that the idea of *unilateral* recognition of a ratification-independent standard of correctness should be rejected as incoherent. Wright, as we shall see, draws this conclusion, as does Kripke, in the face of remarks such as *PI*§185, where Wittgenstein raises the possibility of a "deviant rule

follower". However, once the paradox is read as a rejection of the possibility of manifesting understanding of a condition of correctness, but *only when an individual is considered in isolation*, this kindles the hope, as in Kripke's reading, that Wright's Wittgenstein can make a retrieval of normative notions through appeal to the wider linguistic community. Wright eventually abandons this earlier, crude Communitarianism in favour of a more sophisticated Constructivist response to the paradox, but, as I will argue in Chapter 2, neither Kripke's non-constitutive Communitarianism, nor (updated) Wright's Constructivism enable us to avoid the devastating implications of the paradox.

§1 The Constitutive Nature of Wittgenstein's Discussion of Rules

To say that linguistic meaning is normative is to say no more, but no less, than that, in applying words and other linguistic expressions, we operate according to standards of correct use. Of any use of a word, it can legitimately be asked whether that word has been used according to its meaning. It is a criterion of saying of someone that they understand the meaning of a word that they are able to employ that word according to certain standards of correctness, such that substantial deviation from such standards means deviating from the meaning of the word. This basic normative aspect of the meaning of words is clearly foregrounded in the remarks which initiate Wittgenstein's investigation of rules and of rule-following. At §138 of the *Investigations*, Wittgenstein draws attention to the normativity of language: that the meaning of a word — the example used is the word “cube” — necessitates a (more or less) determinate *use* of the word. The meaning of a word fixes certain normative obligations on the correct application of that word. How we respond to an utterance of the word, or the behavioural applications we make of the word in particular circumstances, are to be judged, in part, according to that word's meaning (are to be judged, in other words, according to correctness-conditions).

At §§143-155, Wittgenstein turns the focus of his investigation from use of the word “cube” onto the nature of behaviour which manifests understanding of the expression of a rule — the example discussed is the rule for the extension of an arithmetical series. Nevertheless, the focus of the discussion remains the *normative* features of a rule – the fact that there are correct and incorrect responses to the expression of a rule. Indeed, the reason Wittgenstein uses a mathematical example, as does Kripke, is that it better serves to throw this normative aspect into high relief, given that action which manifests understanding of the meaning, or significance of the addition function, for instance, determines a completely precise response at any particular stage of its application: a sum ‘yields one precise value for each pair of numerical arguments’ (Kripke 1982: 82); or, as in the case of the development of an arithmetical series, having reached “...998, 1000...” in the development of the series “Add 2”, only writing “...1002, 1004, 1006...” will conform with the rule.

The focus of Wittgenstein's discussion continues to alter at various points in the text: thus at §149, consideration is given to the ability to recite the alphabet, and, from §§156-178, Wittgenstein turns his attention to "reading" (in the sense of bare translation of written symbols into phonemes). That which unites each of these disparate activities is that they are all normative, in the sense that they each involve behaviour undertaken according to a standard of correctness: there is a correct way to recite the alphabet, and right verbal response to written text; there is as a way a word ought to be used or understood, and a way one ought to react to the expression of a rule in a way that manifests understanding of the meaning or significance of the rule.

Intimately tied to the issue of the normativity of meaning is that making an assertion or judgement involves a normative relation with the facts that would make the assertion true or false. As Michael Luntley puts the point:

Our meaning things with our words requires that there be a right / seems right distinction for our use, else our use is not a meaningful use.

So, if I utter the words

(1) Grass is blue

with their conventional meaning I am *obliged*, on being presented with a grass sample to withdraw my utterance. For in uttering (1) I am bound by the meaning of the utterance to acknowledge that there are conditions which would render the utterance correct and conditions that would render it incorrect. That is just what meaning something with our words is like.

We take on obligations (from Thornton 1998: 34).

Such normative obligations extend wider than assertoric utterances alone, however, and characterise intentional mental states more generally: for instance, one's intention, wish, or hope to go on holiday is fulfilled only in the event that one *goes on holiday*. If one

does not go on holiday then one's intention, wish, or hope remains unsatisfied. As Tim Thornton puts the point:

There is a normative connection between a state and what the state is about or for...Hunger for an apple is not hunger for a punch in the stomach nor for a pear even if the first might stop one feeling hungry and the latter ease one's craving just as well. The connection between a state and what satisfies it is normative in the way that the connection between a rule and an application of it is normative. Rules and intentional mental states *determine* what accords with them (Thornton 1998: 33).

A principal purpose of this chapter, then, is to suggest that the point of Wittgenstein's focus on following according to a rule can helpfully be understood as a means of reflection upon this more general phenomenon — characteristic of human beings, in both deed and speech — of operating according to standards, or conditions of correctness. Or, to repeat Wittgenstein's remark, his investigations are 'directed...towards the "*possibilities*" of phenomena' (PI§90).

Wittgenstein raises the issue of the normativity of meaning by introducing an apparent problem about the fact that assertions and other intentional states are used according to a standard, or standards, of correctness. At PI§437, the following sense of "queerness" is expressed:

A wish seems already to know what will or would satisfy it; a proposition, a thought, what makes it true – even when that thing is not there at all!
Whence this *determining* of what is not yet there? This despotic demand?
("The hardness of the logical must.")

At §138, the remark which begins the "rule following considerations", expression is given to a similar puzzlement over the normative nature of words. As we have seen, it is an undeniable fact that a criterion of one's being deemed to have grasped the meaning of a word is that one is able to apply that word according to certain standards of correct

application, such that deviation from those standards entails deviation from the meaning of the word. Yet it is an equally undeniable fact that we often manifest immediate recognition of a word's meaning: 'we *understand* the meaning of a word when we hear or say it; we grasp it in a flash' (PI§138). When I hear someone make an utterance involving the word "cube", for example, I most often know – unreflectively and without hesitation – what is meant. But how do the normative obligations on one's linguistic repertoire — the applications we make of the word: past, present, or future — relate to the fact that we, for the most part, grasp the word's meaning "in a flash"? Surely, the voice in Wittgenstein's text ponders, 'what we grasp in this way is surely something different from the "use" which is extended in time!' (ibid.). At PI§139, this puzzlement is carried over into a series of anguished questions about the nature of any such relation:

But can the whole *use* of the word come before my mind, when I *understand* it in this way?...And can't these ways of determining meaning conflict? Can what we grasp *in a flash* accord with a use, fit or fail to fit it? And how can what is present to us in an instant, what comes before our mind in an instant, fit a *use*?

Notoriously, on Kripke's reading, Wittgenstein answers, in effect, that there cannot be any connection at all between the manifestation of immediate recognition of a word, on the one hand, and a determinate pattern of its use, extended over time, on the other. This is the "sceptical conclusion" famously attributed to Wittgenstein by Kripke: that 'there is no fact about me that distinguishes between my meaning a definite function by "+" ...and my meaning nothing at all' (Kripke 1982: 21). This is because, as will be discussed further in a moment, a succession of arguments of broadly Wittgensteinian provenance apparently show that 'no matter what is in my mind at a given time, I am free in the future to interpret it in different ways' (ibid: 107). As we shall soon discuss, this leads Kripke's Wittgenstein into the 'incredible and self-refuting conclusion that all language is meaningless' —that there 'can be no such thing as meaning anything by any word' (ibid: 71, 55). And such 'insane and intolerable' repercussions extend over intentional states and representations more generally: there is no such thing as a unique determination of a mental representation: for instance, 'any present intention could be

interpreted so as to accord with anything we may choose to do' (ibid: 60, 55). The apparent upshot of Wittgenstein's reflections, according to Kripke, is that 'we follow a rule as we do without reason or justification' (ibid: viii). In short, the very notion of operating according to a standard of correctness is, supposedly, incoherent: 'there can be,' Kripke concludes, 'neither accord, nor conflict' (ibid: 55). In the third section of the present chapter, I will urge that Kripke is mistaken, and that Wittgenstein does not accept any such paradoxical and self-refuting claims (whatever "accepting" such "claims" could even amount to). First, however, we need to turn to the considerations which lead Kripke to such a surprising, and deeply problematic, conclusion.

Anybody possessed with the most rudimentary grasp of mathematics, Kripke begins, is capable of using the word "plus" and the symbol "+" to denote the mathematical function, addition. Kripke highlights the fact that one is trained into competent use of such words, symbols, and functions by virtue of only a limited number of examples. Moreover, since being trained into competence, one has calculated only a limited number of sums in one's life thus far. These facts are dramatized by Kripke's sceptic's invitation for us to suppose that our arithmetical experience thus far has been confined to operating with integers of a value less than 57. Despite these facts, however, Kripke is keen to highlight that 'one point is crucial' to one's coming to grasp the rule for addition: 'the rule determines my answer for indefinitely many new sums that I have never previously considered' (ibid: 7). This rightly serves to bring out the undeniable normative aspect of the rule for addition: the fact that understanding the meaning of the rule commits one to determinate applications when applying the rule in various situations — commits one, in general terms, to act in accordance with a standard, or standards, of correctness. Thus Kripke acknowledges 'an obvious Kantian flavor' to Wittgenstein's problem (ibid: 62, fn. 48).

In the very next sentence, however, Kripke abandons this crucial insight, and replaces it with another; one that, moreover, contradicts the one crucial point! He goes on to say: 'the whole point of the notion that in learning to add I grasp a rule' is that 'my past intuitions regarding addition determine a unique answer for indefinitely many new cases in the future' (ibid: 8). So, it is not now the addition function which determines all future

applications; rather, it is my past intuitions, or perhaps we could say past intentions, regarding this rule which determines the steps to be taken. Yet these should not be confused. No doubt, if I understand what the rule for addition demands of me, such that my past intuitions of the rule have conformed to these demands, then there is no essential divide between my past intuitions concerning the rule and the demands required by that rule, such that conforming my present behaviour to my past intuitions will, by proxy, also entail that my present applications of the rule are in conformity with its demands. If, however, at any time in the past I have misunderstood, or have failed to understand, the rule's normative demands, then, during such moments, my past intuitions regarding the rule would not, by definition, have been in conformity with the – undeniably 'crucial' – demands of the rule for addition. And so, if I intend to manifest understanding of the meaning of the addition rule during my present application of the word "plus", then I need to make sure that that my present behaviour conforms to the normative demands of the rule — and precisely *avoid* conforming my behaviour to my previous (and *ex hypothesi* mistaken) intuitions regarding the rule.

Aside from such considerations, however, one's past intuitions or intentions regarding the rule for addition are beside the point in relation to the question of whether the steps one takes in applying a rule, such as "Add 2", are in conformity with the normative dictates of the rule. Moreover, as I hope will be clear by the third section of the present chapter, Kripke later admits that any particular considerations about past intentions in relation to a rule are irrelevant to Wittgenstein's problem, as 'it is completely general and can be applied to any rule or word' (ibid: 58). The point about past intentions, then, is a red herring — the one crucial point, as Kripke initially acknowledged, is an apparent problem about the normative dictates of a rule, or of action manifesting understanding of a condition of correctness (this is the "Kantian flavor" Kripke detects). In what follows, I will try to keep the focus on the central issue of normativity: of how there can be rules for the correct use of a word, or of how there can be criteria for the correct application of a rule — which is anyway the important issue for Wittgenstein — and not the particular, and comparatively marginal, issue of whether my present employment of a word conforms to my past intentions so to employ it.

Assuming one's competence in the use of the addition function, upon being confronted with a sum, even one never encountered before, one will usually be confident in providing the answer. Even assuming that one's arithmetical experience does not thus far extend to adding numbers higher than 57, upon being asked to compute "68+57", one will no doubt confidently —"in a flash"— provide "125" as the correct answer. Kripke next invites us to imagine a sceptic who aims to cast doubt upon this certainty. The sceptic does not doubt that I mean the addition function by my present use of the word "plus" — at least not at this stage. Rather, the sceptic invites me to explain how I know that, when I used the word "plus", or the symbol "+" in the past, they were used to denote, not the plus function, but the *quus* function, defined as yielding the same values as addition with all sums involving pairs of integers less than 57, but otherwise yielding 5 as value. Under the supposition that I have never previously tackled sums involving numbers beyond 57, then all my previous applications of the word "plus" provide evidence, it is supposed, both for the hypothesis that, during my previous applications, I was operating with the addition function, and the hypothesis that I was, in actual fact, operating according to the *quaddition* function. Given the supposition that my being presented with a sum involving an integer with a value higher than 57 is an arithmetical novelty for me, then my previous use of "+" and of "plus", on the evidence of my previous linguistic behaviour—considered as mere behaviour—leave it undetermined which of the competing hypotheses should be favoured. So, what justifies my certainty that I was, in the past, adding and not quadding?

Furthermore, falsely assuming that the answer I give to an addition problem in the present must conform, not to the rule for addition, but to my previous intuitions regarding that rule, the sceptic now challenges me to justify my present confidence that the answer I *ought* now to give is "5", not "125". The sceptic is proposing, then, that I 'have made a mistake, not in mathematics, but in the supposition that I had accorded with my previous linguistic intentions' (ibid: 9). Under the supposition that the meaning of my present use of "+", or of "plus" must conform to my previous intentions, then the sceptic is doubting that I am able to justify the correctness of "125" as my present response.

We will recall that Boghossian wishes to highlight the fact that ‘the crucial issue for Kripke’ is the normativity of meaning (Boghossian 1989: 511). And, as I have suggested, this is the crucial issue for Wittgenstein too². Whilst I wholeheartedly agree that normativity does form the core of the discussion, Boghossian nevertheless plays down the extent to which Kripke’s own treatment of the problem does serve, unfortunately, to obscure this crucial feature. That Boghossian ignores this characteristic of Kripke’s presentation of the problem is shown in his criticisms of Colin McGinn, whom he accuses of confusing the conception of normativity that is in question in Kripke’s text (ibid: 511-513). For McGinn: ‘The notion of normativeness Kripke wants captured is a transtemporal notion...(Kripkean) normativeness is a matter of meaning now what one meant earlier’ (McGinn quoted at ibid: 512). This Kripkean notion of normativeness, McGinn explains, amounts to the claim that my present use of a word is deemed ‘*correct* in the light of my earlier employment of that word’, and with the concomitant notion of ‘linguistic incorrectness’ as consisting in ‘using the same word with a different meaning from that originally intended (and doing so in ignorance of the change)’ (quoted in ibid: 511). McGinn rightly points out that this is not a picture of using the same concept incorrectly, but amounts to using a *different concept* (using the quaddition function, say, rather than an incorrect use of addition)³. Boghossian agrees with McGinn about this point, but complains that ‘the “normativity” requirement defined by McGinn has nothing to do with the concept of meaning *per se* and is not the requirement that Kripke is operating with’ (ibid: 512).

I agree completely with Boghossian about this conception of “ transtemporal normativity” having nothing directly to do with the meaning of a word or rule; nevertheless, he overlooks the fact that, to be fair to McGinn, the issue over normativity in Kripke’s text is fudged: for whilst Kripke does indeed state that the crucial issue concerns the question

² Boghossian, like Blackburn, recognises also that the focus on the normativity of language is not, as I earlier suggested in relation to Wittgenstein, meant to be freighted with any philosophical load, nor is it meant to imply any substantial commitments to any formal theory of meaning:

The normativity of meaning...[is] simply a new name for the familiar fact that, regardless of whether one thinks of meaning in truth-theoretic or assertion-theoretic terms, meaningful expressions possess conditions of *correct* use. (On the one construal, correctness consists in *true* use, on the other, in *warranted* use). (Boghossian 1989: 513).

³ We will return to discuss this crucial point in Chapter 3, §2.

over how the meaning of a word or rule can constitute a standard of correctness — the conception of normativity which Boghossian rightly wishes to keep focussed on: ‘the rule determines my answer’. Nevertheless, as we have seen, in the very next sentence, Kripke does then go on to say that ‘the whole point of the notion that in learning to add I grasp a rule’ is that ‘my past intuitions regarding addition determine a unique answer for indefinitely many new cases in the future’ (Kripke 1982: 7, 8) — the conception dismissed by Boghossian as ‘McGinn’s understanding of normativity’ (Boghossian 1989: 512). Moreover, Kripke continues to couch the problem in such “transtemporal” terms throughout his text. So, whilst I agree wholeheartedly with Boghossian that what he calls “McGinn’s understanding of normativity” is not to the point — the real issue is how a word or a rule can constitute a standard of correctness — it must be conceded to McGinn that the “transtemporal notion of normativity” is prominent throughout Kripke’s book, so isn’t something that McGinn has smuggled into the debate surreptitiously.

Setting aside the confusion that my present use of the word “plus”, in order to be deemed correct, must accord with my previous linguistic use of the word (*regardless* of whether I was then using it correctly – that is to say, to denote addition), then the challenge the sceptic confronts us with is to produce some fact about me which constitutes my having meant addition and not quaddition during my previous applications of the word. The crucial point is that, despite the talk of previous intentions, the issue is not the epistemological one of how I can presently *know* what my previous linguistic intentions were, but the *constitutive* question over how I could even be said to have been, during my previous applications of the word “plus”, in a state of meaning one thing rather than another (addition rather than quaddition). Thus, our answer to the sceptic, Kripke stipulates, ‘must give an account of what fact it is (about my mental state) that *constitutes* my meaning plus, not quus’ (Kripke 1982: 11; emphasis added). But this is precisely what the sceptic is suggesting we cannot provide:

Now if the sceptic proposes his hypothesis sincerely, he is crazy; such a bizarre hypothesis as the proposal that I always meant quus is absolutely wild. Wild it indubitably is, no doubt it is false; but if it is false, there must be some fact

about my past usage that can be cited to refute it. For although the hypothesis is wild, it does not seem to be *a priori* impossible (ibid: 9)⁴.

At this stage, Kripke continues, someone may wish to protest that, although I have only thus far encountered a limited number of addition problems, nevertheless, through my – admittedly finite – training in the use of the word “plus” and its cognate words, I have ‘internalized instructions for a *rule* which determines how addition is to be continued’ (ibid: 15). It is surely the fact that I am following *this* internalised rule, and not any other, that determines my behaviour. For example, Kripke refers us to an explicit and repeatable counting rule ‘in its most primitive form’: when called upon to add x and y , I count out x marbles into a heap, and count y marbles into another; I then combine the two heaps and count out the total number of marbles which form the new unified heap. The result of this procedure determines $x+y$. The actual imbibed rule, or algorithm internalised through learning, Kripke admits, maybe ‘more sophisticated and practically applicable than the primitive one just described’ (ibid: 16). Yet however primitive or complex we imagine the ingested algorithm, the important point is that:

It is engraved on my mind as on a slate. It is incompatible with the hypothesis that I meant quus. It is this set of directions, not the finite list of particular additions I performed in the past, that justifies and determines my present response (ibid: 15-16).

But now, according to Kripke, it is open to the sceptic to provide the following response: it may be true that “count”, as I used the word in the past, referred to the act of counting. But I applied the word “count”, as I applied the word “plus”, in only a finite number of cases. Thus, my present interpretation of my past application of “count” can be brought into doubt, as was the case with my present interpretation of my past application of “plus”. Perhaps, the sceptic mischievously suggests, that by “count” I

⁴ In Chapter 3 I will argue that conceiving the challenge of “the sceptic” as a “false hypothesis” betrays a deep misunderstanding of the nature of the constitutive paradox—which embodies a *conceptual confusion*, not a false hypothesis.

previously meant *quount*, defined as the same rule as counting, except in cases where one of the sub-heaps contains more than 57 items, in which case one must give the answer “5”.

Kripke’s sceptic deploys a similar succession of arguments, of broadly Wittgensteinian provenance, and concludes that ‘nothing in my mental history establishes whether I meant plus or quus’ because ‘everything in my mental history is compatible both with the conclusion that I meant plus and with the conclusion that I meant quus’ (ibid: 21).

Kripke’s sceptic claims to allow us *carte blanche* over the range of facts which can be appealed to in answer to the question over what determines my meaning one thing or another — from *both* the domain of observable “external” behaviour *and* from the domain of “internal” mental events (ibid: 14-5; 55-8). Nevertheless, nothing from the finite range of behavioural and mental facts will serve the purpose, the sceptic argues, because *ex hypothesi* I have never previously engaged with sums involving numbers larger than 57; and, since both “plus” and “quus” yield the exact same values for sums involving figures no larger than 57, then all facts about myself – both behavioural and mental – can be interpreted as providing evidence for *either* my having meant addition *or* my having meant quaddition during my previous performances. Either interpretation, it seems, conforms to the available evidence. And so now we appear to be in the position that any behavioural fact, and any mental fact we can appeal to, fails to constitute a determinate meaning: ‘nothing in my mental history or past behaviour – not even an omniscient God would know – could establish whether I meant plus or quus’ (ibid: 21). Furthermore, it won’t do to cite a further mental item that serves to fix one interpretation rather than another - such as for instance, a particular rule – because, as was discussed a moment ago, any such further mental item is *itself* subject to further interpretation. And any attempt to fix this second-order interpretation by a further interpretation is itself vulnerable to further interpretation. Such a line of response thus leads to a hopeless infinite regress of interpretations. It thus does appear that ‘no matter what is in my mind at a given time, I am free in the future to interpret it in different ways’ (ibid: 107).

Again, it is important not to become confused here by the talk of what fact constitutes my *previous* use of the word “plus”. The issue of past intentions is irrelevant to the thrust of the problem, which can be directed against appeal to an internalised mental rule which is supposed to be the source of my present applications of a word: ‘For the point is perfectly general: if “plus” is explained in terms of “counting”, a non-standard interpretation of the latter will yield a non-standard interpretation of the former’ (ibid: 16). And now, again, ‘the sceptical move can be repeated’ (ibid: 17). That the problem casts its shadow over grasping a determinate meaning, sense, or mental content in general is shown in the following summary of the problem, where Kripke has now left behind talk of past intentions regarding meaning:

But ultimately the sceptical problem cannot be evaded, and it arises precisely in the question how the existence in my mind of any mental entity or idea can *constitute* “grasping” any sense rather than another. The idea in my mind is a finite object: can it not be interpreted as determining a quus function, rather than a plus function? Of course there may be another idea in my mind, which is supposed to constitute its act of *assigning* an interpretation to the first idea; but then the problem obviously arises again at this new level (ibid: 54).

Leaving aside the question of my past intentions regarding the use of a word makes it easier to see that the problem is about the very possibility of conditions of correctness. The problem concerns how we can make sense of the fact that there could *even be* such a thing as a word, or rule, used according to a standard of correctness — that is, with a determinate meaning — regardless of the tense of our speech⁵. Boghossian puts the basic point well:

⁵ Thus, on p.22, Kripke ‘reruns the sceptical argument about my *present* intention to follow the plus sign...’ The train of reasoning which begins with casting doubt on my previous intentions regarding the word “plus” gets us anyway, but in an unnecessarily roundabout way, to the point of doubt about my presently meaning something determinate by the word: ‘if there can be no fact about which particular function I meant in the *past*, there can be none in the *present* either’ (Kripke 1982: 13). This follows given that, if the sceptical reasoning establishes that there can be no meaning-constituting fact about my previous behaviour, then we can easily imagine applying the same sceptical reasoning tomorrow, which would then retrospectively

Suppose the expression “green” means *green*. It follows immediately that the expression “green” applies *correctly* only to *these* things (the green ones) and not to *those* (the non-greens). The fact that the expression means something implies, that is, a whole set of *normative* truths about my behaviour with that expression: namely, that my use of it is correct in application to certain objects and not in the application to others (Boghossian 1989: 513)⁶.

To repeat, the problem is not primarily epistemological: ‘The sceptic does not argue that our own limitations of access to the facts prevents us from knowing something hidden’ (Kripke 1982: 39). The question of whether I can *know* whether something is correct presupposes that we are already operating with a condition of correctness, or according to a determinate meaning, which is the very thing being put into doubt: That ‘there is any fact of the matter as to which [rule — addition or quaddition] I meant’ in my use of the word “plus” is precisely what is doubted by the sceptic (ibid: 41). That the issue is constitutive, rather than epistemological, is dramatized by the sceptic’s claim that even ‘an omniscient being, with access to *all* available facts, still would not find any fact that differentiates between the plus and the quus hypotheses’ (ibid: 39)⁷.

confer the same sceptical conclusion upon my present meaning of “plus”. And, of course, if there can no such present fact constitutive of determinate meaning then there can be no such fact in the future either, so that the conclusion of the sceptical reasoning is that ‘[t]here can be no fact as to what I mean by “plus”, or any other word at any time’ (ibid: 21).

⁶ This remark continues, however:

...This is not, as McGinn would have it, a relation between meaning something by an expression at one time and meaning something by it at some later time; it is rather, a relation between meaning something by it at some time and its *use at that time*. The normativity of meaning turns out to be, in other words, simply a new name for the familiar fact that...meaningful expressions possess conditions of correct use’ (Boghossian 1989: 513).

Whilst I agree with the general thrust of this remark, as I pointed out above, given that Kripke clearly does trade off the “transtemporal” conception of normativity throughout his text, Boghossian is unfair to McGinn here.

⁷ It is precisely because the problem is constitutive – there are, supposedly, no facts which establish one’s meaning something determinate by one’s words – that Kripke’s sceptic rejects any imagined appeal – perhaps ‘under the influence of too much philosophy of science’ – to the notion that the hypothesis that I meant plus rather than quus ‘is to be preferred as the *simplest* hypothesis’:

Now simplicity considerations can help us decide between competing hypotheses, but they obviously can never tell us what the competing hypotheses are. If we do not understand

The fundamental issue, then, concerns normative notions: ‘How can I *justify* my present application of...a rule, when a sceptic could easily interpret it so as to yield any of an indefinite number of other results?’ (ibid: 17; emphasis added). That the problem is concerned with the very possibility of behaviour undertaken according to conditions of correctness is signalled by the fact that Kripke’s sceptic stipulates that any candidate fact produced in response to his problem ‘must, in some sense, show how I am *justified* in giving the answer “125” to “68 + 57”’ (ibid: 11; emphasis added). And this applies beyond the narrow issue of previous linguistic intentions, to the understanding of language *tout court*: ‘We all suppose that our language expresses concepts – in such a way that, once I “grasp” the concept, all future applications of it are determined (in the sense of being uniquely *justified* by the concept grasped’ (ibid: 107; emphasis in original). And ‘no “hypothetical” state could justify such a requirement’ because, and to reiterate, ‘it seems that no matter what is in my mind at a given time, I am free in the future to interpret it in different ways’ (ibid: 40, 107). Any hypothesis which is offered about the mental item which putatively underlies my mere external behaviour – past, present, or future - and which supposedly determines the rule I am following, can at best be conjecture, and so cannot serve to justify the determinacy of my meaning (whether, for instance, I mean addition or quaddition when I use the word “+”).

what two hypotheses *state*, what does it mean to say that the one is “more probable” because it is “simpler”? If the two competing hypotheses are not genuine hypotheses, not assertions of genuine matters of fact, no “simplicity” considerations will make them so (Kripke 1982: 38).

§2 Kripke and Boghossian Contra Dispositionalism

It is because an answer to the sceptic must show why I am *justified* in giving one answer rather than another that Kripke rejects a possible response to the sceptic which claims that my confidence that “125” and not “5” is the correct response is owing to the fact that, by virtue of the training I received in addition, or maybe as a result of certain innate natural propensities, I am *disposed* to give the one answer rather than another (ibid: 22-37). According to this line of objection, the problematic assumption of the paradoxical reasoning is that the meaning-constituting fact ‘must consist in an *occurrent* mental state’, whereas such facts are ‘to be analysed *dispositionally*’ (ibid: 22). According to proponents of this line of response, dispositional facts about language users enable us to differentiate between the hypothesis that one means addition from the hypothesis that one means quaddition. Thus, to mean addition by “+” is to be disposed, when asked to perform computations of the form “ $x + y$ ”, to give the *sum* of the two numbers; whereas by contrast, to mean quaddition by “+” is to be disposed, when asked to perform computations of the same form, to give the *quum* of the two numbers.

Such a naïve dispositional account, however, leaves no space for the crucial logical distinction between the answer I *should* give, as distinct from the answer I *would* give. According to this basic dispositional account, the answer I ought to give simply collapses into the answer that I do actually give - performance and “correctness” are simply equated with each other. If the meaning of an expression is equated with how I am disposed to use it, then this entails that *any* way in which I am disposed to use an expression thereby counts as the meaning of that expression. But this has the consequence that the basic dispositional theory fails to distinguish the hypothesis that I mean “plus” to denote the addition function from the hypothesis that I used the word to denote the quaddition function —which was the very distinction the theory was meant to uphold. This is due to the fact that equating performance with correctness leads to an unacceptably disjunctive conception of the meaning of the word “plus”: which word can be said to denote the addition function, whenever I am so disposed to use it; *and* to denote the quaddition function, whenever I am so disposed to use the word; *and also* any other application whatever I may be disposed to make of the word, *ad infinitum*...

The most serious objection to such a crude dispositional account, however, is that the equation of performance with correctness renders incoherent the very notion of error. The simple dispositional account leaves no room at all to distinguish mistaken responses from the response one ought to give. To put the point crudely, imagine a person suffering from Tourette's Syndrome who, when asked to perform "68 + 57" is, because of his condition, disposed to shake violently and shout obscenities. According to the basic dispositional account, there is no way to distinguish between this dispositional propensity of his and the fact that the response he ought to give is "125". Given this equation of performance with "correctness", the very notion of "mistakes", even one as radical as the Tourette's case, is simply incoherent. As Wittgenstein remarks at *PI* §258, in such cases, we lack an adequate 'criterion of correctness': 'One would like to say: whatever is going to seem right to me is right. And that only means that here we can't talk about "right."' This is unacceptable, as we surely must leave open the possibility of distinguishing error from accord, which is simply a basic requirement for making sense of any behaviour undertaken according to a condition of correctness. Kripke summarises the force of these objections as follows:

The point is *not* that, if I meant addition by "+", I *will* answer "125", but rather that...I *should* answer "125". Computational error, finiteness of my capacity, and other disturbing factors may lead me not to be *disposed* to act as I *should*, but if so, I have not acted in accordance with my intentions [to mean addition by "+"]. The relation of meaning and intention to future actions is *normative*, not *descriptive* (ibid: 37).

The arguments broached above are decisive against the crude dispositional theory, but what about the prospects of one not so basic? Could a more sophisticated variant of dispositionalism adequately meet the challenges posed to a theory of meaning in face of Kripke's paradox? In an insightful discussion of this question, Boghossian has argued powerfully, and to my mind convincingly, that dispositional theories are inadequate to the task of providing a satisfactory account of the normativity of meaning in face of the Kripkean paradox. This is because, he argues, dispositionalism is unable, simultaneously,

to maintain the two basic requirements which any such theory must meet in order to be minimally adequate as an account of the normativity of meaning. The first basic condition is that any such theory must specify, from in amongst all the dispositions I may have for application of the word “plus”, only ‘those dispositions which are *meaning-determining*’ (Boghossian 1989: 532). This is in order that the more sophisticated dispositional theory avoids the overly liberal disjunctiveness which shipwrecked the crude version. The second basic condition which must be met is that this specification of meaning-determining dispositions, on pain of vicious circularity, must be made in terms which do not presuppose intentional or semantic notions. As we shall soon discuss, Boghossian makes a convincing case for the conclusion that it is impossible for dispositional theories to satisfy both these conditions simultaneously (ibid: 538). Either the theory ends up presupposing what it was meant to explain (knowing what “plus” means), or the very notion of determinate meaning vanishes altogether in face of a disjunctive infinity of interpretations —the very consequence the dispositional theory was meant to enable us to avoid.

As we saw above, the crude theory is too liberally disjunctive: the meaning of “plus” is simply equated with the entire range of dispositions which I may possess in regard to the application of the word. But what if the meaning of the word “plus” were equated, not with the entire range of my dispositions in respect of the word, but only with a particular limited set of dispositions —only those dispositions which cause me to use “plus” according to its correctness conditions (that is to say, to denote the addition function)? In order to satisfy the first basic requirement of dispositionalist theories mentioned above, such an account would need to specify, from in amongst all the dispositions which can be brought into relation with my use of the word “plus”, only ‘those dispositions which are *meaning-determining*’; that is to say, the theory ‘must characterize...a property M such that possession of M is necessary and sufficient for being a disposition to apply an expression in accord with its correctness conditions’ (Boghossian 1989: 532). If meaning is equated, not with the entire range of my dispositions so to employ an expression, but only with ‘*the set of dispositions with respect to that expression that possess M*’, then the theory would no longer be unacceptably disjunctive: as Boghossian points out, ‘dispositions with that property will be guaranteed to be dispositions to apply the

expression correctly' (ibid.). That is to say, specifying dispositions with property M would only select those dispositions which lead me to use "plus" to denote the addition function. On the other hand, unlike the crude dispositional theory, which rendered incoherent the very possibility of error, a theory of select dispositions seemingly provides room for error: 'those dispositions not possessing M will not be dispositions to apply the expression to what it means and will be free, therefore, to constitute dispositions to apply the expression falsely' (ibid.). In other words, my disposition to use "plus" to denote the *quaddition* function would not fall within the set of dispositions with respect to that word that possess M.

These considerations, of course, immediately raise the question of whether there really is such a property M. But, as Boghossian observes, that question is, at least for Kripke, beside the point: for 'even if there were a dispositional predicate that logically covaried with a meaning predicate, the one fact could still not be identified with the other, for they are facts of distinct sorts' (ibid.). As we saw a moment ago, Kripke affirms that the specification of a disposition, no matter how selective, is descriptive: it specifies what a person does do, or is disposed to do; whereas, meaning is normative —that is to say, the fact of meaning implies standards of correctness which *ought* to be followed. The import of this distinction is that, even assuming we were successfully able to isolate a set of appropriately selected dispositions which perfectly mirror the correct extension of an expression, nevertheless such a description simply presupposes a standard of correctness (which we may grant, *per example*, is accurately tracked by the select set of dispositions).

So, it seems that a fully satisfactory reduction of the meaning of an expression must do more than simply accurately track the extension of the expression, but must also, as Boghossian notes, 'reveal that what [the reduction] is specifying is an *extension* —namely, a *correctness* condition' (ibid: 533). Yet this is what Kripke claims a dispositional theory — which can only ever be descriptive — cannot do. As Boghossian puts the point on Kripke's behalf:

There might be dispositions that logically covary with the extensions of expressions; so that one could read off the dispositions in question the expressions' correctness conditions. But the dispositional fact does not amount to the meaning fact, because it never follows from the mere attribution of any disposition, however selectively specified, that there are facts concerning *correct* use; whereas this does follow from the attribution of an extension. To be told that "horse" means *horse* implies that a speaker ought to be motivated to apply the expression only to horses; whereas to be told, for instance, that there are certain select circumstances under which a speaker is disposed to apply the expression only to horses, seems to carry no such implication (ibid.).

This objection, it seems to me, is decisive against dispositional theories. But Boghossian expresses caution over whether these considerations, taken on their own, apply generally across all dispositional theories of meaning. He leaves it an open matter whether there could possibly be 'a non-circular specification of how [a] person would *ideally* respond, as compared with how he actually responds', and which would account for the normativity of meaning.⁸ Given such caution, Boghossian concludes that there is 'no way to settle the matter in advance of the consideration of particular dispositional proposals' (ibid.). Nevertheless, Boghossian does offer an alternative argument for the conclusion that dispositional theories, in general, are inadequate to the task of overcoming Kripke's paradox. Assuming the holistic character of belief, Boghossian argues that it is impossible for dispositional theories of meaning to satisfy both conditions simultaneously —to specify the properties which constitute only those dispositions which are meaning-determining, but in non-semantic or non-intentional terms.

⁸ 'Perhaps', Boghossian reflects, 'the M-dispositions are those dispositions that a person would have when his cognitive mechanisms are in a certain state; and perhaps it can be non-question-beggingly certified that that state corresponds to a state of the *proper* functioning of those mechanisms' (ibid.). We will turn to consider such a proposal at various junctures throughout Chapters 2 and 3.

In order to avoid the devastating implications of the crude dispositional theory, a successful account of the possibility of using a word according to a specific meaning in dispositional terms must specify some property which determines that some select dispositions only are meaning-determining, and which are distinguished, therefore, from those dispositions which lack this meaning-determining property. Thus, for instance, Boghossian considers dispositional theories which appeal to optimality conditions, such that 'M is the property of being a disposition to apply (an expression) *in a certain type of situation*':

The idea behind such proposals is that there is a certain set of circumstances—call them “optimality conditions”—under which subjects are, for one or another reason, incapable of mistaken judgements; hence, we may equate what they mean by a given (mental) expression with the properties they are disposed to apply the expression to, under optimal conditions' (ibid: 537).

Thus, for someone to mean horse by “horse” is for that person to be disposed to call only horses “horse”, under circumstances where conditions are optimal⁹. When those optimal conditions pertain, it is supposed, it would be impossible for that person to believe that there is a horse in front of him otherwise than when there is indeed such a creature present: 'under those conditions, “horse” will get tokened (in the belief mode) only in respect of the property it expresses' (ibid: 538). This seemingly allows us to avoid the problems of overly disjunctive dispositional theories, which are unable to rule out dispositions to apply the term “horse” to cows in bad lighting, or to painted zebras, and so on *ad infinitum*... Such a theory also seems to allow for the possibility of error: mistakes arise when conditions are not optimal. According to such a theory of optimality

⁹ Boghossian is taking on the overall project of such theories, although he acknowledges that there are various proposals for what constitutes the conditions which are “optimal” in the relevant sense. Thus, for instance, he cites Fred Dretske, for whom 'optimal conditions are the conditions under which the meaning of the expression was first acquired'; or refers to teleological responses, such as those put forward by David Papineau and Jerry Fodor, amongst others, for whom 'optimality conditions are those conditions—defined by evolutionary biology—under which our cognitive mechanisms are functioning just as they are supposed to' (Boghossian 1989: 537).

conditions, then, in order to determine what any expression means, we must specify those particular properties which, when the relevant conditions are optimal, dispose a person to apply the expression “horse” only to horses. The specification of such properties (in optimal conditions), it is supposed, would provide us with a reconstruction of meaning facts in terms of dispositions. Such an account, it is supposed, would thus avoid vicious circularity by being couched in decidedly non-semantic and non-intentional terms. But this is precisely what Boghossian denies is possible: ‘there could not be naturalistically specifiable conditions’, he states, ‘under which a subject will be disposed to apply an expression only to what it means; and hence, that no attempt at specifying such conditions can hope to succeed’ (ibid: 537).

In reaching this conclusion, Boghossian points us to the essentially holistic character of belief-formation. That is to say, ‘under normal circumstances,’ belief-formation depends upon a whole network of already held beliefs and convictions: ‘what contents a thinker is prepared to judge will depend on what other contents he is prepared to judge’ (ibid: 539). Thus, for instance:

Neil may come to believe *Lo, a magpie*, as a result of seeing a currawong, because of his further belief that that is just what magpies look like; or because of his belief that the only birds in the immediate vicinity are magpies; or because of his belief that whatever the Pope says goes and his belief that the Pope says that this presented currawong is a magpie. And so on (ibid: 539-40).

Now, as we have seen, any minimally adequate dispositional theory has to specify a situation (“optimal conditions”) under which a person will be incapable of mistaken judgements about whether there is a magpie in the vicinity and, on pain of vicious circularity, must do so in non-semantic and non-intentional terms:

But the observation that beliefs are fixed holistically implies that a thinker will be disposed to think *Lo, a magpie* only in respect of an indefinite number of *non-magpies*, provided only that the appropriate background beliefs are present. Specifying an optimality condition for “magpie”, therefore, will involve, at a minimum, specifying a situation characterized by the absence of *all* the beliefs which could potentially mediate the transition from non-magpie to *magpie* beliefs (ibid: 540).

Without any such specification we could never be sure that we have adequately limited, from in amongst an indefinite number of possible dispositions, only those dispositions which lead a person to believe *Lo a magpie* when there is indeed such a bird present. Any adequate account of meaning would have to guarantee that there is, in the specified situation, a total absence of every one of the mediating beliefs which may lead a person to think *Lo a magpie* in regard to any non-magpie:

Since, however, there looks to be a potential infinity of such mediating background clusters of belief, a non-semantically, non-intentionally specified optimality situation is a non-semantically, non-intentionally specified situation in which it is guaranteed that none of this potential infinity of background clusters of belief is present (ibid.).

But how are we to provide such a guarantee? ‘What is needed,’ Boghossian points out, ‘is precisely what a dispositional theory was supposed to provide: namely, a set of naturalistic necessary and sufficient conditions for being a belief with a certain content’ (ibid.). The specification of optimality conditions, we will recall, was required in order that a theory of meaning enables us to select from a potential infinity of dispositions only those which are meaning-determining. But, given the considerations relevant to the holism of belief-formation, such a specification would need to provide a systematic way of weeding out all the inappropriate beliefs from in amongst the potentially infinite network of mediating beliefs. Thus, Boghossian draws the devastating conclusion:

But, of course, if we had *that* we would already have a reductive theory of meaning—we would not need a dispositional theory! Which is to say that, if there is to be any sort of reductive story about meaning at all, it cannot take the form of a dispositional theory (ibid.).

In other words, either the dispositional theory must presuppose that a person knows what the word “magpie” means, such that she applies the word only to magpies — which is the very ability the theory was meant to explain— or it simply falls back into so-called “meaning scepticism”, whereby an infinity of possible dispositions or interpretations can seemingly be made compatible with the meaning of an expression —which is precisely the implication such a theory was supposed to enable us to avoid. Boghossian claims that these issues ‘are general and apply to any dispositional theory whatever’, and this seems to me to be right, given that, as Boghossian points out: ‘All [meaning-determining] dispositional properties are such that their exercise...is contingent on the absence of interfering conditions, or equivalently, on the presence of ideal conditions’ (ibid: 528, 529). Regardless, in Chapter 2, we will consider responses to the constitutive paradox from commentators who otherwise explicitly repudiate dispositional theories—Kripke and Crispin Wright—but whose thought falls prey to this very same dilemma, and this because, despite their repudiations, their respective pictures of linguistic behaviour are, at bottom, essentially dispositional.

§3 Generality of the Putative Problem

Hopefully, it is clear by now that the focus of Wittgenstein's discussion (and the same can be said of Kripke's, once the confusions over previous linguistic intentions have been cleared away) is the possibility of manifesting understanding of a condition, or conditions, of correctness—for instance, using a word according to its meaning. Kripke, as we have seen, mostly focuses his discussion around a mathematical example—specifically, providing the value of an arithmetical sum. And Wittgenstein, too, discusses the development of an arithmetical series. The advantage of mathematical examples, as I have already mentioned, is that they serve to bring certain crucial implications of normative authority – of conditions of correctness – into especially sharp focus: acting in accordance with an understanding of the meaning of the addition function, for instance, determines a completely precise response at any stage of application. But, as I hope now to show, the point under consideration is quite general, and the apparent problem applies to the meaning of any word, rule, intentional act, or norm-governed activity whatsoever. Precisely because the putative problem is constitutive, the paradoxical train of thought can be repeated on behalf of any practice, or ability, which involves acting in accordance with a standard of correctness. Thus, Kripke refers to Wittgenstein's 'general problem about rules' (Kripke 1982: 111, fn. 86).

That the addition function determines a completely precise response at any stage of its application demonstrates that, as Kripke rightly points out, Wittgenstein cannot be indicting 'the rule of addition [as] somehow *vague*, or [that it] leaves some cases of its application undetermined' (ibid: 82). Wittgenstein, of course, recognises that the words we use in most instances—and unlike in the case of the addition function— do not have a 'fixed and unequivocal use for me in all cases' (PI §79). Nevertheless, this does not mean that a word whose use is comparatively vague or undetermined does not involve conditions of correctness of its use: 'are we to say that we do not really attach any meaning to [a] word, because we are not equipped with rules for every possible application of it?' (PI §80). For instance, Wittgenstein imagines giving someone the order "Stand roughly here" (PI §88): granted, there isn't a precise determination of where the person is meant to stand, comparable to the determinate precision of the arithmetical case, but having said that, even without such a comparatively precise standard of

exactness, there is nevertheless here a condition of correctness—if the person nonchalantly walks off, for instance, then the order remains unfulfilled.

Even though the arithmetical function determines a specific response to an innumerable number of possible imagined applications of the rule, it is worth noting here that Wittgenstein's reflections in this area are not concerned to make a particular point about infinity. For instance, at PI §149, Wittgenstein reflects on what it is that grounds the determinate steps required of 'knowing the ABC', and so, as McDowell remarks, the thrust of Wittgenstein's discussion cannot be 'essentially connected with the fact that there is no application for the idea of getting to the end of extending a number series':

There *is* such a thing as getting to the end of reciting the ABC.

Wittgenstein's point is not about infinity. It is about the relation between knowing the ABC, or understanding the principle of a number series, on the one hand, and the actions one performs in reciting the ABC, or extending the number series, on the other' (McDowell 2009a: 80, fn. 3).

The issue at hand, then, concerns the "normativity of language"—the fact that using a word according to its meaning commits one to (more or less determinate) conditions of correctness. As Kripke admits, 'these problems apply throughout language and are not confined to mathematical examples' —they are 'completely general and can be applied to any rule or word' (Kripke 1982: 19, 58). Consequently, Kripke confesses, the problem 'of course applies to predicates of sensations, visual impressions, and the like, as well' (ibid: 20). For instance, Kripke's sceptic challenges: 'Perhaps by "green", in the past I meant *grue*...'; and any attempt to fix the meaning of "green" by stipulating its use only to items of the same color as *this* <pointing to a blue sample> for 'the sceptic can reinterpret "same color as the same *schmolor*, where things have the same *schmolor* if...' (ibid.)¹⁰. Wittgenstein's paradox is, Kripke states, that 'anything in my head leaves it undetermined *what* function "plus" (as I use it) denotes (plus or quus), what "green" denotes (green or grue), and so on'; and 'his paradox shows, among other things, that every explanation of

¹⁰ Kripke transposes the same reasoning to the concepts of sensation: 'let "pickle" apply to pains before t, and tickles thereafter...!' (Kripke 1982: 62, fn. 49).

a rule could conceivably be misunderstood' (ibid: 82). Thus, Kripke admits that 'Wittgenstein's problem is that it appears that he has shown *all* language, *all* concept formation, to be impossible, indeed unintelligible' (ibid: 62).

It is worth noting also that the problem is not confined to explicitly verbal, or discursive responses to the expression of a rule. Thus, at *PI* §85, Wittgenstein repeats the paradoxical reasoning in the case of responding (non-verbally) to a sign-post:

—Does the sign-post leave no doubt open about the way I have to go?
Does it shew which direction I am to take when I have passed it; whether along the road or the footpath or cross-country? But where is it said which way I am to follow it; whether in the direction of its finger or (e.g.) in the opposite one?—And if there were, not a single sign-post, but a chain of adjacent ones or of chalk marks on the ground—is there only *one* way of interpreting them?

Furthermore, Wittgenstein clearly sees that the problem generalises over all intentional notions too, such as wishes, desires, beliefs, and so on. Thus, for instance, Wittgenstein raises the paradox in the context of an order and the specific circumstances that bring the order to fulfilment:

When we give an order, it can look as if the ultimate thing sought by the order had to remain unexpressed, as there is always a gulf between an order and its execution. Say I want someone to make a particular movement, say to raise his arm. To make it quite clear, I do the movement. This picture seems unambiguous till we ask: how does he know that *he is to make that movement*?—How does he know at all what use he is to make of the signs I give him, whatever they are?—Perhaps I shall now try to supplement the order by means of further signs, by pointing from myself to him, making encouraging gestures, etc. Here it looks as if the order were beginning to stammer.

As if the signs were precariously trying to produce understanding in us. —But if we now understand them, by what token do we understand? (PI §433; *cf.* PI §461).

It is clearly in reference to such remarks that Kripke claims: ‘the important thing for Wittgenstein is that my present mental state does not appear to determine what I *ought* to do in future’ (Kripke 1982: 56)¹¹. Finally, at §437, connects the problem, not only with desire and its fulfilment, but with the possibility of a thought, or a declarative sentence, having a determinate content:

A wish seems already to know what will or would satisfy it; a proposition, a thought, what makes it true – even when that thing is not there at all!
Whence this *determining* of what is not yet there? This despotic demand?
(“The hardness of the logical must.”)

Thus, consistent with his claim that Wittgenstein accepts the implications of the paradox, Kripke concludes that Wittgenstein rejects ‘the natural presupposition that meaningful declarative sentences must purport to correspond to facts’ (Kripke 1982: 78-79). But, as Kripke also recognises, the problem arises as to how *any* mental representation, of whatever kind, can be understood as having a determinate content, owing to the allegation that ‘the components of such “mental representations” do not have interpretations that can be “read off” from them in a unique manner...So *a fortiori* there is no such...unique interpretation of a mental representation’ (ibid: 85).

In summary, then, the upshot of the preceding arguments, Kripke maintains, is the following “sceptical paradox”:

There can be no such thing as meaning anything by any word. Each new application we make is a leap in the dark; any present intention could be

¹¹ Notice that this remark (as do those of Wittgenstein) concerns a *present* state of mind. Setting up the problem does not require the mental acrobatics involved in connecting intentions, expectations, and commands made in the present, with my previous aims and intentions. Moreover, it would be more correct to say that the important issue for Wittgenstein is that my present mental state *does* determine what I *ought* to do in future (I think Kripke misses the irony in Wittgenstein’s remarks).

interpreted so as to accord with anything we may choose to do. So there can be neither accord, nor conflict (ibid: 55).

Furthermore, Kripke claims that Wittgenstein himself accepts the conclusion of the “sceptical paradox”. Despite Kripke’s early qualification that his ‘paper should be thought of as expounding neither “Wittgenstein’s” argument nor “Kripke’s”’: rather Wittgenstein’s argument as it struck Kripke, as it presented a problem for him’ (ibid: 5), he is later quite explicit in attributing acceptance of the sceptical paradox to Wittgenstein. Thus, although Kripke does acknowledge that Wittgenstein might disapprove of the precise form in which Kripke presents the sceptical paradox, nevertheless, he immediately goes on to state, ‘I choose to be so bold as to say: Wittgenstein holds, with the sceptic, that there is no fact as to whether I meant plus or quus’ (ibid: 70-1). And similar attributions can be found elsewhere: ‘[Wittgenstein] accepts his own sceptical argument’ (ibid: 68); ‘he agrees with his own hypothetical sceptic that there is no such fact [about meaning something by a word], no such condition in either the “internal” or the “external” world’ (ibid: 69).

Kripke calls upon the first paragraph of §201 of the *Investigations* as apparent evidence for Wittgenstein’s acceptance of the sceptical argument:

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

If we lack a notion of accord or conflict, then we lack conditions of correctness. And given that meaningfulness depends on conditions of correctness, we seem to lack a notion of meaning anything by our words. On Kripke’s interpretation, then, Wittgenstein seems to

be faced with the ‘incredible and self-defeating conclusion, that all language is meaningless’ (Kripke 1982: 71). But given the preceding discussion, if Kripke’s claim that Wittgenstein accepts the paradox were true, then that would commit him to the radical claim that not only all language, but *all* concept use, understanding of *all* rules, signs, and symbols, *all* mental representations, *all* intentional thought and action — in short, all thought and action involving conceptions of correctness — are not only impossible, but incoherent! The paradox is such that it is left entirely mysterious how the sceptic has so much as managed to communicate the problem to us at all! Truly: ‘The ladder must finally be kicked away’ (ibid: 21). As Baker and Hacker rightly point out, Kripke’s “sceptical paradox” ‘is not scepticism at all, it is conceptual nihilism, and, unlike classical scepticism it is *manifestly* self-refuting’ (Baker & Hacker: 1984: 6).¹² Even for someone who lacked Wittgenstein’s commitment to the avoidance, in one’s philosophical writings, of unnecessary controversy and stating only what everyone admits (PI§128), such a position can only be an embarrassment.

Certainly, if Kripke were correct that Wittgenstein did accept the implications of the paradox (whatever “acceptance” here could possibly mean) then, truly, the Austrian philosopher wasn’t exaggerating when, during a reflection upon his work in philosophy he said to himself: “‘I destroy, I destroy, I destroy...’” (Wittgenstein: 2006: 19e). No wonder Kripke expresses astonishment at the implications of the “sceptical” paradox which, he confesses, ‘struck me with the force of a revelation’ (Kripke 1982: 1). The desperate (but seemingly unavoidable) implications of the paradox is that ‘the entire idea of meaning’ simply ‘vanishes into thin air’; and our perfectly everyday ability to understand the meaning of spoken and written words appears ‘completely mysterious’ (ibid: 22, 51). Thus we can surely sympathise with Kripke when he confides that: ‘Sometimes, when I have contemplated the situation, I have had something of an eerie feeling’ (ibid: 21).

¹² Boghossian accuses Baker and Hacker of failing to understand that Kripke’s problem is concerned with constitutive scepticism, not empirical scepticism (Boghossian 1989: 515-516, fn. 14). But I think this unfair, as suggested by the remark just quoted.

As almost all commentators have pointed out, however, in the exact same remark which Kripke cites as confirmation that Wittgenstein accepts the sceptical paradox, he clearly goes on to state that the paradox should be rejected as based on misunderstanding:

It can be seen that there is a misunderstanding here from the mere fact that in the course of our argument we give one interpretation after another; as if each one contented us at least for a moment, until we thought of another standing behind it. What this shews is that there is a way of grasping [or of understanding] a rule which is *not* an *interpretation*, but which is exhibited in what we call “obeying the rule” and “going against it” in actual cases (PI§201).

The key to understanding this passage, then, is to reflect appropriately on the implications of Wittgenstein’s assertion that our thought is based on a misunderstanding—referred to later in the same numbered remark as ‘a tendency to say: any action according to the rule is an interpretation’—and to give an account of what Wittgenstein means by “a way of grasping a rule which is *not* an *interpretation*”.¹³ This is a matter of some controversy, and we will need to return to consider the import of Wittgenstein’s remark in greater detail in Chapter 3. But the passage provides pretty clear evidence that, *pace* Kripke, Wittgenstein rejects the paradoxical reasoning as based on misunderstanding. Further *prima facie* evidence is provided by the following remark from *Zettel* §448: ‘We want to replace wild conjectures and explanations by quiet weighing of linguistic facts’. As is clear from the original context, this remark concerns the mature Wittgenstein’s whole approach to philosophical practice and methodology — it seems curious, to say the least, that he should base his whole philosophical enterprise on the quiet weighing of something – linguistic facts – which he elsewhere claims are non-existent, or even incoherent!¹⁴

¹³ In Chapter 3, I will argue that it is partly a failure to pay proper heed to the implications of Wittgenstein’s diagnosis of misunderstanding in our thought process that is a critical source of failure to understand Wittgenstein’s insight (there is a tendency, instead, to try to place substantial explanatory weight on the obverse of the misunderstanding: ‘there is a way of grasping a rule which is *not* an interpretation’).

¹⁴ Consider *OC*§306: ‘And isn’t it an empirical fact—that *this* word is used like *this*?’; and also *PI* §79: ‘—Say what you choose, so long as it does not prevent you from seeing the facts’. It is, again, clear from the

Until we find ourselves in a position to discuss the matter further, let us only observe that Wittgenstein, at PI §428, remarks that the conditions of correctness embodied in intentional thought do not strike us as problematic during the ordinary course of thinking, but only during our distorted reflection upon that employment. As if in response to Kripke's "eerie feeling" in face of the apparent mysteriousness of mental content, Wittgenstein remarks:

"This queer thing, thought" – but it does not strike us as queer when we are thinking. Thought does not strike us as queer when we are thinking, but only when we say, as it were retrospectively: "How was that possible?" How was it possible for thought to deal with the very object *itself*? We feel as if by means of it we had caught reality in our net.

context that Wittgenstein is here talking of *linguistic* facts. The remark continues, in parenthesis: 'And when you do see them [the linguistic facts] there is a good deal that you will not say.'

§4 Kripke and Wright on “Non-misinterpretable Interpretations”

It should be clear enough by now that the apparent problem raised Wittgenstein’s is, as Kripke recognises, ‘completely general’ (Kripke 1982: 62). Crispin Wright, whose reading, as we shall see, mirrors that of Kripke in certain crucial respects, acknowledges the completely general nature of Wittgenstein’s discussion: ‘the principal issues to do with rule-following impinge on every normatively constrained area of human thought and activity: on every institution where there is right and wrong opinion, correct and incorrect practice’ (Wright 2001: 1). Wright’s reading of Wittgenstein merits attention, not only because he has devoted much published work to these very same problems but, more interestingly, because, despite the fact that Wright developed his interpretation independently of Kripke, both readings (despite certain variations in emphases) converge in many respects.

‘For the believer in objectivity’, Wright states, ‘human opinion in no sense constitutes truth; truth is in no sense dependent on human opinion’ (Wright 2001b: 33). Thus, Wright is justly keen to acknowledge that our intuitive notion of objectivity is bound together with the notion of *investigation-independence*: the idea that ‘confronted by any decidable, objective issue, there is *already* an answer which – if we investigate the matter fully and correctly – we will arrive at’ (ibid.). For instance, I make the judgement that “the cat is on the mat” because I believe there to be a particular cat on a particular mat. Now, either there is a particular cat on a particular mat, in which case my judgement is true; or, alternatively, it *only seems* to me that there is a particular cat on a particular mat – that is to say, in actual fact, the mat lacks a cat, and my judgement is false. Such normative constraints on our thought and utterances are undeniable if we are to maintain any kind of grip on notions such as meaning, rationality, and objectivity at all. And yet, according to Wright, Wittgenstein ‘is, rightly or wrongly, sceptical about investigation-independence’ (ibid.). As Wright points out, the ‘idea of investigation-independence...leads us to look upon grasping the meaning of an expression as grasping a general pattern of use, conformity to which requires certain determinate uses in so far unconsidered cases’ (ibid.). And yet, not unlike Kripke, Wright claims that Wittgenstein

‘apparently rejects the idea that the meaning of an expression is anything which is properly seen as constraining a certain sort of future use on it’ (Wright 2001a: 14).

The basis of this claim finds its source, as does Kripke’s reading, in passages such as §§185-188 of the *Investigations* where Wittgenstein raises the possibility of a “deviant rule-follower”. There, we are asked to imagine a pupil who is being trained into competence of the development of an arithmetical series for the first time (say, “+2”) and, not unlike in Kripke’s discussion, we are invited to imagine that the exercises and tests thus far performed by the pupil have not ventured beyond a certain limit —say, not beyond 1000. Nevertheless, on the evidence of his performance so far, we are satisfied with the pupil’s competence with the rule. ‘Now’, Wittgenstein continues, ‘we get the pupil to continue the series...beyond 1000—and he writes 1000, 1004, 1008, 1012’. Perhaps, Wittgenstein reflects: ‘It comes natural to this person to understand our order with our explanations as we should understand the order: “Add 2 up to 1000, 4 up to 2000, 6 up to 3000 and so on”’¹⁵:

We say to him: “Look what you’ve done!”—he doesn’t understand.

We say: “You were meant to add *two*: look how you began the series!”

—He answers: “Yes, isn’t it right? I thought that was how I was *meant* to do it.”

—Or suppose he pointed to the series and said: “But I went on in the same way.”

—It would now be of no use to say: “But can’t you see....?”—and repeat the old examples and explanations (*PI* §185).¹⁶

¹⁵ The remark continues: ‘Such a case would present similarities with one in which a person naturally reacted to the gesture of pointing with the hand by looking in the direction of the line from finger-tip to wrist, not from wrist to finger-tip’ (*PI* §185).

¹⁶ This expression is taken up again at *PI* §231: “‘But surely you can see....?’” That is just the characteristic expression of someone who is under the compulsion of a rule’ — which, of course, is precisely what is under question here.

Such repetition is clearly useless, as it is precisely these pre-millennial examples and explanations of the rule “Add 2” which have been misunderstood. Each and every one of the (oral or written) expressions and examples of the rule “Add 2” proffered thus far (taken individually or as an aggregate), we might suppose, are open to a variety of alternative interpretations. An utterance of “Add 2”, say, doesn’t in and of itself, we suppose, determine which series of signs the pupil needs to write down, such that the signs which the pupil does in fact write down are in accord with the rule expressed. The expression of a rule is, we might presume, at bottom nothing but the emission of sounds (if expressed verbally) or a mere pattern of inscribed shapes in some physical medium (if given in writing). It is tempting now to suppose that the pupil will need to *interpret* the teacher’s utterance of “Add 2”, such that this interpretation determines the correct series of signs which are to be written down in response to the rule. Thus, we are led to picture the pupil as having selected one such interpretation from a variety of candidate interpretations — one which, as it so happens, deviates from the correct interpretation of the rule, as it was initially meant by the teacher’s expression. (We need not imagine that the pupil here *explicitly* selects one interpretation from in amongst a range of considered alternatives. It is enough that we consider the series of signs the pupil writes down - perhaps unreflectively and instinctually – as embodying, so to speak, or behaviourally manifesting, *his interpretation* of the rule from in amongst the imagined background of alternative interpretations —not considered explicitly, in this instance).

Furthermore, if we suppose that the pupil must (perhaps unreflectively) *interpret* each of the teacher’s pre-millennial rule-expressions, offered up for example and instruction, then the available evidence for hitting on the correct interpretation – the spoken or written rule formulations offered thus far – leave it undetermined which is the correct series of signs to write down, from in amongst other imagined possibilities. All examples of the rule offered thus far have *ex hypothesi* been given only up to 1000, and so each of the oral or written instructions (taken individually or aggregately) are compatible, we might suppose, not only with the “correct” interpretation (the one which was initially meant by the teacher’s instruction to “Add 2”), and the pupil’s deviant interpretation (“Add 2 up to 1000, 4 up to 2000, 6 up to 3000 and so on”), but also with a seemingly infinite expanse of other possible (and deviant) interpretations: “Add 2 up to 1000, 3 up

to 2000, 4 up to 3000”; or “Add 2 up to 1000, then stop” — or any other possible interpretation, seemingly compatible with the instruction given thus far, one cares to dream up.

‘How,’ then, Wittgenstein ponders in PI§186, ‘is it decided what is the right step to take at any particular stage?’. Or, as we perhaps might put it, what rules out all other interpretations of the order as deviant?

—“The right step is the one that accords with the order—as it was meant”

But, comes the retort:

—So when you gave the order +2 you meant that he was to write 1002 after 1000—and did you also mean that he should write 1868 after 1866, and 100036 after 100034, and so on—an infinite number of propositions?

We will need to return to consider this response in a moment, as it gestures towards certain broadly Platonist pictures of a rule, appeal to which can be tempting as a putative way of avoiding the threat of a regress of interpretations. The Platonic conception, if you like, forms one horn of a dilemma, the other horn of which is constituted by the regress of interpretations argument, presently under discussion. As the interlocutor anyway disavows this line of response — at this stage of the reasoning at least — let us return to the response that the interlocutor does now give:

“No: what I meant was, that he should write the next but one number after *every* number that he wrote; and from this all those propositions follow in turn.”

But now we appear to be in trouble. In an attempt to secure the notion of the correct interpretation of the teacher's expression of the rule, as it was meant, the interlocutor invokes a further rule for fixing the connection between the initial expression of the rule, "Add 2", on the one hand, and behaviour which manifests understanding of the rule, on the other. The correct interpretation of the original expression, we want to say, is not any of those other candidate interpretations that can be imagined, but the rule which can be expressed thus: "write out the next but one number from the series of natural numbers which you have already mastered," or some such. The way I meant my instruction "Add 2", we want to say, is in accord with only *this* rule and not any other.

—But that is just what is in question: what, at any stage, are we to call "being in accord" with that proposition (and with the *mean*-ing you then put into the proposition—whatever that may have consisted in)?

On the supposition that our initial formulation of the rule – whether an oral or written instruction – is open to a variety of possible interpretations, then what bars us from supposing that this further expression of the rule – in speech or inscribed – is not itself open to a variety of possible interpretations? The original expression of the rule, we have supposed, does not in and of itself determine which steps are in accord with the meaning of the rule. So, on what basis do we suppose that this further expression of the rule *does* in and of itself determine the correct steps? Any further expression of a rule we care to give will be in the same boat as the original rule-expression. And, now, any hope we may nurture of offering another rule to fix the interpretation of this further rule will be dashed on the rocks of a looming, and vicious, regress. It is thus left open to doubt whether the interlocutor's rule for interpreting a rule does, or indeed even can, serve as a means for correcting the misunderstanding, or misinterpretation. As Wittgenstein remarks at PI§198: 'every interpretation, together with what is being interpreted, hangs in the air; the former cannot give the latter any support' — or, perhaps at very best, as *PI* §201 has it, the interpretation offers support, but 'only for a moment, until we thought of yet another standing behind it'. The inevitable, and devastating, conclusion for this line of reasoning is thus made at PI§201:

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

Once we have embarked upon this train of thought, one particularly tempting way to fix a determinate rule or pattern, but in such a way as to avoid a regress, is to appeal to a further rule, or to an interpretation, which is somehow invulnerable to further interpretation. In other words, this is to picture an interpretation which allows us to fix a connection between the verbal or written expression of a rule, such as "Add 2", on the one hand, and a determinate pattern of application made in accordance with that very same rule, on the other. On pain of vicious regress, however, this interpretation must, at the same time, itself be, as it were, immovably fixed: it cannot possibly be interpreted otherwise than as connecting the verbal or written expression of the rule with precisely *these*, and only *these* actions and responses which are in accord with the rule. This is so that the possibility of further deviant interpretations is foreclosed, and the subsequent threat of regress avoided. Of course, we are tempted to say, the mere verbal utterances or written inscriptions of the rule (" +2" or "Add 2", and so on...) don't determine the correct series of signs which need to be written down. What *does* determine the correct steps, however, is the *meaning* of the rule, conceived as an interpretation which is itself immune to misinterpretation. As McDowell notes, with such a picture we are tempted into a 'conception of meaning as just like an ordinary interpretation' – in that it fixes a determinate pattern of application – 'except that it is somehow not itself susceptible to interpretation' – and so is in this sense utterly unlike, as we have repeatedly seen, an ordinary interpretation (McDowell 2009a: 83). McDowell refers to Wittgenstein's remark in the *Blue Book*, which clearly gives expression to this temptation to picture a very special species of interpretation: 'What one wants to say is: "Every sign is capable of interpretation but the *meaning* mustn't be capable of interpretation. It is the last interpretation"' (Wittgenstein 1969: 34)¹⁷.

¹⁷ See also, PI§230: '...it is my *last* arbiter for the way I am to go.'

This picture of a special, regress-stopping interpretation is particularly tempting in a mathematical context. Thus, Wittgenstein speaks of ‘the beginning of a series’ as ‘a visible section of rails invisibly laid to infinity’ and the imagery of ‘infinitely long rails’ corresponding ‘to the unlimited application of a rule’ (*PI* §218)¹⁸. Such rigid rails, Wright notes, are pictured ‘as somehow constituted quite independently of any propensities for any judgement or reaction of ours’ (Wright 2001: 3). Such ‘absolutely hard conceptual constraints’ are pictured as providing the ‘ultra-objective’ standards over against which our practical applications of the rule or concept can be judged: such an imagined rule, as Wright puts it, ‘conducts you like a gangway with rigid walls’ (Wright 2001a: 11). Following according to the rule of a series is thus pictured as ‘the tracing of steps which are, in some sense, *already* drawn and which we have no rational option but to acknowledge once presented to us’ (Wright 2001: 3). “All the steps are already taken”, Wittgenstein remarks at *PI* §219, ‘means: I no longer have any choice.’

Recall this stage of discussion at *PI* §186 where the interlocutor protests, in face of the pupil’s deviant interpretation:

—“The right step is the one that accords with the order—as it was meant.”

But are we to imagine that each and every one of the possible steps of the series “add 2” was somehow traversed in my mind when I meant the order?

—So when you gave the order +2 you meant that he was to write 1002 after 1000—and did you also mean that he should write 1868 after 1866, and 100036 after 100034, and so on—an infinite number of propositions?

¹⁸ See also, *PI* §229: ‘I believe that I perceive something drawn very fine in a segment of a series, a characteristic design, which only needs the addition of “and so on”, in order to reach to infinity.’

As we will recall, the interlocutor rejects this response in order, instead, to offer a further rule (“write down the next but one number...”) for interpreting the initial rule (“Add 2”). However, this was before the full force of the regress of interpretations was felt (and was a decisive step towards it), and so, in face of this threat, it might now be more tempting to reach for the regress-stopping mythology, whose idea is that, as Wittgenstein puts it at PI§188:

[T]hat act of meaning the order [“Add 2!”] had in its own way already traversed all those steps: that when you meant it your mind as it were flew ahead and completed all the steps before you physically arrived at this or that one.

Thus you were inclined to use such expressions as: “the steps are already taken, even before I take them in writing or orally or in thought.” And it seemed as if they were in some *unique* way pre-determined, anticipated—as only the act of meaning can anticipate reality.

This is clearly the picture of the “ultra-objective” rails, constituted in total independence of any steps taken, but which nevertheless determine those steps “like a gangway with rigid walls”, as Wright says. Once we have been seduced into this picture, those fulgurations of understanding – “I’ve got it!”, “Now I can do it...!”, “Now I can go on...”, etc. – begin to look like the visible indication that we have – somehow – been put in touch with these ultra-objective rails, wherever they may lie hidden — whether in our minds or in some abstract Platonic realm (cf. PI§191). It is as if all the steps, as Wittgenstein says at PI§193, ‘had to be really – in a mysterious sense – already present.’ At §PI184, Wittgenstein transposes the case of “grasping a whole series in a flash” with sudden remembrance of a tune:

I want to remember a tune and it escapes me; suddenly I say “Now I know it” and I sing it. What was it like to suddenly know it? Surely it can’t have occurred to me *in its entirety* in that moment!—Perhaps you will say: “It’s a

particular feeling, as if it were *there*”—but *is* it there? Suppose I now begin to sing it and get stuck?—But may I not have been *certain* at that moment that I knew it? So in some sense or other it was *there* after all!—But in what sense?

Kripke expresses the same puzzlement in relation to knowing how to operate with the addition function over all possible applications. ‘Ordinarily,’ Kripke supposes, ‘in computing “68 + 57” as I do, I do not simply make an unjustified leap in the dark’ (Kripke 1982: 10). I am inclined to think that my present grasp of this determinate rule - when I ‘unhesitatingly and automatically’ give the answer “125” as my response - is, as Kripke puts it, ‘unquestioned and unquestionable’ (Kripke 1982: 15, 14): ‘Just this is the difference between someone who computes new values of a function and someone who calls out numbers at random’ (ibid: 17). Our confident mastery of addition is such that ‘we think of ourselves as *guided* in our application of it to each new instance’ (ibid; original emphasis). There is ‘one and only one answer [that] is dictated as the one appropriate to “68 + 57”’ (ibid: 17-18). Given this fact, then, it is as if my capacity to follow this rule is ‘engraved on my mind as on a slate’ (ibid: 15). But the paradoxical reasoning appears to have made a total mystery of the very possibility of such a determinate rule: ‘Whence the *determining* of what is not yet there? This despotic demand?’ (PI§437). Consider, for instance, PI §197, where Wittgenstein discusses the intention to play a game of chess:

Don’t I know, then, which game I want to play until I *have* played it? Or are all these rules contained in my act of intending? So is it impossible for me to be certain of what I am intending to do? Is it experience that tells me that this sort of game is the usual consequence of such an act of intending? So is it impossible for me to be certain of what I am intending to do? And if this is nonsense – what kind of super-strong connexion exists between the act of intending and the thing intended?

‘Perhaps,’ Kripke speculates, in one last gasp attempt to avoid the sceptic’s final triumph, ‘we may try to recoup, by arguing that meaning addition by “plus” is a state even more *sui generis* than we have argued before’:

Perhaps it is simply a primitive state, not to be assimilated to sensations or headaches or any “qualitative states”, nor to be assimilated to dispositions, but a state of a unique kind of its own. Such a move may in a sense be irrefutable, and if it is taken in an appropriate way Wittgenstein may even accept it (Kripke 1982: 51).

And yet:

But it seems desperate: it leaves the nature of this postulated primitive state – the primitive state of “meaning addition by ‘plus’” – completely mysterious. It is not supposed to be an introspectible state, yet we supposedly are aware of it with some degree of certainty whenever it occurs. For how else can each of us be confident that he *does*, at present, mean addition by “plus”? (ibid)

As the voice within the text of the *Investigations* is moved to cry out: ‘...but in a *queer* way,’ the use does seem to us to be ‘in some sense present’! (*PI* §195). Kripke registers his sympathy with this urge: ‘we surely suppose – unreflectively – that something like this is indeed the case. Even now I have a strong inclination to think this somehow must be right’ (Kripke 1982: 52). But, he continues to ponder: ‘What can that sense be? Can we conceive of a finite state which *could* not be interpreted in a quus-like way? How could that be? ...[T]he nature of the supposed “state” is left mysterious’ (ibid: 52-53).

Wittgenstein demonstrates his sensitivity to the seeming hopelessness of such an appeal: ‘You have no model for this superlative fact...’ (*PI* §192). We can again, perhaps,

sympathise with Kripke when he exclaims: ‘Sometimes, when I have contemplated the situation, I have had something of an eerie feeling’ (Kripke 1982: 21).

Wittgenstein clearly rejects such pictures as a ‘mythological description of the use of a rule’ (PI§221)¹⁹. Wright, too, is justly concerned to reject this picture of an “ultra-objective” rule. Once such a picture is accepted, Wright remarks, ‘it encourages, if it does not make absolutely inevitable, a drift into the idea that each of us has some sort of privileged access to the character of her own understanding of an expression’ (Wright 2001b: 35-6). Each person must hit on to the right invulnerable interpretation for herself, and thus, ‘we move towards the idea that understanding an expression is a kind of “cottoning on”—that is, a leap, an inspired guess at the pattern of application’ which is to be followed:

It becomes almost irresistible to think of someone who is learning a first language as if she were forming general *hypotheses*. “Cottoning on” would be forming the right hypothesis; and failing to do so would be forming the wrong, or no hypothesis. And the leap involved would just be that with the best will in the world we – the instructors – cannot do better than leave an indefinite variety of hypotheses open for selection (ibid: 35).

¹⁹ Wittgenstein does not, however, deny that *any* sense can be made of such claims as “But I already knew at the time when I gave the order, that he ought to write 1002 after 1000”. See PI §187, for instance, where Wittgenstein is perfectly happy to affirm the sense of such claims, expressed in conditionals of the form: “If I had then been asked what number should be written after 1000, I should have replied ‘1002’”: ‘And that I don’t doubt. This assumption is rather of the same kind as: “If he had fallen into the water then, I should have jumped in after him.”’ We will be in a better position to interpret the import of such passages in Chapter 3. It is clear enough from §187, however, that a version of the “superlative” model of such claims is to be rejected: ‘For you don’t want to say that you thought of the step from 1000 to 1002 at that time—and even if you did think of this step, still you did not think of other ones.’ (In Chapter 2, §c I will argue that Wright conflates the *sui generis* fact with the superlative mythology, such that rejection of the latter *eo ipso* entails rejection of the former; it is hoped that it will be clear by then that this is the source, too, of Kripke’s failure to conceive the state of using a word according to a determinate meaning as anything but “eerie” and mysterious.)

It now looks as if it can be no more than a 'strong presumption, when sufficient evidence has accumulated', that the interpretation we hit on 'is shared communally' (ibid: 36). Nevertheless, each individual, Wright continues, has certain knowledge 'of a particular idiolectic pattern of use which she intends her use of an expression to subserve' (ibid.). 'I cannot know for sure what rule another is following', Wright contends (this is because, it is supposed, I can, at best, only ever have inductive grounds for the rule the other has in mind, based on the evidence of his behavioural responses to the rule); nevertheless, 'I can be certain, we would ordinarily allow, of the rule I am following and of what it requires me to do' (ibid: 22). What is recognised as the correct characterisation of an expression 'may mean different things in different mouths', Wright continues, yet there is 'a pattern which for each of us is transparent to himself' (ibid: 27; 30)²⁰. Correlatively, in exchange for each being granted the freedom to choose whatever meaning we wish for the rule, we also grant this courtesy to each of our fellow "rule-followers": 'However many rules we give him, he can supply an alternative interpretation of what he is doing' (Wright 2001a: 12). Wittgenstein too reflects this reasoning at *PI* §293: 'If I say of myself that it is only from my own case that I know what the word "pain" means—must I not say the same of other people too?'

The picture of "cottoning on" finds expression in Wittgenstein's text too: "Don't you get [the other person] to *guess* the essential thing?" (*PI*§210). And the picture of "idiolectic understanding" is clearly present in the relevant parts of Kripke's dialectic. Given that the rule that the teacher is trying to impart is seemingly open to an endless array of possible interpretations (or hypotheses), and the sheer breadth of possible interpretations cannot be narrowed by further rules - such as further explanation of meaning - as such further rule-expressions are themselves seemingly open to a breadth of further interpretations (or hypotheses), then there must come a point where I just have to simply go by the way the rule strikes me at any one time, without justification or explanation: 'Nothing justifies a brute inclination to answer one way rather than another' (Kripke 1982: 15). Thus, Kripke's individual rule-follower is, like Wright's, entitled to his own purely idiolectic

²⁰ Wittgenstein gives expression to this temptation in *Remarks on the Foundations of Mathematics* I-113: 'However many rules you give me – I give a rule which justifies *my* employment of your rules.'

response to the rule: 'By definition, *he* is licensed to give, without further justification, the answer that strikes him as natural and inevitable'; and his confident inclination to feel that he has "cottoned-on" to the right rule 'is to be regarded as primitive' (ibid: 88, 91).

This, Kripke claims, is the significance of Wittgenstein's remark that 'I obey the rule *blindly*; which is to say no more than '*This is how it strikes me*' (PI §219). The idiolectic rule-follower has 'exhausted the justifications', 'has reached bedrock', and is only 'inclined to say: "This is simply what I do" (PI §217). Kripke claims that this point is made in the very first section of the *Investigations*, where we are asked to imagine a customer who walks into a shop and hands the shopkeeper a slip marked "five red apples", which leads to the following exchange:

"But how does he know where and how he is to look up the word 'red' and what he is to do with the word 'five'? —Well, I assume that he *acts* as I have described. Explanations come to an end somewhere (PI §1).

'But *are* the steps then *not* determined by the [arithmetical function]...?' a voice from the *Investigations* protests (PI §189). No doubt we confidently respond "125" upon being presented with our sum. But this confidence of appropriate understanding now appears, on reflection, to be misplaced. On Kripke's reading, Wittgenstein's own answer to this question is entirely negative: 'there is an important class of cases where a use of language properly has no justification other than the speaker's inclination on that occasion' (Kripke 1982: 74, fn. 63). The sceptic's arguments appear to entail 'that it is logically impossible (or at least that there is a considerable logical difficulty) for there to be a state of "meaning addition by 'plus'" at all' (ibid). My seemingly unshakeable confidence that in uttering "125" I am giving the correct response is, according to the paradoxical reasoning, based on nothing more than my 'simply following an unjustifiable impulse'; my grasp of the rule nothing more than 'a mere jack-in-the-box unjustified and arbitrary response' (ibid: 18, 23).

As Wright summarises the issue: ‘The hypothetico-deductive picture thus encourages us to accept as a matter of course the rider of certain knowledge of the character of one’s own understanding of an expression’ (Wright 2001b: 36)²¹. But the fundamental problem with such a picture of a ‘personal semantic contract’, Wright goes on to note, is that it provides no basis for the distinction, crucial for our ordinary notions of meaning, understanding, and objectivity between something’s merely seeming to one to be correct, on the one hand, and its actually being so, on the other (ibid: 31, 36). Recall, again, Wittgenstein’s remark: ‘One would like to say: whatever is going to seem right to me is right. And that only means that here we can’t talk about “right”’ (*PI* §258). “Correctness” simply collapses into whatever application I am disposed to make of a concept: I simply find myself ‘with a sincere disposition’ to apply a particular concept or expression in a new case, and, as Wright says, ‘that is the whole of the matter’ (Wright 2001b: 30).

Wright is candid about the fact that this argument constitutes a demand that we jettison the idea that there are ratification-independent facts, the recognition of which we are committed by determinate patterns of linguistic usage. The upshot of this position, given the rejection of ratification-independence, seems little different from Kripke’s Wittgenstein’s rejection of ‘the natural presupposition that meaningful declarative sentences must purport to correspond to facts’ (Kripke 1982: 78-79). Similarly, the rejection of ratification-independent patterns of linguistic usage seems to obliterate the possibility of using a word according to a determinate meaning: that is to say, Wright’s

²¹ The essential connection between the Platonic mythology and idiolectic understanding is brought out particularly clearly in the following reflection of Kripke’s:

Do I not, in elementary mathematics, grasp rules such as that for addition, which determines all future applications? Is it not in the very nature of such rules that, once I have grasped one, I have no future choice in its application...*And is not the grasping of a mathematical rule the solitary achievement of each mathematician independent of any interaction with a wider community.* True, others may have taught me the concept of addition, but they acted only as heuristic aides to an achievement – the “grasping of the concept” of addition – that puts me in a special relation to the addition function... *[I]n grasping a mathematical rule I have achieved something that depends only on my own inner state, and that is immune to Cartesian doubts about the entire external material world*’ (Kripke 1982: 79-80; emphases added).

position, in its essential relation to the paradox, seems little different from Kripke's "meaning scepticism" (or, more accurately: "conceptual nihilism").

Yet, this is not the conclusion drawn by Wright. The upshot of the argument is, according to him, that 'there cannot be such a thing as *first-personal privileged recognition* of the dictates of one's understanding of an expression, irrespective of whether or not that understanding can be shared' (Wright 2001b: 37; emphasis added). Although, as we have said, Wright believes we must excise the intuitive notion of ratification-independence from our conception of objectivity, this only matters, he concludes, when we consider an individual in splendid isolation from the rest of her linguistic fellows:

So long as the "recognition" [of a determinate, ratification-independent pattern] is unilateral, it cannot legitimately be claimed to be recognition, nor can that state of affairs of which it is supposed to be recognition be legitimately claimed to obtain (ibid: 38).

By making this move (which he also attributes to Wittgenstein), Wright raises the hope that an appeal to the wider community will serve to re-constitute conditions of correctness —required by any satisfactory account of meaning, understanding, and of objectivity— by picturing the dispositions of the linguistic community as constituting the standards by virtue of which individual linguistic dispositions may be judged. What is most significant, however, is that, although working on his reading independently of Wright, Kripke makes essentially the same move on Wittgenstein's behalf. Despite Kripke's suggestion that Wittgenstein accepts the implications of the "sceptical paradox" —all language is meaningless! —Kripke, like Wright, ultimately concludes that the implications of the paradox only apply to an individual considered entirely apart from his linguistic fellows:

If our considerations so far are correct...[then] if one person is considered in isolation, the notion of a rule as guiding the person who adopts it can have *no* substantive content... As long as we regard him as following a rule “privately”, so that we can pay attention to *his* justification conditions alone, all we can say is that he is licensed to follow the rule as it strikes him (Kripke 1982: 89).

And, as with Wright, this subtle movement from the ‘insane and intolerable conclusion’ that all language is meaningless to the position that the conclusion holds only for an individual considered in isolation, opens the door for Kripke’s appeal to the community as providing some way of making the required contrast between being right and only seeming to be right—a condition of correctness. This forms the basis of Kripke’s reading of the dialectic of Wittgenstein’s so-called “private language argument”. Recall that the conclusion of the “sceptical paradox”, which Kripke connects with *PI* §201, is that:

There can be no such thing as meaning anything by any word. Each new application we make is a leap in the dark; any present intention could be interpreted so as to accord with anything we may choose to do. So there can be neither accord, nor conflict (Kripke 1982: 55).

Notice that here the conclusion of the paradoxical reasoning is stated as being that the very idea of ‘meaning anything by any word’ is impossible; and the same claim is made for the notion of correctness-conditions: ‘there can be neither accord, nor conflict’. Notice also that this conclusion appears to hold for the very notion of language-use *tout court*, not only the unilateral linguistic behaviour of one individual alone; thus, the problem is stated in reference to the collective of language users: ‘Each new application *we* make is a leap in the dark’, because it ‘could be interpreted so as to accord with anything *we* may choose to do’ (emphases added). But the implications of the paradox undergo a subtle, but significant, alteration during Kripke’s reading of the “private language argument”:

Wittgenstein...states a sceptical paradox...[H]e accepts his own sceptical argument and offers a “sceptical solution” to overcome the appearance of paradox. His solution involves a sceptical interpretation of what is involved in such ordinary assertions as “Jones means addition by ‘+’”. The impossibility of private language emerges as a corollary of his sceptical solution of his own paradox...It turns out that the sceptical solution does not allow us to speak of a single individual, considered by himself and in isolation, as ever meaning anything’ (ibid: 68-69).

So now talk of the collective implications of the sceptical paradox – the applications ‘we make’, which can be interpreted ‘so as to accord with anything we may choose to do’ – have been abandoned in favour of the conclusion that the implications of the paradoxical reasoning only threatens the possibility of meaningful linguistic transactions *when the individual language-user is considered in isolation*. Moreover, it is not now language-use *tout court* (‘meaning anything by any word’), nor the very possibility of correctness-conditions (‘there can be neither accord, nor conflict’), that is seemingly rendered problematic. Now, rather, the problematic linguistic utterances are limited to attributions of determinate meaning to individual speakers, and Kripke refers to ‘Wittgenstein’s sceptical conclusion: no facts, no truth-conditions, correspond to statements such as “Jones means addition by ‘+’”’ (ibid: 77). This subtle shift from the conclusion of the “sceptical paradox” – all language is meaningless! – to the non-factuality of ascriptions of determinate meaning to individual speakers (“sceptical conclusion”) is never really argued for, and, as the collective reference in the initial formulation of the “sceptical paradox” attests, it does not necessarily follow from the paradoxical chain of reasoning, which appeared to render the very possibility of behaviour – whether individual or collective - undertaken according to a condition of correctness as incoherent²². Of course, if “all language is meaningless”, then, no doubt, attributions of meaning to individual speakers (“Jones means addition by ‘+’”) are thereby

²² Thus, as we saw earlier, Kripke acknowledges that the paradox threatens ‘the natural presupposition that *meaningful declarative sentences* must purport to correspond to facts’ – and not that non-factuality applies to ascriptions of determinate meaning to the utterances of speakers only (Kripke 1982: 78-79; emphasis added).

rendered meaningless —but, then, and as Kripke elsewhere recognises, so are all other forms of language!

As we shall consider in Chapter 2, the main point of the “sceptical solution” which constitutes Kripke’s Wittgenstein’s “private language argument” is that, although attributions of determinate meaning to individual speakers are supposed to lack factual content (the “sceptical conclusion”), nevertheless, utterances of the form “Jones means addition by ‘+’” still play a role in our linguistic transactions, and are, it is supposed, governed by - if not truth-conditions - then by “conditions of assertability”. These conditions, according to Kripke, make essential reference to the linguistic behaviour of the community as a whole. This move underlies Kripke’s “community appeal” whereby the collective behaviour of the linguistic crowd somehow underwrites the legitimacy conditions according to which the linguistic performance of an individual speaker may be judged. I say “somehow” underwrites conditions of legitimacy as, unlike Wright’s early crude Communitarianism, Kripke does not imagine the behaviour of the linguistic community as *constituting* the standards by virtue of which an individual speaker may be judged. As I will argue, however, such differences are irrelevant as Kripke’s community appeal, no less than Wright’s, fails to escape the devastating implications of the paradox. But that is to look ahead to Chapter 2, whereas now it remains to conclude the discussion of this present chapter.

Conclusion

Wittgenstein raises an apparent problem that can be repeated over all and any behaviour which manifests understanding of some form of correctness-condition. The problem appears to impugn the very possibility of acting in accordance with the dictates of a rule, or of using words according to determinate sense or meaning; and seemingly renders incoherent the very notion of believing, wishing, or desiring something specific. This is because any written or spoken expression of a word or rule, considered as a mere inscription or vocalisation, can be interpreted in any way we like; just as any mental item, according to this chain of reasoning, can be ‘interpreted so as to accord with anything we may choose to do’ (Kripke 1982: 55). At *PI* §201, Wittgenstein summarises the upshot of this paradoxical chain of reasoning as seemingly impugning the very possibility of conditions of correctness: ‘there would be neither accord nor conflict’. In the very next sentence, however, Wittgenstein clearly rejects the paradoxical chain of reasoning as based on a misunderstanding.

Kripke wishes to highlight both the generality and the constitutive, or normative, nature of Wittgenstein’s paradox. This is one blessing of Kripke’s reading. A further blessing is that Kripke highlights that the normative nature of the problem dooms any attempt to avoid the paradox which makes appeal to linguistic dispositions. This is because, as Boghossian makes clear, such responses are incapable of simultaneously satisfying both requirements which any adequate dispositional theory of meaning must satisfy – to specify only those dispositions which are meaning-determining, and to do so in strictly non-semantic and non-intentional terms. Either the theory ends up presupposing what it was meant to explain (knowing what “plus” means), or the very notion of determinate meaning vanishes altogether in face of a disjunctive infinity of interpretations —the very consequence the dispositional theory was meant to enable us to avoid.

Despite such blessings, however, Kripke’s treatment of the problem is, at the same time, cursed by a tendency to obscure the constitutive nature of the problem by cloaking its presentation in “transtemporal” form. The reason Kripke presents the problem in this manner, despite elsewhere highlighting its constitutive nature, is, of course, that presenting the “sceptical” paradox “straight on”, as it were (that is, without the

intellectual acrobatics connecting previous linguistic intentions with present performance) simply reveals, on its face, the absurdly self-refuting implications of the paradox: “all language is meaningless!” In order to communicate the problem, or even ‘to converse with me at all’, as Kripke acknowledges, the sceptic and I ‘must have a common language’ (Kripke 1982: 11-12). Kripke confides that he had made repeated attempts to formulate ‘a precise statement of the problems and conclusions’: ‘Although one has a strong sense that there is a problem, a rigorous statement of it is difficult’ (ibid: 5). Kripke thus initiates his presentation of the problem by focusing on *previous* linguistic intentions in order to make it plausible that the sceptical scenario can so much as be expressed, given that the sceptic must use words in order to doubt that words have any meaning:

If we are querying the meaning of the word “plus”, how can we use it (and variants, like “quus”) at the same time? So I suppose that the sceptic assumes that he and I agree in our *present* uses of the word “plus”: we both use it to denote addition. He does *not* – at least initially – deny or doubt that addition is a genuine function, defined on all pairs of integers, nor does he deny that we can speak of it. Rather he asks why I now believe that by “plus” in the *past*, I meant addition rather than quaddition’ (ibid: 12).

As we saw earlier (see, especially, footnote 4), and as is implied in the passage above (“at least initially...”), later in Kripke’s presentation the sceptic anyway retracts the initial immunity here granted to the legitimacy of present meaning: the sceptical paradox is recognised as impugning the meaning of ‘any...word at any time’ (ibid: 21). In addition, the initial immunity granted to the addition function is also retracted (only the *metalinguistic* certainty that the word “plus” refers to the addition function, and not the legitimacy of the function itself, is doubted by the sceptic— “at least initially...”). Kripke’s presentation of the problem in epistemological, sceptical garb, then, merely detracts attention from what he elsewhere identifies as the crucial issue: the possibility of understanding conditions of correctness (the “Kantian flavor”). Unnecessary mental gymnastics aside, however, Kripke elsewhere admits that Wittgenstein ‘might simply ask

“How do I know that I should respond ‘125’ to the query ‘68+57’?” or “How do I know that ‘68+57’ comes out 125?” (ibid: 12-13).

Perhaps the most notorious feature of Kripke’s treatment of the apparent problem is how vividly he makes the ‘insane and intolerable’ implications of the constitutive paradox (misleadingly termed “sceptical”). Precisely because the problem is constitutive, should the paradoxical reasoning be left to stand, its implications would indeed devastate all language, all concepts, all mental notions, etc.—any behaviour at all manifesting understanding of a condition of correctness. That the paradoxical reasoning leads to such evidently self-refuting and absurd implications is precisely the reason Wittgenstein rejects the paradoxical train of thought: ‘It can be seen that there is a misunderstanding here from the mere fact that in the course of our argument we give one interpretation after another; as if each one contented us at least for a moment, until we thought of another standing behind it’ (*PI* §201). ‘What this shews’, Wittgenstein concludes, ‘is that there is a way of grasping a rule which is *not* an *interpretation*’ (ibid).

The significance of these considerations, and the import of Wittgenstein’s rejection of the paradox, will be considered in Chapter 3. For the time being, however, it is enough to note that Kripke, by virtue of his interpretation of the “private language” dialectic, actually obscures what he earlier correctly accentuated—the ‘insane and intolerable’ implications of the paradox. As we have continually seen, linguistic behaviour, considered as mere behaviour, can be interpreted in any way we please. And any further interpretation which we may be tempted to appeal to in an attempt to fix any one initial interpretation is itself open to a seemingly infinite expanse of possible interpretations. This way, vicious regress threatens. Once we have embarked on this train of thought, one particularly tempting way to fix a determinate rule or interpretation, but in such a way that the threat of regress is avoided, is to imagine an interpretation which is somehow itself invulnerable to further interpretation. ‘What one wants to say’, as Wittgenstein remarks in the *Blue Book*, ‘is: “Every sign is capable of interpretation but the *meaning* mustn’t be capable of interpretation. It is the last interpretation”’ (Wittgenstein 1969: 34).

Once such a picture is accepted, it leads inevitably to the idea, as Wright remarks, ‘that each of us has some sort of privileged access to the character of her own understanding of an expression’ (Wright 2001b: 35-6). But the fundamental problem with the picture of such a “personal semantic contract” (‘I know what rule / am following, and what it requires *me* to do’: *ibid*: 22) is that it fails to uphold the required distinction between something only seeming to me to be correct, on the one hand, from its actually being so, on the other: ‘One would like to say: whatever is going to seem right to me is right. And that only means that here we can’t talk about “right”’ (*PI* §258). The very notion of “correctness” simply collapses into whatever application I am disposed to make in response to the expression of a rule: ‘my grasp of the rule nothing more than ‘a mere jack-in-the-box unjustified and arbitrary response’ (Kripke 1982: 23). Such considerations encourage both Kripke and Wright to distract attention from the ‘incredible and self-defeating’ conclusion of the paradox – all language is meaningless! – by claiming that the devastating implications of the paradoxical train of thought applies only to individual rule-followers, considered in splendid isolation. The hope kindled by this move is that some kind of appeal to the wider linguistic community can somehow underwrite the required standards of correctness which stand judgement over unilateral linguistic performance. Yet, as will become apparent in Chapter 2, to which it is now time to turn, such hope of avoiding the paradoxical implications —and of thereby preserving the possibility of behaviour appraisable according to conditions of correctness— is forlorn.

Chapter 2 From Conceptual Nihilism to Superlative Mythology (and back again...)

'However many rules you give me—I give a rule which justifies my employment of your rules...'

Remarks on the Foundations of Mathematics: I, §113

'If I say of myself that it is only from my own case that I know what the word "pain" means—must I not say the same of other people too?'

Philosophical Investigations: §293

'Does this mean, e.g., that the definition of "same" would be this: same is what all or most human beings with one voice take for the same?—Of course not.'

Remarks on the Foundations of Mathematics: VII, §40

Introduction

In the previous chapter, I tried to bring out the constitutive and general nature of the paradox which Kripke teases out of Wittgenstein's discussion of rules and of rule-following. And this led to the dilemma delineated at *PI* §185-188: *either* semantic and mental concepts appear to vanish under the weight of infinite interpretability, on the one hand, *or* the relevant aspect of meaning and mind appear as mythological regress-stopping, superlative rules (invulnerable, "non-misinterpretable interpretations"), on the other. The aim of this chapter is quite simple. It is to show that neither Kripke's attempts to live with the paradox (the "sceptical solution") nor Wright's attempt to avoid its devastating implications (his Constructivist "straight solution") succeed in their aims. As I will argue, both responses, in different ways—and despite all the intellectual dexterity and ingenuity on display—continue to reverberate between the two horns of the dilemma. Despite Kripke's and Wright's explicit rejection of dispositionalist theories of

meaning in face of the paradox, both of their respective pictures present linguistic behaviour as, at bottom, essentially dispositional. And for so long as putatively linguistic behaviour is conceived as essentially dispositional propensities to react, then such pictures remain continually haunted by the mythology of ultra-objective superlative concepts, or the fantasy of rules as rails extending to infinity. Thus, at the end of the chapter, we will be no further away from Wittgenstein's dilemma, with its attendant sense of impenetrable mystery on behalf of our mental and linguistic lives.

In §1 we will consider Kripke's brand of non-constitutive Communitarianism. The hope that a broadly Communitarian appeal could come to our rescue was kindled, as we saw at the end of Chapter 1, by Wright's and Kripke's shared move of taking the conclusion of Wittgenstein's paradox to apply only to individual speakers, considered alone. Kripke's Wittgenstein, as we have already noted, aims to avoid the absurd and self-refuting implications of the "sceptical paradox"—all language is meaningless!—by accepting, instead, the watered-down "sceptical conclusion", which claims that ascriptions of meaning to speakers ("Jones means addition by '+'") lack factual content. Now, 'if this is our framework,' Kripke contends, then 'we can only conclude that sentences attributing meaning and intention nevertheless have a role or utility in our lives': ascriptions of meaning operate according to "conditions of assertability", which involve reference to the wider linguistic community (Kripke 1982: 75). In this manner, it is supposed, whilst ascriptions of meaning and intention are deemed to be non-factual (the "sceptical conclusion"), it can nevertheless be claimed that they do not lack a linguistic function (the "sceptical solution").

Kripke offers a stark vision of linguistic life: individuals merely follow their brute inclinations to react in response to a word or rule. Kripke expends much intellectual dexterity in an effort to avoid presupposing shared linguistic agreement amongst these solitary babbling brutes: according to Kripke, appeal to recognition of agreement would fall foul of the sceptical challenge ("How does one know it isn't *quagreement*...?") But, as we shall see, "conditions of assertability" is a bit of a misnomer, given that there are, supposedly, no facts to be asserted about the meaning according to which a speaker, say,

is using a word. Rather, it is simply the case that an individual will not be (brutely) inclined to call another's utterance "incorrect" unless there is some kind of divergence between their respective dispositions to react to a word or rule. Despite the ingenuity on display in Kripke's attempt to sketch a decidedly *non-constitutive* Communitarianism, however, the "sceptical solution" still ends up presupposing what it was meant to avoid—recognition of *agreement* in the linguistic performance of speakers. That is to say, even Kripke's stark picture assumes that one can recognise whether the linguistic pattern another is following converges, or diverges, from one's own linguistic patterns. But this presupposes that one can recognise an instance of the same type of linguistic behaviour—which is what was meant to be explained to us in the first place. Therefore, we are returned once more to another instance of the familiar dilemma: either meaning vanishes altogether, or we are presupposing meaning in our purported explanations of meaningful behaviour.

More or less concurrently with Kripke's celebrated treatment of Wittgenstein, Crispin Wright was developing his own reading, and one which converged with that of Kripke in many respects. During Wright's early struggles with Wittgenstein's paradox, his favoured response was a constitutive brand of Communitarianism: the standards by virtue of which the linguistic inclinations of individuals may be judged are *constituted* by the aggregate inclinations of the linguistic crowd. Yet, the reason Kripke pulls his intellectual contortions in order to sketch a decidedly *non-constitutive* brand of Communitarianism is owing to the fact that constitutive Communitarianism so obviously presupposes the notion of *agreement* in linguistic reactions (a fate to which Kripke's efforts anyway eventually succumb—in order to form a community there needs be some kind of *sameness* in linguistic reaction: using "+" to mean addition, rather than quaddition, say).²³ Wright does eventually come to abandon his earlier crude Communitarianism, for, he recognises, the rejection of the possibility of "unilateral" recognition of a communal

²³ For more thorough objections to constitutive Communitarianism, see for instance: Blackburn (2002a), Boghossian 1989: 534-536, Hattiangadi 2007: 144-149. To paraphrase an argument Blackburn makes: if standards of correctness have been evacuated from the linguistic lives of individuals, then what use is calling on a whole crowd of such creatures?

mistake...is a definite error' (2001: 7).²⁴ He acknowledges that: 'Consensus cannot constitute the requirements of a rule, because we do, on occasion, actually *make use* of the notion of a consensus based on ignorance or mistake' (2001e: 168). Furthermore, Wright comes to admit that Wittgenstein does not endorse Communitarianism as a response to the paradox (2001f: 187-188), citing the following remark, where Wittgenstein explicitly disavows such a strategy: 'Does this mean, e.g., that the definition of the same would be this: same is what all or most human beings...take for the same? – Of course not' (RFM: VII, §40). One could also cite *On Certainty* §2: 'From its *seeming* to me—or to everyone—to be so, it doesn't follow that it *is* so.' Or, one could add, and to borrow Dewi Phillips's paraphrase of Flannery O'Connor: 'whether one follows the herd should depend on what the herd is doing' (Phillips 2001: 240).

The dilemma for Wright, however, is that, as he candidly admits: 'there only seem to be these two options—on the one hand, some form of community appeal, or, on the other 'the Platonist line' (ibid: 5). In response to the failure of his earlier crude Communitarianism, Wright, as we shall see in §2, develops a more sophisticated, two-pronged Constructivist response to the paradox. The "third-personal" branch of Wright's Constructivism concerns the reality, or metaphysics of rules, whilst the "first-personal" branch is concerned with epistemological issues around following according to a rule.

Following according to a determinate rule is, according to Wright's "third-personal" prong, constituted by *activity*— activity which is conceived, as we shall see in §2a, as propelled by "subrational propensities", or "classificatory dispositions". Yet this activity is also communally shared: thus our rule-following practices require an institutional framework: our subrational nature, Wright contends, work 'courtesy of a co-operative *institutional* environment. There has to *be* such a thing as adding correctly before any physiological condition can constitute the ability to add' (2001f: 155). Unlike Wright's

²⁴ In order to keep repetition to a minimum, for the duration of this chapter, but for this chapter only, I will refer in this form to papers from Wright's collection of published works on the topic of Wittgenstein on following a rule and of private language, *Rails to Infinity*. I will continue to use the regular Harvard system of reference for all other commentators.

earlier response to the paradox, however, such communal institutions do not *constitute* the standards by virtue of which individual behaviour is to be judged. Rather, the social institutions are *somehow* constituted by the “subrational propensities” of individual practitioners. This picture now looks suspiciously similar to that of Kripke: individuals simply indulging their own brute inclinations to react, and the “community” as merely an aggregate of such babbling brutes. Therefore, Wright acknowledges that, on this picture, there is inevitably a drift towards ‘the idea that the requirements of a rule, in any particular case, are simply *whatever we take them to be*’ (2001f: 167; original emphasis). Showing how this can be avoided is the hope of future constructive endeavour (which endeavour is passed over by Wright’s Wittgenstein who, Wright implies, is adept at arguing for what does *not* constitute the requirement of a rule—*not* communal agreement, *not* an autonomous superlative rule—but more reticent at explaining what *does* constitute such requirements).

Yet, as I will argue in §2a, any such hope for future explanatory analysis seems unlikely of fulfilment, however, given that the appeal to “subrational propensities” takes us, once more, back to the dilemma with which we began. *Either* such items are conceived as merely brute, or purely physical, propensities: but then the determinacy of meaning and mental content becomes unavailable once more, as such states or processes could only sustain contingent connections with appropriate applications of a word or rule, or performances which accord with the identity of an intention. The regress of interpretations then looms large, and with it, conceptual nihilism. *Or*: such propensities are conceived as inherently classificatory, or meaning-determining, in which case our explanatory ambitions fall under the weight of presupposition. The ubiquity of this dilemma is then imported into our very terminology: thus we are introduced to ambiguous terms such as “subrational propensities” and “classificatory dispositions”. Moreover, rather than dispel the mystery, we simply end up transferring the sense of queerness: thus “How is it possible for people to use words according to a determinate meaning, or form intentions with a specific content?” effectively becomes the arguably more queer: “How is it possible that subrational propensities are able to perform such mental and linguistic feats?”

The “third-personal” branch of Wright’s response is, admirably, founded on an attempt to respect Wittgenstein’s assertion that the paradox is based on a misunderstanding: ‘an inclination to say: any action according to the rule is an interpretation’ (*PI* §201). As we shall discuss further in Chapter 3, however, the fundamental problem is that, rather than give due consideration to the nature of this misunderstanding, Wright, instead, begins thinking about the positive substance to the obverse of the misunderstanding: ‘there is a way of grasping a rule which is *not* an *interpretation*’ (*PI* §201). This failure to give due weight to Wittgenstein’s misunderstanding leads Wright to make a fateful assumption in his reading of *PI* §201, an assumption which embodies a form of the very misunderstanding Wittgenstein is concerned to warn us against: Wright reads Wittgenstein as ‘allowing, apparently, that there is indeed an ineliminable multiplicity of conflicting possible interpretations’ of the expression of a rule (2001d: 124). The implication of this fateful assumption is that Wright thus identifies rationality itself to be a species of interpretation. Thus, when Wright turns to consider a positive account of the obverse of the misunderstanding diagnosed by Wittgenstein—that ‘there is a way of grasping a rule which is *not* an *interpretation*’ (*PI* §201)—the substance Wright reads into this is that ‘something other than the exercise of interpretative (rational) faculties enters into the capacity to “read” another’s behaviour’ (2001e: 124). And this *something other* is, according to Wright, “subrational propensities” or “classificatory dispositions” which supposedly guide our linguistic and mental behaviour.

The “first-personal” (epistemological) prong of Wright’s reading of Wittgenstein starts off in similar fashion to the first, in that we are led once more into a particular version of Wittgenstein’s misunderstanding: the “inclination to say: any action according to the rule is an interpretation”. As we shall see in §2b(i), Wright justifiably objects to the implicit reductionist prejudice of Kripke’s sceptic—that we cannot simply directly cite the meaning-constituting fact (“I meant, or mean addition by ‘plus’”), but that the response must be essentially inferential. Despite pressing his face right up to Wittgenstein’s misunderstanding in this way, however, Wright nevertheless assumes that any response to the sceptic must still be some species of *interpretation*: that is to say, ‘the substitution of one expression of the rule for another’ (*PI* §201). Not unlike those challenged by the sceptic, one cannot directly cite the fact that one meant addition by “plus”: rather, Wright

asserts, ‘the challenge is to recall some *independently characterised fact*’, on pain of begging-the question against the sceptic (2001g: 176). And the *something other* with which we need to rebut the sceptic is, Wright suggests, facts about intentional mental states, particularly intention itself (“I intended to use “plus” to denote addition”).

What makes appeal to intentional mental states seem promising is that they share crucial features in common with linguistic meaning: avowals of intention are, like self-knowledge of the meaning of one’s own words, *immediate* or *non-inferential*. Moreover, no less than the meaning of a word, which one may call upon to apply in an indefinite amount of situations, intentions are “infinitely fecund”, as Wright puts it. Yet, although such parallels between meaning and intention provide hope for some non-question-begging direct response to the sceptic, the very same similarity casts a dark shadow over any such promise—for, as we saw in Chapter 1, the same paradoxical chain of reasoning can be brought to bear against the content of intentional mental states, no less than against the determinacy of meaning. Most crucially, despite the fact that intentions are “non-inferential” (that is to say, immediate), appeals to intentions in order to rescue the determinacy of linguistic meaning are, nevertheless, still *interpretational*, in the sense described above.

In preparation for his proposed direct rebuttal of the sceptic, Wright aims to clear away some deep puzzles connected with intentional mental concepts. Any minimally adequate account of intentional mental states must account for their authoritative (first-personal) avowability, together with the fact that such concepts are constitutively connected with events, states of affairs, or behaviour which are determined by their content. The trouble for Wright, however, is that the relevant mental concepts seem to hover “puzzlingly and unstably” between two epistemological paradigms: knowledge by observation and knowledge by inference. On the one hand, we are drawn towards the conviction that each epistemological paradigm can account for the authoritative avowability and the constitutive connectivity of intentional mental concepts, and also, *at the very same time*, our thought is drawn to the conviction that neither epistemological paradigm is able to account for this pair of essential features. This ostensibly unavoidable antinomic tension

leads Wright, and his Wittgenstein, to abandon the idea that intentional mental states enjoy existence independent of our practices of avowal and, consequently, to abandon the idea that there are intentional mental states, properly speaking, to be made available to cognitive awareness.

But as I will argue in §2b(ii), the real source of the difficulties is not so much the antinomic tension, but the two putative paradigms of the mental realm upon which Wright's reflections are based (and which deform Wright's view of the respective paradigms of epistemological access). On the one hand, Wright conceives of a dispositional-behavioural, or physicalist paradigm of the mental, whose avatar are character traits which are constitutively connected with certain patterns of behaviour: thus *courage*, say, is constitutively connected with courageous behaviour. On the other hand, Wright distinguishes a putative paradigm of the mental whose avatar are occurrent mental states.

Wright claims that, in attempting to explain intentional mental concepts we are drawn to the inferential paradigm of the mental, owing to the fact that dispositions are supposedly able to account for the constitutive connectivity of intentional states. But, as I will argue in §2b(ii), Wright cannot unproblematically help himself to the undeniable constitutive connections between, say, *courage* and courageous behaviour in marking off this putative paradigm of the mental. This is because Wright's characterisation of such character traits itself "hovers puzzlingly and unstably": either such dispositional characteristics are conceived as merely brute, or purely physical in which case they can only sustain *contingent* connections with the appropriate behaviour; or, our explanatory ambitions are deflated by having to presuppose the kinds of constitutive connection we were hoping to explain. Thus, we are thrown once more back into another form of the familiar dilemma: either the constitutive connection is lost, with seemingly no hope of its retrieval (and we are on the road again to conceptual nihilism); or, we are led to posit queer *quasi-physical facts* which, unlike ordinary physical states, are special *constitutive-connection-bearing* or *meaning-determining* physical states (therefore: either our explanations, were they forthcoming, are not scientific, or we are implicitly recommending a radically

revisionist conception of physical facts). As we shall see, this ubiquitous dilemma is, again, reflected in terminology which is ambiguous: “the dispositional psychological state”. Therefore, Wright’s claim that we are drawn to the epistemological paradigm of inference in order to explain constitutive connections does not hold water, and so this loosens one of the threads of antinomic tension.

As we will discuss further in §2b(ii), McDowell makes analogous arguments against Wright’s conception of the contrasting putative paradigm of the mental, that for which occurrent sensations provide the avatar. The fundamental problem with this putative paradigm is that Wright conceives occurrent mental states as “pure” or “genuine” states: those which, according to Wright, make ‘no demands upon the conceptual resources of the subject’ (2001i: 298). McDowell casts doubt on any Constructivist sympathies Wittgenstein may have by pointing out that the *Investigations* contains clear instances where Wittgenstein unproblematically registers that intentional occurrences (a mental image of something red, for instance) can be available to conscious awareness. McDowell also brings Wittgensteinian argument to bear against Wright’s conception of “pure” or “genuine” occurrent states, which fails to preserve the essential constitutive connection between the concept *pain*, say, and expressions of pain, or pain behaviour. Thus, the argument of Wright’s Wittgenstein for the Constructivist conclusion: if an intention were purely an occurrence in consciousness, then it could only be contingently connected to subsequent performance. But, intentions enjoy *constitutive* connections with subsequent performance; therefore, intentions cannot be occurrences in consciousness. But this argument only follows from the starting conception of “pure” mental occurrences, which are not already constitutively connected with their expression (pain and pain behaviour, for instance). But once such a picture of occurrent mental states has been revealed as the villain of the piece, then the ground has been removed from the radical Constructivist conclusion, with its implication that, as McDowell points out, talk of being struck by a thought, or grasping meaning “in a flash” is incoherent. So, once more, the antinomic tension begins to relax once the putative paradigms of the mental on which it is based are revealed as the real source of the problem.

Ending finally, in §2c, with Wright's positive Constructivist proposals, I will show that we are, once more, returned to familiar dilemmas. Despite the fact that both Kripke and Wright intend to distance themselves from dispositionalist theories of semantic and mental content, their respective pictures present, at bottom, individuals reacting to rules in whatever way their brute dispositions, or their subrational propensities incline them. Thus Wright speaks of "deep contingencies", such that each of us is 'ceaselessly, but subcognitively moved'—'who knows exactly how...' (2001i: 313; 2001f: 140). Again, this fits Kripke's picture, where the individual's inclinations to react are 'to be regarded as primitive' (Kripke 1982: 91). In both Kripke's and Wright's respective pictures, then, communal agreement is seemingly perched on the edge of a frighteningly precarious knife-edge, depending as it does on the sheer contingency that the brute inclinations of each individual continue to march in step. And, once again, not unlike Kripke's "community," who allow individual speakers provisional membership until such a time as their brute inclinations begin to diverge from theirs, individual avowers of intention in Wright's picture are granted "deference" to their own self-ascribing inclinations until such time as they are given reasonable grounds to retract such default concessions. But such reasonable grounds are said to be such things as whether the instinctive self-ascriber goes on to *do what she said she intends to do*, and so we are once more presupposing what we set out to explain (including concepts such as "reason", "understanding", "standard of correctness", etc.). So again, once more: either brute dispositional propensities (and so threat from regress); or presupposition of what we were hoping to explain. The problem for Wright, however, is that, as we shall see in §2c, he can only conceive occurrent phenomena of consciousness in superlative, mythological fashion, such that rejection of this picture is *eo ipso* rejection of the idea that occurrent mental phenomena can be made available to self-conscious awareness.

Throughout this present chapter we will be confronted by different instances of the same general form of dilemma (delineated by Wittgenstein at *PI* §§185-188). On the one hand, as Wright protests, the 'price of objective meaning' seems to be 'an *absolute conception* of truth: a conception absolved from all practical controls' (2001c: 77). But, on the other hand, as Kripke articulates, the vehicles of language 'are visible (or audible or palpable), concrete phenomena—marks or diagrams on paper' (Kripke 1982: 106). But conceived

merely as such, then such concrete phenomena are seemingly open to infinite interpretation. And so, on pain of lapsing into Platonistic fantasy, we are thrown once more into the threat of a regress of interpretations, and its attendant conceptual nihilism.

§1 Kripke's Community Appeal

As we saw in the first chapter, Wittgenstein, according to Kripke, rejects 'the natural presupposition that declarative sentences must purport to correspond to facts' (ibid: 77-78). The grounds for such rejection is that the paradox has apparently unmasked as incoherent the notion of 'a natural relation of interpretation between a thought in someone's mind and the "fact" it "depicts"' (ibid: 85). Given that the paradoxical train of thought seemingly forces us to conceive any such relation of interpretation in a Platonistic, "superlative" manner, then we are led to conclude that declarative sentences cannot pick out determinate facts. Consistently, this should commit Kripke's Wittgenstein to rejection of the idea that, say, the assertion "the cat is on the mat" can be said to pick out a determine state of affairs—*the cat's being on the mat*. But, as we have already seen, this is not something Kripke's Wittgenstein commits himself to. Instead, it is claimed that the paradoxical reasoning only serves to impugn *attributions of meaning and intention* to speakers, as distinct from other forms of linguistic expression. This is one move made by Kripke's Wittgenstein in an attempt to wriggle free from the devastating implications of the paradox.

Passing over the fact that it is not entirely clear on what basis this distinction between attributions of meaning, on the one hand, and all other forms of linguistic expression, on the other, has been made (over and above the hope that it will enable us to avoid the implications of the paradox), the next move Kripke makes on Wittgenstein's behalf is to claim that, despite the fact that attributions of meaning —curiously singled out from in amongst all other forms of linguistic expression— lack factual content, nevertheless such specific forms of assertion are not devoid of function within our linguistic lives. According to Kripke, if we are to speak properly of such utterances ("Jones means addition by 'plus'") we should not think about what must be the case for the sentences to be true but, rather, give due consideration to their "conditions of assertability": 'conditions when a move (a form of linguistic expression) is to be made in the "language game"' of ascribing meaning (ibid: 74). 'All that is needed to legitimize assertions that someone means something', Kripke's Wittgenstein claims, 'is that there be roughly specifiable circumstances under which they are legitimately assertable' (ibid: 77-78).

Leaving aside for the time being the fact that this picture, by invoking the notion of *legitimacy* of assertion, seemingly trades in normative notions—that there are conditions of correctness pertaining to the *appropriate* circumstances of assertion—Kripke claims that, once such conditions are specified, we must pay attention to the *role* or *utility* in our lives of our practices of making such assertions (and their denial) under these conditions: ‘granted that our language game permits a certain “move” (assertion) under certain specifiable conditions’, we must ask, ‘what is the role in our lives of such permission?’ (ibid: 75). According to Kripke’s Wittgenstein, it transpires that the role of such assertions and their conditions ‘involve reference to a community. They are inapplicable to a single person considered in isolation’ (ibid: 79). This move, of course, is made available by the concurrent move we saw Kripke’s Wittgenstein make in the previous chapter, where it is claimed that the implications of the paradox only apply to an individual considered in isolation. But now, armed with conditions of communal assertability, Kripke’s Wittgenstein can nevertheless claim that to ‘use a word without justification does not mean to use it without right’ (*PI* §289).

‘It is part of our language game of speaking of rules’, Kripke’s Wittgenstein contends, that an individual speaker will ‘follow his own confident inclination’ that a certain way of acting (responding “125”, rather than “5”) ‘is the *right* way to respond’; that is to say, ‘the “assertability conditions” that license an individual to say that, on a given occasion, he ought to follow his rule this way rather than that are, ultimately, that he does what he is inclined to do’ (ibid: 87-8). It is a further consequence of the sceptical conclusion, however, that for so long as we consider such a speaker in isolation, this entirely exhausts what we can say about his linguistic usage. We can say that he acts unhesitatingly, and without further justification, when he responds “125”, rather than “5”, and that *he* is confident that this is the “right” way to respond (“I mean addition by ‘plus’”). However, there is no scope to explain this normatively; that is, to explain his use on the basis of conditionals such as “if he means addition by ‘plus’, the answer to ‘68 + 57’ *must*, or *should*, be ‘125’”. When an individual speaker is considered in isolation, such conditionals, and the concomitant notion of one’s linguistic behaviour being in *accord* with the authority of a determinate rule, ‘can have *no* substantive content’ (ibid: 89). This is because, as we have seen, there is no way to preserve the required distinction

between being right, and only seeming to be right—conditions of correctness: ‘Nothing justifies a brute inclination to answer one way rather than another’ (ibid: 15).

Things look different, it is claimed, once Kripke’s Wittgenstein has made his “community appeal”. Jones’s disposition to give particular answers to particular computations, and his general inclination to feel confident that he has “got it”, are, according to Kripke, ‘to be regarded as primitive’ (ibid: 91). Nevertheless, ‘Smith need *not* accept Jones’s authority on these matters: *Smith* will judge Jones to mean addition by “plus” only if he judges that Jones’s answers to particular addition problems agree with those *he* is inclined to give’ (ibid: 91). No doubt, if the answers to particular addition problems that Jones is inclined to give only occasionally differ from those which Smith is inclined to give, then Smith is entitled to interpret Jones as, at least, still following the “proper” procedure (Smith would interpret such occasional deviations, we may suppose, as “mistakes”). However, if Jones consistently fails to give responses which agree with Smith’s own (or if Jones is not inclined to give the same answers for sums involving very small numbers; or if a deviation is so wildly out of kilter with the response Smith is inclined to give, so that it would be difficult to class these as “mistakes” - Jones answers “5” when asked to add “68 + 57”, say), then Smith will be justified in judging that Jones does not mean addition by “+”. Even supposing that Jones did mean addition by “+” in the past, the evidence of such present deviations ‘will justify Smith in judging that he has lapsed’ (ibid.). (Presumably, in such cases of systematic deviation, given Jones’s assumed confidence that he means addition by “plus”, together with his awareness that Smith’s answer deviates from his own, Jones will also think himself justified in judging that *Smith*, given his answer “125”, does not mean addition by “+”).

We say of somebody that he is following a rule if, in enough cases, it is claimed, his brute inclinations to respond in a particular way accord with our own brute inclinations, and deny it of him when they do not. These, it is claimed, are the conditions which justify utterances such as “Jones means addition by ‘+’”. At the level of a particular linguistic community, then, the “language game” of ascribing determinate meaning to speakers serves to demarcate those individuals whose linguistic inclinations have so far been in

convergence with the rest of a particular community from those speakers whose inclinations have not. Such ascriptions, on this view, serve to (provisionally – more on this in a moment) admit those individuals who are disposed to give a certain answer to a certain addition problem as a member of the community whose other members share the same inclinations:

[A]n individual who passes such tests in enough other cases is admitted as a normal speaker of the language and member of the community. Those who deviate are corrected and told (usually as children) that they have not grasped the concept of addition. One who is not an incorrigible deviant in enough respects simply cannot participate in the life of the community and in communication (ibid: 92).

Given these conditions of utterance, then, the role or utility of the practice of ascribing determinate meaning is that they supposedly enable us to make discriminations between those speakers who have shown that they can be trusted to use “+” in the same way that we do, from those speakers who cannot be so trusted. If Smith goes into a shop and makes a request to Jones the shopkeeper which necessitates Jones’s performance of an arithmetical computation, Smith will expect Jones to proceed by *adding* not quadding. If Smith does not expect that he will do this —maybe he has been tipped-off by Evans that “Jones does not mean addition by ‘+’” (meaning Jones has been known to deviate unacceptably from the addition function on too many occasions in the past)— then Smith will likely not entrust the computation to him (and may well be brutally inclined to take his custom elsewhere!)

In its most general sense, then, the idea is that ascriptions of meaning serve to discriminate between those who are trustworthy members of our particular linguistic community from those who are not: ‘Our entire lives depend on countless [linguistic] interactions, and the “game” of attributing to others the meaning of certain concepts or rules, thereby showing that we expect them to behave as we do’ (ibid: 93). Those who

cannot be so trusted to behave appropriately (speakers deemed by the community to be following deviant rules) are thereby excluded from such transactions.

Now, the crucial point of these considerations is that conditions under which it is assertable that “Jones *believes* that he means addition by ‘+’” can be said to be distinguished from conditions under which it is assertable that “Jones actually does mean addition by ‘+’”. Jones may possess an unshakeable confidence that his behaviour accords with the rule for addition; however, the wider community may perceive that his use deviates from theirs²⁵, and so the conditions under which it is assertable that “Jones means addition by ‘+’” remain unfulfilled. The distinction between behaviour which merely seems to accord with a rule and behaviour which does in fact accord with the rule – conditions of correctness - is thus apparently underwritten by the wider linguistic community. Considered in this sense, assertability conditions supposedly allow others to discern the “incorrect” rule-following practices of a particular speaker, and, unlike instances where a subject is considered in isolation, ‘these will *not* be simply that the subject’s own authority is unconditionally to be accepted’ (ibid: 89).

It is crucial to understand that the claim here is not that communal agreement *constitutes* the normativity of a rule:

We cannot say that we all respond as we do to “68 + 57” *because* we all grasp the concept of addition in the same way, that we share common responses to particular addition problems *because* we share a common concept of addition (ibid: 97).

²⁵ But if there’s no such thing as unilateral recognition, how is this possible? Only via some kind of “group-think”?!

Suppose that no member of a particular linguistic community has ever before performed computations involving numbers higher than 57, but that the members are now called upon to perform sums involving greater numbers. It would then be open to the (constitutive, not epistemological) sceptic to ask how we know, in the event that these putative adders are called upon to enter such uncharted numerical waters, that the responses the speakers are inclined to give will be in agreement. *Ex hypothesi* no one has so far ventured into this numerical beyond. How do we know that there has so far been only *quagreement* between members of the community, which refers to the situation whereby individual speakers are disposed to give the same answers to sums involving numbers no higher than 57, but are each disposed to give different quus-like answers for sums involving larger numbers? In other words, as Kripke perceives, “Respond to an addition problem exactly as others do!” ...falls foul of Wittgenstein’s strictures on a “rule for interpreting a rule” (ibid: 146, fn. 87).

So, the situation is *not* that there are no facts about meaning when an individual speaker is considered in isolation, but that, when we bring a linguistic community into the picture, determinate meaning is *constituted* by facts of communal agreement, so that each individual would be subject to the normative authority of that socially-constituted meaning.²⁶ For Kripke’s Wittgenstein, even at the community level there are, it is supposed, no facts about meaning. The community plays a purely negative role. Thus:

What follows from these assertability conditions is *not* that the answer everyone gives to an addition problem is, by definition, the correct one, but rather the platitude that, if everyone agrees with a certain answer, then no one will feel justified in calling the answer wrong (ibid: 112).

According to this view, then, the communal agreement provides a necessary restriction on the ‘game of attributing to one of its members the grasping of a certain concept’ (ibid:

²⁶ This is the form Wright’s initial, crude community appeal takes, but which he later comes to abandon.

95). But, it is not that this “game” allows us to positively assert that a certain speaker is using a concept, such as addition, “correctly” (where “correct means something like “according to the standards of the community”). Remember, given acceptance of the “sceptical conclusion” there is no such thing as meaning to be correctly ascribed. Rather, it is simply the case that: ‘If Jones does *not* come out with “125” when asked about “68 + 57”, we cannot assert that he means addition by “+” (ibid.). (Or, more accurately, we will lack the brute inclination to do so). Furthermore, according to this proposed picture of meaning, the legitimacy of shared linguistic conventions—amongst those, at least, who continue to remain in step—appears to be a radical hostage of fortune to the contingent vagaries of the merely brute dispositions of individual speakers. It is just a matter of radical contingency that — ‘extreme cases of uneducability or insanity aside’ (Kripke 1982: 96) — our actual community does possess some general regularity in respect of the linguistic behaviour of its members. No doubt such agreement was fostered by the similar linguistic training that speakers received, but no appeal can be made to the authority of a rule when deviants are disposed to depart from the responses recommended by this training. Remember there is no room in this picture for shared commitments to the authority of anything outside the brute inclinations of individuals. Thus agreement amongst individual inclination of response is nothing but ‘a brute empirical fact,’ and we owe it to a radically fortuitous gift of fate that we are not ‘reduced to a babble of disagreement’ (ibid: 109, 91).

Given these considerations, it seems misleading to talk of *assertability* conditions at all. The situation is, rather, that until such time as the speaker begins to ‘exhibit a pattern of bizarre, quus-like behaviour’ the community merely deigns to take the speaker ‘*provisionally* into the community’ (Kripke 1982: 112; emphasis added). As soon as a speaker’s linguistic inclinations and dispositions begin to deviate unacceptably from the community’s own, people will feel inclined to highlight this discrepancy by saying that the deviant “doesn’t mean addition by ‘+’”, and will withdraw their provisional trust in his ability to engage in the community’s linguistic activities, from which he will, presumably, be banished. Until such an event, however, nobody will feel disposed to say this, and the speaker will be given the benefit of the doubt with regard to the community’s transactions.

This picture casts an almost sinister aspect to ascriptions of determinate meaning. The film *Donnie Brasco* (1997), based on the true story of FBI agent Joseph D. Pistone's undercover infiltration of the New York Mafia (under the adopted pseudonym of the film's title), contains a scene where Benjamin "Lefty" Ruggiero, a low-level career hit-man (played by Al Pacino), inducts his newly befriended young protégé "Donnie" (played by Johnny Depp) into the veiled subtleties of mob discourse. On their way to meet members of Lefty's Mafioso "family", the older man informs Donnie that he will introduce him as a "friend of mine". To the assembled mobsters, this will immediately identify Donnie as an "associate" (someone who has criminal connections with the Cosa Nostra, but who isn't a bone fide member of the "family", or "made guy"). A "made guy", in contrast, would be identified by his being introduced as a "friend of ours". On the picture of language attributed to Wittgenstein by Kripke, "Jones means addition by '+'" in the "language game" of ascribing meaning to a speaker seems uncomfortably close, to this reader at least, to "He's a friend of ours" in mob discourse.

It would be fair to call this a radically revisionist conception of language. Admirers of the philosopher who once declared that philosophy 'only states what everyone admits' (*PI* §599) would therefore be justified in rejecting Kripke's Wittgenstein as "no friend of ours." This proposed "solution" bends and twists in a strained attempt to salvage some conception of the normativity of meaning, but the notion we are left with is, at best, problematically thin. On this account there is no positive evidence which can be appealed to when interpreting someone's behaviour as being in accord with a particular way of responding to the expression of a rule, or of using words in a determinate fashion. There are no factual grounds for asserting that a determinate rule is being followed. Texts and speech are open to infinite interpretation. There is no such thing as correctly ascribing determinate meaning to an individual's utterances; and, so, we lack a notion of meaning to be correctly ascribed. Nor can we apply any normative standards to the avowed intentions of an individual, together with what she goes on to do. At the very best, according to Kripke's picture, all we can do is detect—but how, if there's no such thing as unilateral recognition of a condition of correctness?—that the brute inclinations of an individual do not match with our own brute inclinations: "Jones has used this word incorrectly" can amount to nothing more than a brute inclination to articulate such a

contingent incongruence — and “incorrectly” here really means: not as we do it (“He’s no friend of ours”).

The hope that some kind of community appeal could underwrite correctness-conditions was, as we have seen, kindled by the “sceptical conclusion,” which implies that the paradoxical implications apply only to solitary speakers. This then raised the possibility that a community appeal could underwrite conditions of (so-called) “assertability” for ascriptions of meaning. This is the so-called “private language” dialectic according to Kripke. Thus, if Jones is inclined to use the word “plus” in the same way that I am inclined to use it, then I will desist from judging that we are out of step. Even if we grant, as Kripke does on Wittgenstein’s behalf, that there cannot be any legitimate assertions to the effect that we are positively in step with each other, nevertheless, this picture still presupposes that I can unilaterally recognise that our respective linguistic dispositions in relation to the word are in agreement. Thus, as quoted earlier, Kripke states: ‘*Smith* will judge Jones to mean addition by “plus” only if *he judges* that Jones’s answers to particular addition problems *agree* with those *he* is inclined to give’ (ibid: 91; second and third emphases added). Yet the denial of just this possibility was the initial step which raised the hope of any community appeal. So it seems that the community appeal actually presupposes the possibility of unilateral recognition of a correctness-condition: ‘that there be roughly specifiable circumstances under which [assertions that someone means something] are *legitimately* assertible’ (ibid: 77-78; emphasis added). But this trades on a normative notion —precisely that which was meant to be explained²⁷.

Most crucial, however, is that Kripke acknowledges that this community appeal presupposes the notion of agreement: ‘We have seen that it is part of Wittgenstein’s *general* view of the workings of *all* our expressions to concepts that others can confirm whether a subject’s responses *agree* with their own’ (ibid: 100; final emphasis added). Yet the whole point of the intellectual twists ingeniously turned in order to rescue some

²⁷ Boghossian also recognises that ‘Kripke’s communitarian conditions are parasitic on the solitary conditions, and not the other way round’ (see his admirably pithy argument at Boghossian 1989: 522).

notion of correctness-conditions was initiated in face of the recognition that a constitutive brand of Communitarianism fails to avoid the sceptical challenge (recall: “How do we know it’s not quagreement...?”) Thus the “sceptical solution” is no solution at all and simply returns us to the dilemma it was meant to enable us to avoid: either we must pre-suppose the notion of agreement, or it vanishes altogether. At this point, Kripke protests:

What Wittgenstein is doing is describing the utility in our lives of a certain practice [i.e., non-factual attributions of meaning]. Necessarily he must give this description in our own language. As in the case of any such use of our language, a participant in another form of life might apply various terms in the description (such as “agreement”) in a non-standard “quus-like” way...This cannot be an objection to Wittgenstein’s solution unless he is to be prohibited from any use of language at all (ibid: 146).

But Kripke cannot simply help himself to the notion that language must be used in order to state the problem, and the proposed solution to it. Of course, that is no doubt true. But, as Kripke himself has admitted, what the “sceptical paradox” puts under threat is precisely the idea that *any* use of language is determinately meaningful: it’s putative conclusion is that ‘there is no fact about me that distinguishes my meaning a definite function by “+”...and my meaning nothing at all’ (ibid: 21). And so any proposed solution, if it is to be a solution at all, needs to disarm that threat. Thus, Kripke’s Wittgenstein cannot merely avail himself of the notion that, in stating his putative solution, his words mean something determinate. Of course, we have to grant somebody the ability to use language in order that he can state an attempted solution to the putative problem. That this assumption is not warranted in the case of someone who supposedly accepts the “sceptical conclusion,” however, should simply remind us of how “insane and intolerable” its implications are.

Summarising the position, Kripke asserts: ‘As against...a “Platonist” conception of the problem, Wittgenstein asks us that we...*look*...at the circumstances under which [certain] assertions are actually uttered, and at what roles such assertions play in our lives’ (ibid: 75)²⁸. But, as we have seen, this “sceptical solution” does not succeed. We are thus returned to the dilemma with which we began. As Kripke urges: ‘Platonic objects may be self-interpreting, or rather, they may need no interpretation; but ultimately there must be some mental entity involved that raises the sceptical problem’ (Kripke 1982: 54). The trouble is, given that this mental entity, whatever it may be, is —just like the expressions of a rule, considered merely as spoken sounds or written shapes— open to an infinity of possible interpretations, then, on pain of the possibility of meaning disappearing altogether, it can only show up as an ultra-objective, superlative rule, whose natural image in a mathematical context is an infinitely extending immovable rail, and which somehow serves to guide our linguistic behaviour. Kripke acknowledges that Wittgenstein agrees with the (constitutive) sceptic ‘that there is no “superlative fact” about my meaning addition by “plus” and which determines in advance what I should do to accord with my meaning’ (ibid: 65). In the context of mathematical proofs, Kripke acknowledges that Wittgenstein justifiably rejects the picture of ultra-objective and ‘abstract objects laid up in a mathematical heaven’; ‘they are’, rather, ‘visible (or audible or palpable), concrete phenomena – marks or diagrams on paper’ (ibid: 106). But, as we have continually seen, such audible or palpable concrete phenomena (spoken sounds or written shapes)—considered merely as such—are open to an indefinite number of possible interpretations. And so we now seem to be thrown back once more into conceptual nihilism: ‘it is logically impossible (or at least there is a considerable logical difficulty) for there to be a state of “meaning addition by ‘plus’” at all’ (ibid: 53). And so, despite all the philosophical ingenuity and dexterity of thought displayed in the “sceptical solution”, we are no further than before: either meaning appears like an occult mental feat, or it disappears altogether!

²⁸ As Boghossian highlights, given that ‘assertability conditions may not be understood to provide the content (or truth conditions) of...meaning-attributing sentences’ — as this would be to provide a “straight solution”, which has been rejected as falling prey to the sceptical paradox —then ‘a descriptively adequate account of the *actual* assertion conditions for such sentences is the most one may cogently aim for’ (Boghossian 1989: 520). But given that the most that such “assertability” conditions can amount to is to rule out membership of a linguistic mob (“He’s no friend of ours”) then the “sceptical solution” fails even to hurdle this low bar.

§2a Metaphysical Branch of Wright's Reading of Wittgenstein

In face of the apparent dissolution of the very possibility of behaviour made in accordance with a standard of correctness ('there would be neither accord nor conflict here') —given that any course of action whatsoever can seemingly be given an interpretation which connects it with the expression of a rule— §201 of the *Investigations*, we will recall, concludes:

It can be seen that there is a misunderstanding here from the mere fact that in the course of our argument we give one interpretation after another; as if each one contented us for a moment until we thought of yet another standing behind it. What this shows is that there is a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call "obeying the rule" and "going against it" in actual cases.

Wright wishes to distance himself from the Kripkean reading by respecting Wittgenstein's explicit insistence that the paradox is to be rejected as based on a misunderstanding, and, moreover, he acknowledges that a key to avoiding the paradox, according to Wittgenstein, is to realise that there is a form of understanding a rule which is *not* a species of *interpretation*:

The paradox 'is not accepted by Wittgenstein. His response is not to propose an accommodation with it—a "sceptical solution"—but to discharge what he views as a faulty premise on which it depends: the idea that determinacy of meaning somehow depends upon *interpretation* (that to mean is to have an interpretation in mind.) (2001: 85).

Wright is certainly not wrong in what he says, but everything hinges on how we understand Wittgenstein's contention that the paradox shows that grasp of a rule is, in most instances, *not* an interpretation. And Wright's reading of *PI* §201 is vitiated by a

crucial, but fateful, assumption. Wright begins his reflections with a reading of Wittgenstein which takes him to allow, ‘apparently, that there is indeed an ineliminable multiplicity of conflicting possible interpretations’ of the expression of a rule (2001d: 124). Once this assumption is in place, then it soon leads Wright to conflate the very notion of “rationality” with that of *interpretation*. When we return to discuss the actual import of *PI* §201 in Chapter 3, I will show that Wright’s assumption, and its implications for our conception of rationality as such, embodies the very misunderstanding which Wittgenstein is trying to warn us against in this pivotal remark in his discussion of rule-following. My only aim in this sub-section, however, is to assess the chain of reasoning that Wright develops from the assumption, and to demonstrate that his conception of “a way of grasping a rule which is *not* an *interpretation*” as an appeal to primitive “subrational propensities”, or “classificatory dispositions” fails, on its own terms, to avoid the dilemma which it is called upon to enable us to avoid: either meaning vanishes in face of the problem of underdetermination, on the one hand, or, on the other, our explanatory ambitions for the theory collapse under the weight of our having to presuppose the determinacy of meaning—that which was meant to be explained.

As I have already mentioned, Wright begins his reading of *PI* §201 by assuming that there are, indeed, “an ineliminable multiplicity of conflicting possible interpretations” of any rule-expression²⁹. Given this assumption, then, how, if we are not going to succumb to the sceptical paradox and its attendant conceptual nihilism, are we to make sense of one’s understanding of a rule, or of a linguistic expression in a determinate manner? Wright justifiably recoils, as before, from any notion that understanding an expression according to a determinate meaning is ‘a matter of getting into some form of direct intellectual contact with a Platonic concept, or whatever’ (2001l: 395). In face of a seeming infinity of possible interpretations of the uttered expression, then, the challenge is to indicate how we are able to identify one unique interpretation from in amongst this multiplicity, but without recourse to the idea of a self-interpreting rule, and its

²⁹ Wright is, of course, basing his assumption on the kinds of Wittgensteinian considerations which we have continually invoked, encapsulated in Kripke’s formulation that: ‘no matter what is in my mind at a given time, I am free in the future to interpret it in different ways’ (Kripke 1982: 107).

concomitant image of each idiolectic rule-follower as making an inspired leap to, or guess at, the right “superlative” rule. How is this challenge to be met?

It was in the face of just this dilemma—meaning either vanishes in an infinite cloud of possible interpretations, on the one hand, or meaning must be viewed as an ultra-objective, superlative standard, utterly distinct from our linguistic behaviour (but which, by some occult process, is able to guide that behaviour), on the other—which, of course, led Wittgenstein, on Kripke’s reading, to accept the “sceptical conclusion”. In a *volte face* from his earlier concord with Kripke’s reading, however, Wright now claims that Wittgenstein’s conclusion ‘is explicitly not the sceptic’s, that there is no fact of the matter concerning the character of the subject’s understanding’ (2001e: 124). Rather, the point of Wittgenstein’s appeal, at *PI* §201, to “a way of grasping a rule which is *not* an *interpretation*” is, according to Wright:

...that something other than the exercise of interpretative (rational) faculties enters into the capacity to “read” another’s linguistic behaviour. And the additional something is, crudely, human nature: certain subrational propensities towards conformity of response, towards “going on in the same way”, which alone make possible the formation of the common conceptual scheme within which our rational capacities can be exercised (2001e: 124).

The first thing to note about this passage is that it explicitly equates the notion of rationality with that of interpretation: “interpretative (rational) faculties”. This equation is an implication of Wright’s fateful assumption: that any expression of a rule is open to indefinite interpretation. With this assumption in place, then all rational appraisal of data or evidence is, in turn, seemingly open to indefinite interpretation. That this is how Wright is led to conceive standards of rationality is displayed in his objections to any attempt to narrow down the range of putatively infinite interpretation by suggesting that only a limited number of interpretations of the available data are *rational*. Such a response

would be an effective denial that all is interpretation (Wright's first assumption): to deny, that is, 'that there are always already available indefinitely many hypotheses which, on the basis of certain evidence, may equally reasonably be adopted' (2001a: 15).

A teacher orders a pupil to provide the value of "68+57", let us say. According to the hypothesis that the teacher is here using the expression "+" to denote the addition function, then the correct response to the command is "125"; whereas, according to the hypothesis that the expression "+" is here being used to denote the quus function, then the answer I am obliged to give is "5"; and so on, for an indefinite many possible hypotheses. One possible response to this situation would be to propose that, although there may be indefinite many hypotheses, or interpretations, open in relation to the expression of a rule, 'we are nevertheless confronted with only finitely many genuine *probable* hypotheses —hypotheses which, in some objective way, it would be rational to take seriously, given the available evidence', such that 'all reasonable beings would sooner or later arrive at the same hypothesis' (ibid.). The fundamental problem with such a response, however, is that it merely shifts the original difficulty, as Wright recognises, to a different level:

For how are we supposed to have learned which procedures are rational, which type of hypothesis, although consistent with the data we possess, we may nevertheless rationally eliminate? If it was right to admit at all that we are confronted, in the process of learning any concept, with at any stage indefinitely many possible hypotheses about its correct application, then the same must be made with respect to the concept of a rational inductive inference (ibid: 16).

And now, as Wright goes on to conclude, 'our rationality could not be invoked...to cut the number of possibilities open to us down to size, since it is rationality itself that is supposed

to be being explained to us' (ibid.).³⁰ With such considerations in the background, then, and the first fateful assumption firmly in place, Wright concludes on Wittgenstein's behalf that 'a *merely rational* methodology can indeed yield no determinate conclusions' in regard to the disclosure of the meaning of a speaker's words (2001e: 124).

Given the equation of rationality with interpretation, then, this is, according to Wright, the significance of Wittgenstein's assertion that there is a way of understanding a word or rule 'which is *not* an *interpretation*' —and, so, given the supposed equivalence, a way of understanding a word or rule without any kind of *reason* to do so one way rather than another: 'Coming to understand an expression is not and cannot be a matter of arriving at a uniquely rational solution to the problem of interpreting witnessed use of it —a "best explanation" of the data' (2001l: 395). So how are we to give flesh to the notion of a way of understanding which is not a reason/interpretation? With the principal assumption in place —that there are indeed an indefinite number of rationalisations/interpretations available —and with Communitarianism now off the shelf, and on pain of lapsing into Platonist mythology, Wittgenstein's response to the difficulty, according to Wright, is the idea that 'a whole plethora of natural classificatory dispositions [must be shared] if we are to find each other's linguistic behaviour intelligible' (ibid: 124):

This simple point is of great importance, and armed with it, we can perhaps explain how my former understanding of [an expression] could be salient in a sufficient sample of my behaviour, even if no fully rational ground could be given for discounting various unwelcome alternative accounts of it (ibid: 125).

Wright here tentatively suggests that "perhaps" an appeal to "classificatory dispositions", or, to use the term used in the remark quoted a moment ago, "subrational propensities"

³⁰ Again, this points to the generality of the constitutive problem (discussed in Chapter 1), which seemingly impugns the very possibility of all and any behaviour made in accordance with standards of correctness (in this instance, standards of rationality).

can serve as the required explanation of what constitutes the use of a word according to a determinate meaning. Wright's explanatory aspirations—albeit circumspect—are driven by the fact that there do not seem to be any other available options: crude Communitarianism has failed, and the only other two apparent options—Platonist mythology, or conceptual nihilism—are wholly unpalatable. For the time being, however, this is nothing more than a promise for further constructive philosophical work. Accordingly, Wright elsewhere remarks that the strategic proposal here offered constitutes 'merely a direction' (2001l: 396). So, the most that has been achieved at this stage is that Wright's Wittgenstein has suggested a line of response to the paradox (an appeal to "subrational propensities" or "classificatory dispositions") which *may* prove fruitful. But, whether or not this line of response does turn out to bear any fruit is still left open. Yet no other viable options make themselves available to Wright. As to the actual fecundity of the suggested response, Wright has only got so far as to say, "I'll get back to you..."

We shall return to this (promissory) point in a moment, but it is worth noting here that communal institutions still have a role to play within Wright's reading of Wittgenstein, only not a *constitutive* one. "Subrational propensities" or "classificatory dispositions" can only play their constitutive role—somehow ('perhaps'): "I'll get back to you..."—Wright contends, 'courtesy of a co-operative *institutional* environment. There has to *be* such a thing as adding correctly before any physiological condition can constitute the ability to add' (2001f: 155). According to Wright:

Wittgenstein's contention is precisely that, with the demise of Platonism, there can be such a thing as adding correctly —such a thing as a determinate requirement imposed by the rules of addition —only within a framework of extensive institutional activity and agreement in the judgements which participation in those institutions involves us in making. The very existence of our concepts depends on such *activity* (ibid: 155-6).

I agree with these words wholeheartedly, as far as it goes, but, on the other hand, it matters greatly how we are to understand their import. In a moment I will urge that Wright's understanding of these Wittgensteinian sentiments cannot bear the explanatory weight they are called upon to carry, but in Chapter 3, I will explicate the actual (non-explanatory) weight such thoughts play in Wittgenstein's dialectic. But for Wright's Wittgenstein, given the rejection of the Platonist mythology — 'our internalisation of the same strongly autonomous, explanation-transcendent rules, whose requirements we then succeed, more or less, in collectively tracking' — together with the rejection of Communitarianism, then: 'Language, and all rule-governed institutions', it seems must —somehow ('perhaps'): "I'll get back to you..." — be founded 'in *primitive* dispositions of agreement in judgement and action' (ibid: 167).

It is worthy of note here that, although Wright has disavowed his earlier crude Communitarianism, the basic picture of communal agreement is not essentially different to Kripke's brand of Communitarianism—whereby such agreement is nothing but a 'brute empirical fact' (in this instance, pertaining to the "subrational propensities" of our primitive human nature). It is, quite simply, Wright asserts: 'a *basic* fact about us that our ordinary forms of explanation and training do succeed in perpetuating practices of various kinds —that there is a shared uptake, a disposition to concur in novel judgements involving the concepts in question' (2001f: 167). Institutional practices, Wright suggests, are best viewed 'as an expression of certain basic reactive propensities, primitive classificatory dispositions—a common human (or at least cultural) heritage without which our language would fail' (2001c: 72). Moreover, notwithstanding such quasi-scientific terminology, the basic picture of the ground of such brute agreement doesn't seem to have moved us far beyond Kripke's contention that the rule-follower's 'own confident' and 'brute inclination' to respond 'one way rather than another' is 'to be regarded as primitive' (Kripke 1982: 87, 15, 91).

Communal agreement cannot be explained, as it could if we were prepared to entertain Platonism, through appeal to the 'rules-as-rails mythology'; rather, 'the truth is the other way round: it is the basic agreement which sustains all rules and rule-governed institutions'

(2001f: 167). But, of course, since Communitarianism has also been jettisoned, then this basic agreement (supposedly underwritten by our primitive subrational nature) cannot be said to *constitute* such rules and rule-governed institutions either. Yet, as Wright candidly admits, it is difficult to see how we can escape the drift into this conclusion:

The great difficulty is to stabilise [this aspect of Wittgenstein's thought] against a drift into a fatal simplification: the idea that the requirements of a rule, in any particular case, are simply *whatever we take them to be*. For if the requirements of the rule are not constituted, as the Platonist thought, independently of our reactions to the case, what is there to constitute them *but* our reaction? (ibid; emphases in original).

How, then, is this difficulty to be overcome? It is at this point in the dialectic that Wright accuses Wittgenstein of leaving us somewhat in the lurch. Wittgenstein 'tells us that the requirements of rules exist only within the framework of activities which depend upon basic human propensities to agree in judgement' (ibid: 168). On the other hand, Wittgenstein reminds us that such requirements provide standards by virtue of which those judgements, 'even if they enjoy consensus,' can be deemed to be incorrect:

So we have been told what does *not* constitute the requirement of a rule in any particular case: it is *not* constituted by our agreement about the particular case, and it is *not* constituted autonomously, by a rule-as-rail... But we have not been told what *does* constitute it; all we have been told is that there would simply be no such requirement—but for the phenomenon of actual, widespread human agreement in judgement' (ibid: 168).

It is at such stages that Wright ruminates over Wittgenstein's principled "quietistic" stand in failing to engage in the seemingly necessary constructive endeavour required in response to such questions: 'All along [in Wittgenstein's later philosophy] he spares

himself the labour of providing more satisfactory alternatives to the accounts he aims to demolish or make us uncomfortable with' (2001h: 220). Wright laments Wittgenstein's 'seemingly stubborn refusal to respond to the sense of intellectual vacuum which the apparent negative thrust of the rule-following discussion is apt to create' (2001: viii). Nevertheless, Wright recommends, it is important that 'we do not despair of giving answers to constitutive questions too soon' (2001g: 191). Such constructive endeavour may be 'difficult', Wright urges, but 'not impossible' (ibid). 'Wittgenstein is very inexplicit about his "way of grasping a rule which is not an interpretation"', Wright complains (2001: 395). Wright admits that his own gestures towards "subrational propensities," and so forth, are 'merely a direction. But it is', he registers, 'a hopeful direction' (ibid: 396).

Yet it is difficult to see how the notion of "subrational propensities" can sustain such promise: for, unless we can be shown otherwise, it doesn't move us beyond the very dilemma with which we began. On the one hand, if such propensities are viewed in brute (non-semantic, non-intentional, or non-rational) terms, then the possibility of using a word according to a determinate meaning becomes unavailable in face of a seeming infinity of possible brute dispositions which can be interpreted as being in accord with the word's meaning. On the other hand, if we view the propensities as being able to uniquely determine a specific way of responding, then it appears that these propensities constitute a standard by virtue of which our "merely rational" (read: "merely interpretative") behaviour is guided in response to the expression of a determinate rule, which then merely presupposes what was meant to be explained. Until we have been shown otherwise, the only advance on the original position is that, instead of the picture of each idiolectic rule-follower somehow channelling an ultra-objective rule (and which rule transcends our mere behavioural response to the rule-expression), we now have the picture of our "subrational propensities" channelling such a superlative rule —which is no advance at all! We have seemingly only transferred the picture of 'direct intellectual contact with a Platonic concept' to a picture of *indirect* contact, via the intermediary of our "subrational propensities" (ibid.: 395).

As we discussed in the last chapter, dispositional theories of meaning are unable to simultaneously capture the two requirements which such theories must satisfy in order to disarm the threat of paradox: on the one hand, such a theory must specify only those dispositions which are meaning-determining, whilst, on the other, and on pain of vicious circularity, such a specification must be couched in non-semantic or non-intentional terms. Now Wright's reparsing of "subrational propensities" as "classificatory dispositions", quoted earlier, strongly suggests that he (and his Wittgenstein) tries to have it both ways here. On the one hand, the notion of a *non-classificatory* disposition is of no use on its own, as such brute dispositions fail to uphold the required normativity of meaning. On the other hand, and on pain of presupposing what was meant to be explained, such dispositions cannot be pictured as inherently *classificatory*, or meaning-determining. In other words, the very terminology invites an equivocation which keeps alive the possibility of future constructive research ("I'll get back to you..."). At some moments, the stress needs to be on the *classificatory* nature of these dispositions, so that the possibility of determinate meaning is preserved; whilst, at other moments, and on pain of vicious circularity, the classificatory element needs to be hidden beneath their being conceived as merely brute (that is to say, non-classificatory) *dispositions*³¹.

Similar reflections can be made about the term "subrational propensities": the propensities can't be altogether *rational*, of course, on pain of presupposition; but neither can they be entirely *non-rational*—or, simply "propensities". And so we are effectively left again with the original question: how can these two requirements be combined together without lapsing into conceptual nihilism, on the one hand, and a superlative, Platonic mythology, on the other? Wittgenstein's appeal to the notion of grasp of a rule "which is *not an interpretation*", at least on Wright's reading, has, at best, merely modified the object of puzzlement: "How is it possible for a rule-follower to determine a specific response to the expression of a rule?" now effectively becomes, the arguably even more

³¹ Given such equivocation, together with the fact that, as we saw in the previous chapter, the two requirements cannot be satisfied simultaneously, it is little wonder that, and as Wright elsewhere candidly admits, disarming the threat of the paradox seemingly requires 'constructive philosophical developments which one can approximate but whose completion seems to remain tantalisingly out of reach' (Wright 2001: viii). (Or, if you'd prefer, the "I'll get back to you..." remains forever frustrated—frozen solid in suspended animation between two irreconcilable poles). We will return to this point in Chapter 3.

queer: “How is it possible for subrational propensities (as yet hidden, but to be found somewhere in our human nature: “I’ll get back to you...”) to determine a meaningful use of a word, or rational response to a rule?”

We will return again to consider Wright’s appeal to “subrational propensities” in §2b(ii), and, in Chapter 3, I will explicate the actual import of Wittgenstein’s appeal to “a way of grasping a rule which is *not* an *interpretation*”—which is indeed fundamental to a proper understanding of his response to the threat of paradox. Regardless, although, there seems little prospect that an appeal to “subrational propensities” can serve to disarm that threat, this is not anyway, according to Wright, the crucial move made by Wittgenstein in face of the paradox. Even if this strategy were not unsuccessful, the appeal to “subrational propensities” or “classificatory dispositions”, Wright remarks: ‘is a point to use against Kripke’s sceptic only from an assumed *third-personal* point of view, as it were, with respect to one’s own linguistic past’—that is, given the accumulated data provided by one’s linguistic behaviour thus far (2001e: 125). According to Wright, such a strategy, even were it successful, ‘would do nothing to legitimate the idea of one’s present and former meanings as apt for *avowal*...And it is the legitimation of that idea that a satisfying response to the sceptic must ultimately be found’ (ibid.). This point gestures towards the “first-personal branch” of Wright’s Wittgenstein’s two-pronged response to the paradox, to which we now turn.

§2b Epistemological Branch of Wright's Reading of Wittgenstein

The third-personal branch of Wright's reading concerns what we might term the constitution of linguistic content —the claim that meaning is *somehow* constituted by facts about our human nature ("subrational propensities"), even if we remain in the dark about the nature of such facts. In the first-personal branch of his reading of Wittgenstein, however, Wright turns the focus of the discussion onto the *epistemology* of linguistic and mental content. Now, I argued in the previous chapter that the problem which derives from Wittgenstein's later work raises the spectre of a decidedly *constitutive* scepticism, and does not directly concern epistemological scepticism. That is to say, the problem concerns the *possibility* of meaning, mental content, or of correctness conditions more generally, and does not directly concern how we may be said to come to *know* such phenomena. That being said, and as Boghossian remarks: 'one may agree that the problem is constitutive in character, and yet believe it to have an epistemological dimension' (Boghossian 1989: 516). This is an entirely legitimate contention. Epistemological considerations of rule-following are not independent of the question of the constitution of rules. In fact, they seem essentially bound together: questions over the constitution of rules are essentially bound up with how those rules are known, and vice versa. Yet, as we shall see, Wright takes us in a far more contentious direction by urging that the epistemological considerations yield particular—and radical—constitutive results (and which, as we shall eventually discover, are not unrelated to his problematic speculations about the broadly physicalist constitution of rules in Wright's "third-personal" branch).

In brief, Wright argues that because self-knowledge of one's intentional mental states cannot be known either experientially (via an observational, or introspective paradigm) nor through correlation with dispositional behaviour (via an inferential paradigm) then self-ascriptions of intention do not track independently constituted states of affairs. Rather, he asserts, our judgements about the content of our own mental states *themselves constitute* that content. Wright's views on self-ascriptions of intentional mental content complement his denial that if we are to rescue any notion of the objectivity of linguistic meaning, then our conception must be shorn of the idea that such

objectivity is independent of our judgement or ratification. Wright's Constructivism is motivated by a thoroughly justifiable rejection of the Platonistic mythology. However, as I will demonstrate in §2c, this rejection is held together with a wholly questionable identification of the very idea of ratification-independent objectivity—the idea that self-ascriptions of the meaning of one's words, or of one's intentional mental states, track states of affairs which are independent of our judgements—with that Platonic mythology. Given this identity, then, rejection of the Platonistic mythology is, for Wright, *eo ipso* rejection of ratification-independent objectivity.

§2b(i) Wright Manoeuvres Against the Sceptic

In order to grasp Wright's position, we must begin by considering his diagnosis of the fundamental flaw in the "sceptical argument" as it is developed by Kripke. And this is that when the sceptic challenges one to produce some fact of the matter that constitutes one's meaning something determinate by one's words, on any occasion of utterance, he relies upon 'an unsupported reductionism' (2001f: 146). Specifically, the sceptic implicitly assumes that 'the putative species of knowledge in question' must be 'essentially inferential' (2001d: 111). I cannot simply cite the fact that I meant, or mean, addition by my use of the word "plus" on any occasion of utterance. According to the ground rules of the sceptical challenge: 'States having content are somehow to be constituted out of materials whose description, at the point where they are introduced into the debate with the sceptic, does not presuppose their contentfulness' (2001f: 147). Or as Wright otherwise puts the point: the argument put forward by the sceptic 'now emerges as merely an implicit prejudice against the idea that we may and usually do non-inferentially know of our current meanings and intentions, and may and often do non-inferentially recall them later' (ibid.).

In Chapter 3, I will essentially argue that Wright's charge that Kripke's sceptic adheres to an unsupported reductionism is well-founded, and embodies an important insight for overcoming the threat of paradox. Nevertheless, in the remainder of this chapter, I will argue that Wright's attempt to provide a direct response — albeit avowedly anti-reductionist — to the sceptical challenge fails yet again to avoid the problematic dilemma (in fact, as I will argue in Chapter 3, it is the very attempt to provide a direct rebuttal of the paradox which is problematic). In other words, Wright's Constructivist response to the problem fails yet again to escape the seemingly unavoidable choice between a magical, superlative conception of linguistic or mental content, on the one hand, or conceptual nihilism in face of the threat from interpretational regress, on the other.

As we saw in the previous chapter, Kripke does, in effect, consider the possibility of responding to the (constitutive) sceptic by baldly stating that “I meant (or mean) plus by ‘plus’”:

Perhaps we may try to recoup, by arguing that meaning addition by “plus” is a state even more *sui generis* than we have argued before. Perhaps it is simply a primitive state, not to be assimilated to sensations or headaches or any “qualitative” states, nor to be assimilated to dispositions, but a state of a unique kind of its own (Kripke 1982: 51).

Kripke, we will recall, echoes the anguished voice at §195 of the *Investigations*, which cries out ‘...but in a *queer* way, the use itself is in some sense present’, when he concedes that ‘we surely suppose – unreflectively – that something like this is indeed the case. Even now I have a strong inclination to think this somehow must be right’ (Kripke 1982: 52). Yet Kripke immediately rejects the strategy of appealing to such an irreducible, *sui generis* state as ‘desperate’ (ibid: 51). The notion of such a state seems irredeemably eerie, or queer: ‘Can we conceive of a finite state which *could* not be interpreted in a quus-like way? How could that be? ...[T]he nature of the supposed “state” is left mysterious’ (ibid: 52-53). ‘The idea in my mind of a finite object’, Kripke continues: ‘can it not be interpreted as determining a quus function, rather than a plus function?’ (ibid: 54). Wright sympathises with this sense of queerness: ‘How could there be such a state, available immediately to the subject, apt for authoritative avowal and non-inferential recall,’ and ‘yet which is potentially infinitely fecund, possessing directive content for no end of distinct situations?’ (2001e: 125; 2001d: 111). Yet, despite registering his sympathy with this sense of mystery, Wright decries Kripke’s knee-jerk rejection of the question as ‘merely obscurantist’ (2001e: 125):

Kripke’s discussion contrives to leave the impression that [the question] is rhetorical, that we have not the slightest idea what such a state might be. Whereas a little reflection shows that both these features—non-

inferentiality and indefinite “fecundity”—are simply characteristic of the normal intuitive notion of *intention* (2001d: 111).

In response to Kripke’s brusque rejection of the possibility that the state of meaning something specific by a word is *sui generis* and irreducible as too strange to take seriously, Wright draws attention to the fact that we assign the same supposedly “desperate” properties to our ordinary notion of self-ascriptions of intentional states: knowledge of one’s own intentions ‘is based on inference neither from one’s behaviour nor from any other occurrent aspect of one’s mental life’ (2001e: 126). In fact, we routinely form intentions without any explicit reflection at all, inferential or otherwise: I may simply form the intention to take the dog for a walk, for instance—“in a flash”—without any kind of deliberation. Furthermore, and barring special circumstances (I am in a state of acute inebriation, say, or am a compulsive liar), ‘we are credited with a *special authority* for the character of our own intentions’: that is to say, our answers to the question of our own intentions ‘should be given a special weight’ (2001d: 111-2). If called upon to explain how I know that I intend to take the dog for a walk, for instance, I may simply respond: “because that’s what I intend to do”.

This authoritative aspect of my avowal does not, of course, cancel the defeasibility of my intention in light of what I go on, or don’t go on, to do. There is, of course, a sense in which I am in a privileged position in regard to my own intention. Nevertheless, taking someone else’s dog for a walk, say, or my going for a solitary walk, or even my staying in to clean the oven—any action *at all*, in fact, that does not involve my *taking the dog for a walk*—necessarily entails that my original intention remains unsatisfied. Moreover, intentions can be conceived as “infinitely fecund” in the sense that I may be called upon to fulfil my express intention to defend Fido from harm, say, in a potentially indefinite number of occasions where Fido is placed in danger (or, at least, where I perceive that he is in danger).

As uncontroversial as these considerations are, their essential point, at least for Wright's strategic purposes, is that 'the intuitive concept [of intention] seems to contain the resources for a direct rebuttal' to the Sceptical Argument: 'Since I can know my present intentions non-inferentially, it is not question-begging to respond to the sceptic's challenge to my knowledge of my past intentions to reply that I may simply remember them' (2001e: 126-127). Thus, when called upon to cite some mental fact about my previous linguistic performance which serves to rule out quus-like interpretations, I may simply recall my former intention with respect to my use of the word "plus". Similarly, my present use of the word can be specified by appeal to my present semantic intentions. Now, whilst Wright admits that it is 'not...particularly *comfortable*' to conceive the meaning of our words 'as consisting in...a certain general intention', the crucial point of such a proposal is 'that the Sceptical Argument has absolutely no destructive force against [it]' (2001d: 113).

Wright's proposal, then, is that our intuitive notion of intention might somehow ("I'll get back to you....") be used to directly rebut the sceptic's challenge to cite a meaning-constituting fact of the use of our words, but without begging the question against him. In order for such a strategy to have any hope of success, however, 'a satisfactory philosophy of intention has to validate our claim to non-inferential authority for our present (and previous) intentions without succumbing to the mythology of infinite, explicit, introspectable content' (ibid: 115). Yet, as we saw in the previous chapter, this temptation can be difficult to avoid. For the connection between an intention and the act which fulfils, or fails to fulfil it can seem no less mysterious than the connection between understanding the meaning of a word and appropriate practical application of the word.

Consider again, for instance, Wittgenstein's query at *PI*§197: '[W]hat kind of super-strong connection exists between the act of intending and the thing intended? —Where is the connection effected between the sense of an expression "Let's play a game of chess" and all the rules of the game?"' 'And how, moreover,' Wright adds, 'can it be reasonable to credit a subject—in the cases where we do—with a *special* authority concerning such states to himself...?' (2001e: 136). Despite such mystery, however, according to Wright:

‘One way or another we have to answer, or undermine, the question: how is first-personal authority for intentional states possible? Until we do, we have not really answered the Sceptical Argument’ (2001f: 150).

Before we consider Wright’s strategy in action, it is worth pointing out a couple of doubts as to its potential. The first is that, although Wright does acknowledge the generality of the constitutive problem, the sceptic doesn’t, in any case, only subject the determinacy of meaning to doubt, but explicitly includes intentional mental items: ‘But ultimately the sceptical problem cannot be evaded, and it arises precisely in the question how the existence in my mind of any mental entity or idea can *constitute* “grasping” any sense rather than another’ (Kripke 1982: 54). Recall, also *PI* §437: ‘A wish seems already to know what will or would satisfy it; a proposition, a thought, what makes it true – even when that thing is no there at all! Whence this *determining* of what is not yet there? This despotic demand?’ Wright himself, at certain moments, acknowledges that puzzles about “non-inferential self-knowledge” are ‘difficulties which variously apply not just to meaning, but to understanding, thinking, intending, hoping, expecting, and so on...’ (2001: 85). Facts about mental contents in general are, he avers, ‘surely in exactly the same boat’ as facts about linguistic meaning (ibid: 83).

Moreover, linguistic meaning can be shown to possess parallel features to those upon which, in the case of intentional mental concepts, Wright seizes for his anti-sceptical ambitions. Although there may be occasions on which one has to search around for the right words to say, whilst engaged in everyday discourse speakers are routinely able to immediately and non-inferentially say what they mean (the words simply “roll off the tongue”). Furthermore, although words can be misused, speakers are generally credited with a *special authority* for the meaning of their words (“No, I don’t mean that—what I mean is...”). And words are, of course, “infinitely fecund”, given that there may be no end to the occasions on which I may be called upon to utter a particular linguistic expression. So what promises are held, one wonders, by switching the focus of our investigations from one parallel case to another? If we’ve so far faced a brick wall in regard to

explaining such features in the case of linguistic meaning, then what do we expect will be different in the coordinate case of intentional mental concepts?

The second, more serious, doubt as to the potential fecundity of Wright's strategic proposal relates to his revelation of the sceptic's reductionist prejudices. For, although Wright aims to circumvent the sceptic's implicit insistence that our response to his request to reveal some meaning-constituting fact must be inferential, by probing the non-inferentiality of avowals of intention his strategy still assumes that any direct response to the sceptical challenge must be in terms of an *interpretation*, in the sense of 'the substitution of one expression of the rule for another' (*PI* §201). That is to say, Wright's strategy still implicitly assumes that we cannot supply a direct response to the sceptic by simply pointing out the meaning-constituting fact ("I meant, or mean addition by 'plus'"). 'Rather,' Wright submits, 'the challenge is to recall some *independently characterised* fact'—according to Wright's strategic proposal, facts about one's semantic intentions (2001g: 176; emphasis in original).

§2b(ii) Putative Paradigms of the Mental

Wright explicates the difficulties with which Wittgenstein grapples, difficulties which can arise when we try to make sense of intentional concepts (those such as desire, expectation, intention, belief, meaning and understanding and so forth) as grounded upon a rough-and-ready distinction which Wright (and his Wittgenstein) detects between two putative classes of mental states and processes. The first class, or paradigm of the mental is characterised by occurrent mental states (standardly, sensations and the like); the second paradigm of the mental, on the other hand, is exemplified by dispositional states (emblematically, qualities of character, such as courage, patience, laziness, etc.). The difficulties with the relevant class of intentional mental concepts (intention, belief, meaning and understanding, etc.) arise, according to Wright, owing to their seeming ‘to hover, puzzlingly and unstably, between two [epistemological] paradigms’—knowledge by observation and knowledge by inference (2001f: 149). As I will argue in this subsection, however, the real source of the difficulties is to be found, rather, in the putatively paradigmatic conceptions of the mental realm themselves, upon which the two epistemological paradigms are based.

In order to understand the difficulties which Wright, and his Wittgenstein, detect in regard to intentional mental concepts, we must identify two essential features of such concepts which must be upheld simultaneously by any account of their nature which is even minimally adequate: their *authoritative avowability* and their *constitutive connectivity*. As to the first essential feature, one is, as Wright says, ‘for the most part, effortlessly and reliably authoritative about, say, one’s intentions’ (2001f: 148). Thus, although I need not bring an intention of mine to explicit discursive awareness—whether through public speech, or via internal monologue, say, or “inner” speech—I am perfectly capable of doing so: “I intend to such-and-such”; “I believe that so-and-so”; “I desire this-or-that”, “What I mean to say is...”, and so forth. Moreover, self-knowledge about intentional mental states, and about the meaning of our words is, as Wright puts it, ‘typically immediate’: that is to say, non-inferential (2001: 84). And this immediacy brings us to the second essential feature of intentional mental states, which is that there is a *constitutive connection, or internal relation* between such states and the performances,

events, and states of affairs which are appropriate to the satisfaction, or fulfilment of the determinate content of the particular mental state. As Wright puts the point:

But the striking feature about self-ascription of an intention is that one *identifies* it at the time: there is no such thing as knowing that one has an intention of some sort but not know *what* is intended. Since we are able confidently and authoritatively to self-ascribe intentions which we have never had before, it simply cannot be that we identify our intentions in general by inductive association...’ (2001e: 137).

In summary, then, any minimally adequate self-reflective account of intentional mental states must successfully combine these two essential features of their nature: their first-person avowability together with the fact that they are constitutively connected to those determinately appropriate performances, events, and states of affairs which can be said to conform to the content of the specified intentional state.

Now, two epistemological paradigms are, Wright claims, ‘liable to blinker our response’ to this requirement of combinatorial adequacy: ‘the paradigm of knowledge by inference and of knowledge by observation’ (2001i: 310). As should become clearer as we proceed, the latter paradigm is the epistemological counterpart to the paradigm of occurrent mental states, whilst the former epistemological paradigm naturally conforms, for Wright, to the behavioural-dispositional, or physicalist paradigm. As I mentioned, Wright sees the putative difficulties with intentional mental concepts as owing to their tendency ‘to hover, puzzlingly and unstably,’ between these two epistemological paradigms’ (2001f: 149). Thus, the apparent problem with intentional states is most adequately conveyed in terms of an antinomic predicament. On the one hand, as Wright has it: ‘the difficulty raised by the concepts with which Wittgenstein was preoccupied is that we are pulled in both directions [that is, towards each distinct epistemological paradigm] simultaneously’ (2001g: 177-178). On the other hand, ‘neither [paradigm] is adequate for the class of psychological concepts with which Wittgenstein was most concerned’ (2001f: 149).

According to Wright, we are drawn towards the epistemological paradigm of inference by the constitutive connectivity of intentional mental states. Remarking upon this essential feature of intentional mental concepts, Wright unproblematically asserts: 'the identity of a subject's intentional states is constitutively answerable to her (subsequent) capabilities and behaviour...' (2001: 86). Where matters get problematic, however, is when he continues: '...in a fashion which is broadly analogous to that of dispositional states' (ibid.). Making the same point elsewhere, Wright says of intentional states that 'they answer constitutively to ways in which the subject manifests them, and to that extent conform to the characteristics which...are properly conceived as dispositional' (2001i: 311). In a moment, I will argue that Wright cannot unproblematically help himself to the idea that we can appeal to dispositions in order to uphold the constitutive connectivity of intentional states. But such an appeal can at least make it intelligible why Wright thinks this essential feature pulls us towards the paradigm of knowledge by inference. For, according to Wright: 'The warranted self-ascription of any disposition must attend, one would suppose, the possession of relevant evidence: evidence, in the best case, provided by actual manifestations of the disposition' (2001: 84).

Given that I am about to argue against the appeal to dispositions as such, I do not wish to dwell on the sense in which self-ascriptions of dispositional propensities can be said to be inferential, but it is certain that, as Wright says, their warrant is based 'only on grounds which any third-party could employ': that is to say, the ground of self-attributions of laziness, say, is based on publicly manifest lazy behaviour (2001f: 148). It also does seem true to say that attributions of dispositional propensities are based upon "an accumulation of evidence", in the sense that an attribution of a propensity to laziness would be unfair to someone who manifested lazy behaviour only ever on one occasion! Regardless, the crucial point for present purposes is that the problem, for Wright, 'is that to conceive of one's meanings [or intentions] as consisting in dispositions would in any case be in tension with one's ability to know of them *non-inferentially*' (2001: 84). Yet, as I hope now to show, the problem is not so much that to conceive one's intentions as dispositional makes a mystery of their authoritative first-person avowability; the central problem, rather, is the conception of one's intentional mental states as dispositional.

Wright introduces his conception of the dispositional-behavioural paradigm of the mental as exemplified by:

...qualities of character—like patience, courage, and conceit—which are naturally viewed as constituted in the (broadly) behavioural dispositions of a subject, are fully manifest in things he is inclined to say and do, and advert to no inner phenomenological causes of these inclinations (2001g: 177).

As I have already suggested, however, before we even begin to consider this second, dispositional paradigm in relation to the putatively problematic mental concepts (meaning, understanding, intention, and the like) which, Wright claims, “hover puzzlingly and unstably between [the] two paradigms,” the dispositional paradigm can already be shown to be deeply problematic and unstable on its own terms. Now, of course, it is not wrong *per se* to class such qualities of character as dispositional. The point at which such an attribution becomes problematic, however, is when, as is in the present case, such characteristics are called upon to play a role within a wider *explanatory* framework: in this particular instance, for the purposes of attempting to explain the putatively mysterious properties of intentional mental states. With this goal in mind, we are naturally drawn to conceive such qualities of character— those which are being marked as emblematic of the dispositional paradigm—as *mere* or *brute* dispositional propensities to react: thus, Wright remarks, this paradigm of the mental is ‘an option which Wittgenstein expects, writing when he did, will naturally take a behaviourist shape so that now “it looks as if we had denied mental processes” [Wright is here quoting *PI* §308]’ (2001k: 359). In the current philosophical climate, however, Wright suggests that this paradigm will ‘more likely [be] physicalist’ (ibid.).

But insofar as we are drawn to conceive such dispositions as merely brute, then it seems problematic to call this a paradigm of the *mental* at all (hence: “it looks as if we had denied mental processes”). That this is so is suggested by considerations which parallel those we discussed in relation to dispositionalist theories of meaning in the first chapter,

and which surfaced again, particularly during the discussion at §2a. As Kripke and Boghossian argue: if we are led to regard meaning or mental content in terms of brute dispositional propensities, then we fail to preserve the constitutive connection between the meaning of a word or of the content of an intention, on the one hand, and those types of performances which are determinately *appropriate* to the meaning, or of the intention, on the other. As Wright acknowledges, there is ‘an internal connection’ between ‘aspects of a subject’s (subsequent) doings and reactions which mental states of this kind [i.e., understanding, expecting, intending, hoping, etc.] essentially sustain’ (2001i: 296)³².

Now, of course, qualities of character, such as patience, courage, and conceit are not *intentional* as, say, a state of expectation is intentional. That is to say, being disposed to courageous behaviour cannot be said to be *about* anything, in the sense that being expectant that someone comes to tea is *about* whether or not that person does come over for tea. But there is, nevertheless, an internal relation, or constitutive connection between, say, the concept courage, on the one hand, and publicly observable manifestations of courageous behaviour, on the other. Yet, viewing such dispositional propensities as merely brute (or *purely* physical, if we are swept up in prevailing trends) fails to preserve this essential constitutive connection. The relation between such brute propensities and outward behaviour could only be contingent, or “external” (as a *merely* causal connection, perhaps³³). Thus, we would fall foul of a version of the problem of disjunction, or of underdetermination: *any* manifestation of behaviour could, under some interpretation, be brought into relation with the hypothesised dispositional states.

Let me try to make this point clear by transposing a remark Wright makes about intentional mental states. He writes:

³² For a discussion of the notion of an “internal” connection or relation, and the development of Wittgenstein’s thought in regard to them, see Baker & Hacker 1984: esp. Chapter 3.

³³ I enter this qualification—“*merely* causal”—in order not to prejudice the question of whether internal relations, or constitutive connections can also be causal relations. A proper discussion of this important question is beyond the scope of the present work: (but see, for instance, Thornton 1998: Chapter 6). Nevertheless, given issues of the kind presently under discussion, it should be clear that if any sense can be made of the idea that constitutive connections *can* be causal, they better not be *merely* causal.

...if expecting someone to tea were a matter purely of the occurrence of certain states and processes of consciousness, then it would be a contingency how it correlated with any subsequent goings-on which can be matters of public awareness. But such correlations are not...points of contingency: an expectation is constitutively identified by the distinction it effects between states of affairs which comply with it and those which do not (ibid: 296-297).

Transposing to the case of dispositional qualities of character: if being courageous were a matter purely of the actualisation of certain *brutely* dispositional (or *purely* physical) states or processes, then it would be a contingency how those states or processes correlated with manifestations of behaviour which can be matters of public awareness. But such correlations are not...points of contingency: a courageous disposition is constitutively identified by the distinctions it effects between behavioural manifestations which comply with it (standing one's ground, for instance) and those which do not (fleeing hardship or adversity, say).³⁴ Note that it is not open to Wright to protest here that those concepts which pick out dispositional characteristics, such as *courage*, are not psychological or mental predicates: for in that case, why are such characteristics so much as being canvassed as exemplifications of a possible paradigm of the *mental* at all? Neither is it open to Wright to attempt to avoid this problem by conceiving the relation between dispositional states and their behavioural manifestation as a 'relation...of antecedent state to symptom or trace' (2001f: 149). For what ultimately grounds the distinction Wright draws between the two paradigms is that, in the case of the first—that exemplified by occurrent mental sensations—it is precisely that, according to Wright: 'the relation, we conceive, of for example pain to its expression, can only be that of symptom or trace. By contrast, in the case of the other paradigm—that of the dispositional psychological state—the connection between behavioural display is not symptomatic but

³⁴ In making this transposition I am here following the lead of McDowell who, as we shall soon see, makes a parallel transposition (or 'parody,' as he has it) of Wright's remarks on coming to understand a rule "in a flash" to the case of feeling pain (McDowell 1998d: 302-303). (I am transposing McDowell's transposition, if you like!)

constitutive' (ibid.). Yet, as I am in the process of arguing, it is difficult to see how Wright can help himself to that latter claim without begging the question.

Now, under pressure from the foregoing kinds of considerations, we are intelligibly drawn to conceive those qualities of character being called forth as exemplary of the behavioural, or physicalist paradigm of the mental in a contrasting sense. Driven by our explanatory ambitions, we are given to conceive such dispositional characteristics as merely brute propensities to react, which brute reactions are summoned as grounds for the constitutive connections between mental predicates, on the one hand, and determinate manifestations of publicly observable behaviour, on the other (the constitutive connection between *courage* and courageous behaviour, for instance). However, once we realise that dispositions pictured as merely brute propensities to react cause us to lose sight of such constitutive connections altogether, and with seemingly no hope of their retrieval, then we are intelligibly drawn to a conception of the requisite dispositional characteristics as '*somehow* wholly manifest and available to public view' (2001k: 359; emphasis added). (Hence the remark of Wittgenstein's, invoked by Wright—'And now it looks as if we had denied mental processes' (*PI* §308)—continues: 'And naturally we don't want to deny them').³⁵ Thus, we are tempted to say, only *these*

³⁵ Let me make it clear that the point I am in the process of making is *not* that there is any problem with the conception of dispositional characteristics (courage, conceit, patience, etc.) as essentially bound to a specific mental predicate as such. Precisely not: that there *is* a constitutive connection between determinately courageous behaviour, say, and the concept *courage* is the very relation we are trying to uphold. To deny any such relation would, in effect, be a denial of the psychological concept. What is problematic, however, is the sense of mystery expressed in the reservation "*somehow*": thus, immediately preceding the remark invoked by Wright at *PI* §308, Wittgenstein declares: 'So we have to deny the yet uncomprehended process in the yet unexplored medium.' (That there is such a sense of mystery surrounding "the yet unexplored medium" is, of course, what gets the blood pumping for explanatory ambitions, or promises of constructive philosophical enterprise: "I'll get back to you..."). That the sense of mystery surrounding the constitutive connection is the problem, and not the constitutive connection itself, will be one of the defining themes of Chapter 3 (for a parallel discussion of these matters, see McDowell 2009a: 81-86). Until it becomes time to discuss the matter further, however, consider this exchange (in the context of a parallel query as to the nature of the constitutive connection between understanding the meaning of a linguistic expression and performance, or use) at *PI* §195:

"But I don't mean that what I do now (in grasping a sense) determines the future use *causally* and as a matter of experience, but that in a *queer* way, the use itself is in some sense present."

—But of course it is, 'in *some* sense'! Really the only thing wrong with what you say is the expression "in a queer way". The rest is all right...

kinds of dispositional propensities can manifest *this* kind of behaviour (determinately *courageous* behaviour). But, insofar as we are drawn to such a conception of dispositional propensities, we are now, problematically, presupposing a species of the very genus—internal relations or constitutive connections—which we are setting out to explain.

Furthermore, in parallel to the point that was made earlier (§2a) in relation to Wright's use of the terms "subrational propensities" and "classificatory dispositions", the dilemma intelligibly tempts us towards the use of equivocal terms of expression, such as when, in a remark quoted a moment ago, Wright refers to the behavioural paradigm as that of 'the dispositional psychological state' (2001f: 148). In the present case, as before, the notion of a mere (*non-psychological*) dispositional state is inadequate for our purposes, as we are then unable to sustain the required constitutive connection between the hypothesised dispositional state and what it disposes us to do (manifest determinately courageous behaviour, for instance). As a consequence, we are driven onto the high road leading to conceptual nihilism. On the other hand, on pain of presupposing some form of constitutive connection (which presupposition would extinguish our explanatory ambitions) we are debarred from conceiving such hypothesised states as inherently *psychological* (such that the putative dispositional state is already internally related to the concept *courage*, for instance). That is to say, in order to keep our explanatory ambitions on life support, at some moments we are drawn to picture a "*dispositional psychological state*" (where dispositions here are merely brute propensities to react), whilst, at other moments, we are drawn to picture a "*dispositional psychological state*" (where *appropriate kinds* of dispositional propensities are here being implied)³⁶.

³⁶ Of course, ambiguity in this general vicinity need not be restricted to compound expressions: the standalone terms "disposition" and "propensity" can be ambiguous in the context of reflecting on the difference between rational and non-rational animals, given that it can intelligibly be said of both brutes and humans that they can display dispositional propensities— although mere brutes lack conceptual awareness. (And, for that matter, the term "psychological" can be ambiguous in this context too). Thus, Wright connects his claim that sensations make no demands on our conceptual resources to the observation that 'very small children and animals...have pains and tickles, ...can become anxious or elated' (2001i: 298). Perhaps another way to clarify the point I am driving at is by borrowing McDowell's expression: 'a mere uncomprehending disposition to react' (2009b: 101, fn. 3). Now, in addition to the three qualities of character listed in the quoted remark with which I introduced the second paradigm (patience, courage, and conceit), Wright's catalogue of dispositions emblematic of the behavioural paradigm include: 'intelligence...[or] endurance' (2001e: 136); 'wisdom, tolerance, laziness' (2001f: 148);

This is a predicament which we have already encountered before, of course, but I wish now to add a further point to those which have already been discussed. As Wright observes, the dispositional paradigm which is here being problematically relied upon in an attempt to explain intentional mental states, could be pictured in behavioural terms, but, in light of prevailing philosophical trends, is more likely to be pictured, ultimately, in terms of purely physical facts. So conceived, dispositional propensities would provide us with the tools with which, we imagine, an account of intentional mental states could be proffered, which comfortably conforms to a broadly scientific outlook. Yet, as we have been discussing, insofar as we are led to conceive dispositional propensities in terms of the kind of purely physical events and processes which are investigated by physical scientists, we lose sight of the kinds of constitutive connections which are the focus of *our* investigations, and without any hope of their retrieval.

Given that any swing entirely in the other direction (conceiving dispositional propensities as already constitutively connected to determinate patterns of behaviour) would involve us in presupposing what we are hoping to explain, however, in order to sustain our explanatory ambitions we are led to equivocal conceptions of dispositional states—“subrational propensities”, “classificatory dispositions”, “the dispositional psychological state”, and so forth. Or, similarly, our gestures towards dispositional states are vague, or tentative—as when Wright, for instance, asserts: ‘the ascription of intentional states to a subject appears answerable to what he goes on to say and do in

‘irritability or modesty’ (2001i: 311). Perhaps we could plausibly grant that laziness, endurance, or irritability could be conceived as *brutely* or *merely* (that is to say, uncomprehendingly) dispositional: as it can be intelligibly said of a sloth that it is disposed to laziness, for instance; or it can be said of a horse or a camel that it is disposed to endurance; or any mere brute can be said to be disposed to irritability if, say, it is tired, hungry, or thirsty. But there seems little plausibility in the suggestion that someone is disposed to wisdom, say, yet *uncomprehendingly*. Notice that it does no good to protest here that a person can be uncomprehendingly wise, in the sense that she just happens not to be aware that she possesses these qualities of character (that she is behaviourally manifesting her disposition to modesty, let us say). For this is *not* the sense in which mere dispositional propensities are being conceived, within the explanatory context at issue, to be uncomprehending. Rather, the sense under examination is one in which these dispositional propensities of hers, on pain of presupposing what we are setting out to explain, are *in principle* beyond her powers of comprehension (her powers of mental, or conceptual awareness). I might, as a contingent matter of fact, not be aware of my propensity to laziness. Yet—if I became more self-aware—I could, in principle, bring my predisposition to laziness within the bounds of my comprehension. It can plausibly be said of a sloth that it is predisposed to laziness; but what plausibility is there in saying that it could ever become aware of this fact? (*cf. inter alia* McDowell 1996, esp. Lecture VI & Postscript; 1998c; 2010).

the *broadly identificatory or constitutive fashion* in which the ascription of dispositions and capacities is so answerable' (2001i: 311; emphasis)³⁷. Yet, in hedging our bets in this fashion, what we are imagining here cannot, of course, be instances of the regular kind of purely physical facts which are examined by physical scientists. The facts we are imagining cannot be explicitly of *that* kind, on pain of having to face the problem of underdetermination, and, from there, being led on to conceptual nihilism³⁸. So, essentially, what we are being led to imagine is a mysterious kind of *quasi-physical fact*: something which broadly resembles an ordinary physical fact except, unlike an ordinary physical fact, is a special *content-bearing or constitutive-connection-maintaining* physical fact³⁹.

Such mystery, of course, immediately raises the question of how we are to make sense of such special—"queer" or "spooky"—posited states and processes: "subrational propensities", "classificatory dispositions", and so forth. Yet what we have effectively done now is shift the focus of our puzzlement. Our investigation began, recall, with questions of the sort: "How is it possible that people are able to use words according to a determinate meaning?" or "How is it possible that people can form intentional mental states with specific mental content?" But now the reformulated question is more along the lines of: "How is it possible that quasi-physical (or quasi-casual) facts determine meaningful linguistic behaviour, or specify mental content?"⁴⁰

³⁷ Compare *PI* §149: 'If one says that knowing the ABC is a state of the mind, one is thinking of the state of a mental apparatus (perhaps of the brain) by means of which we explain the *manifestations* of that knowledge. Such a state is called a disposition...'

³⁸ Thus, *PI* §149 continues: '...But there are objections to speaking of a state of mind here, inasmuch as there ought to be two different criteria for such a state: a knowledge of the construction of the apparatus, quite apart from what it does'. That is to say, the relation between the state and behavioural manifestations could only be contingent, or "external" (most likely: merely causal).

³⁹ Compare *PI* §196: 'In our failure to understand the use of a word [or, we might add, an intentional mental state] we take it as the expression of a queer *process*. (As we think of time as a queer medium, of the mind as a queer kind of being.)'

⁴⁰ Thus recall *PI* §195: '...in a *queer* way, the use itself is in some sense present...' In the current context, this could be reformulated along the lines of: '...in a *queer* way, the meaning or content itself is in some sense present in some kind of *quasi-physical fact*...'

In the context of a parallel discussion of analogous mysterious states and processes, at *PI* §158 Wittgenstein imagines a possible response along the following lines: But isn't the mysteriousness of such states and processes 'only because of our too slight acquaintance with what goes on in the brain and in the nervous system? If we had more accurate knowledge of these things we should see what connections were established by [our] training [into language, etc.]...' Yet such an objection would miss the point. For in positing mysterious quasi-physical, or quasi-causal facts what we are in effect doing is: either not providing a scientific explanation at all; or we are, by implication, effectively suggesting that we need to radically reconceive what we are willing to class as a scientific fact, such that the (duly broadened) empirical sciences are now able to welcome such mysterious, quasi-physical facts into its stable. We will return to discuss such matters again in Chapter 3, but let us turn now to consider the second of Wright's putative paradigms of the mental.

Turning now to the other essential feature of intentional mental concepts—their authoritative avowability—Wright claims that the relevant states 'conform to the paradigm of sensation and other "observable" phenomena of consciousness' (2001i: 311). Now, of course, it is true to say that one cannot "observe" one's intention to make a Caesar salad, say, in the sense in which one can observe the setting sun. Nor, indeed, can one be said to be aware of an intention in the same sense in which one's toothache can be maintained as the focus of one's conscious attention. Nevertheless, it seems simple common sense to say that one *can* be self-consciously aware that one has such an intention—which awareness, moreover, can take in the determinate *identity* or *content* of the intention: in this instance, precisely the intention to *make a Caesar salad* (and *not* the intention to make a lentil curry, nor a Welsh rarebit, nor to take the bins out, and so on *ad infinitum*). Not only this, but, as McDowell rightly points out, it can also be the case that an intentional state can be an *occurrent* phenomenon of consciousness: as when one is 'struck by a thought'—one's instantaneous realisation that one forgot to turn off the oven, say (McDowell 1998d: 303).

Examples of occurrent mental states, and which provide the avatar for this paradigm of the mental, include aches and pains, tickles and tinnitus— mental states which ‘may have a definitely dated onset and departure, and which may be interrupted by breaks in consciousness’ (2001f: 149). As mentioned, the crucial point for Wright’s purposes is that such occurrent mental states are, like the putatively problematic class of intentional states, ‘authoritatively avowable’: ‘The subject is authoritative about such states because, since they are events in his consciousness, he is in the nature of the case conscious of them’ (ibid.). So far so good. Where Wright’s characterisation of this paradigm of the mental starts to become deeply problematic, however, is when he labels such occurrent mental states as ‘*genuine* episodes and processes in consciousness’ (2001g: 177; emphasis added). And what Wright means by “genuine” here is revealed when he unpacks this claim by referring to such phenomena as ‘pure states of consciousness’, owing to the fact that they are ‘states whose occurrence’, according to Wright, ‘at least in a central class of cases, makes no demands upon the conceptual resources of the subject’ (2001i: 298)⁴¹.

We will return to consider McDowell’s arguments against Wright’s conception of “genuine”, or “pure” states of consciousness in a moment. But first we need to consider Wright’s reasons for thinking that, despite the fact that we are pulled towards the introspective paradigm (knowledge by “observation”), this paradigm is inadequate to account for intentional mental states. Remarking upon the class of occurrent states of consciousness—paradigmatically (although, as I have suggested, problematically) exemplified, in Wright’s conception, by our awareness of sensations: awareness of one’s toothache, for instance—Wright correctly observes that: ‘Such states are authoritatively avowable...The subject is authoritative about such states because, since they are events in his consciousness, he is in the nature of the case conscious of them’ (2001f: 149). This seems, once more, basic common sense—trivially true, in fact. (Although, of course, it matters a great deal how we come to understand such common-sense statements). Now,

⁴¹ Wright’s qualification—“at least in a central class of cases”—is there to leave space for ‘genuinely episodic states whose occurrence is conditional on the conceptual resources’ of the subject — ‘having a tune run through one’s head’, for instance (2001g: 177, fn. 17). Thus, Wright acknowledges that the distinction he (and his Wittgenstein) draws isn’t everywhere sharp.

intuitively, the claim Wright here makes in application to occurrent mental states in particular should—so long as matters are handled with care and qualifications entered, where necessary—have equal application to *all* states and occurrences of consciousness, including intentional mental states. Intuitive, perhaps—yet Wittgenstein, at least according to Wright’s reading, denies that such is the ground of the authority of our self-ascriptions of intention.

Remarking on the fact that we are naturally drawn to the epistemological paradigm of introspection in order to explain the authoritative avowability of intentional states, Wright says: ‘Cartesians would view the authority [of avowals of intention] as having the same kind of basis which it finds for a subject’s authority concerning his or her occurrent sensations. The subject has privileged access to the state, is immediately aware of it in consciousness’ (2001e: 128). Depending on how we construe our conscious awareness of occurrent sensations, of course, and of what we mean by “privileged access”, there is nothing problematic about this claim as such—again, depending on how we unpack these statements, they merely express the common sense conception of the grounds of our cognitive access to states and occurrences in consciousness. Yet Wright, together with his Wittgenstein, deny that ‘the authority standardly granted to a subject’s own beliefs, or expressed avowals, about his intentional states...is...a *consequence* of the nature of those states, and of an associated privileged relation in which the subject stands to them’ (2001i: 312). In other words, according to Wright, the ‘authority with which our self-ascriptions of meaning [and] intention...assume is not based on any kind of cognitive advantage, expertise or achievement’ (2001e: 137-138). Self-knowledge of one’s own intentions is not a matter of being in touch with a ‘truth-conferring state of affairs, something manifest to consciousness but distinct from the judging, whose content is precisely that such a state of affairs obtains’ (2001i: 299).

Despite these claims, Wright insists that he is, nevertheless, not defending a form of Kripkean irrealism about intentional content. Although Wright proposes that we abandon the common sense idea that our self-ascriptions of intention reflect the nature of judgement-independent intentional states, this rejection is not meant to impugn the

content of intentional states as such. Rather, his proposal is that we reject the intuitive notion that intentions themselves, once introspectively identified, determine those events, states of affairs and performances which constitute satisfaction of the original intentional state: that we 'try to make sense of the idea that knowing one's own beliefs, desires and intentions is not really a matter of "access to"—being in cognitive touch with—a state of affairs at all' (ibid: 312). Once abandoned, Wright recommends that we replace the intuitive conception with a radically revised picture, one according to which our self-ascriptions of intention *themselves* determine or constitute the content of those intentions: 'the authority standardly granted to a subject's own beliefs, or expressed avowals, about his intentional states is a *constitutive* principle': 'it enters primitively into the conditions of identification of what a subject believes, hopes and intends' (ibid.).

In the context of discussing the objectivity of linguistic meaning, Wright urges that: 'We must endeavour to see...content...as plastic in response to speakers' continuing performance' (2001c: 72). Not unlike the case of our conception of intentional mental states, Wright's recommendation is bound to the claim that we reject the intuitive conception of the objectivity of linguistic meaning: 'The competent use of basic vocabulary, with whatever degree of confidence, should not be viewed as reflecting cognition of the requirements of objective meanings' (ibid.). Wright's recommendation in the present context, then, is that we extend this metaphor of the plasticity of linguistic content to apply also to the content of intentional mental states: the proposal is that we view both semantic content, and intentional mental content more generally, 'as ever open to ongoing linguistic behaviour' (ibid.).

In defence of the claim that the epistemological paradigm of observation is inadequate to the purpose of providing an account of intentional mental states, Wright's Wittgenstein offers a phenomenological argument and an *a priori* argument, each complementing the other. As to the former, Wright claims that the upshot of the rich and subtle seam of phenomenologically perspicacious observations which are undoubtedly to be found in the *Investigations* is that meaning, understanding, and the identity of intentional mental states are 'persistently fugitive if sought within the sphere of consciousness' (2001f: 148).

‘There is nothing introspectable,’ Wright insists, ‘by which intending, understanding, believing, etc. even *seem* to be distinguished’ (2001i: 296). If this were the consequence of the relevant reflections in the *Investigations* then it seems implausible, to say the least. As McDowell points out, if this were correct, then ‘the upshot of the discussion initiated by [PI §138 would be] that one cannot grasp meaning in a flash’ (McDowell 1998d: 304). Of course, McDowell further observes, ‘standing states of understanding are not occurrent phenomena of consciousness’ (ibid: 303). Having registered that point, however, McDowell continues, ‘grasp of a meaning [can] be an occurrent phenomena of consciousness’: for instance, when we instantaneously grasp a word’s meaning, or when, say, one visualises something, or when one is struck by a thought (ibid: 304). On the other hand, as McDowell points out: ‘It is not by observation that one knows what one means, or meant’ (McDowell 2009a: 91). Again, having said that, however, one *can* be self-consciously aware of what one means, or meant by one’s verbal utterances.

Wright would counter, however, that neither he nor his Wittgenstein are denying that introspectable, occurrent events may enter into one’s self-conscious awareness in connection with grasping the meaning of a word, say, or of understanding the rule for an algebraic series, or when one has an intention to do something. Thus, Wright’s Wittgenstein would allow that a picture of a cube may well come to mind when someone utters the word “cube” in one’s hearing, say, or a formula may enter one’s consciousness when one is ordered to develop a decimal expansion, or one may well become consciously aware of distinctive experiences whilst reading a text, and so on. Nevertheless, Wright’s point—and this brings us to the *a priori* argument—is that, ‘even if we could identify such phenomena of consciousness (grasping meaning “in a flash”, say)’, the constitutive connection *itself*—that which serves to distinguish appropriate linguistic performance from behaviour which does not accord with the word’s meaning—‘can’t be “read off” from [the subject’s] conscious phenomenology at the time [the meaning of the word, say, is understood]’: ‘nothing of which the subject is conscious intimates [the constitutive connection] to him’ (2001e: 136).

‘To take a familiar instance’, Wright explains:

...if coming to understand “in a flash” the rule governing a decimal expansion were a matter *purely* of the occurrence of certain events in consciousness, then it would be a point of contingency that people who so come to understand something are able subsequently to deliver appropriate kinds of performance. But this is not a contingency; it is a conceptual requirement on the propriety of describing someone as having understood that they are able to perform properly. Ascribing understanding is ascribing (something very like) an ability (2001i: 296).

The *a priori* argument which Wright is attributing to Wittgenstein is, then, that even though we may allow that certain events *may* occur in consciousness in relation to one’s understanding—an image flashes through the mind, say, or one experiences a feeling of elation when it suddenly dawns on one (“in a flash”) how to continue an arithmetical series, and so on. Nevertheless, the crucial point for Wright’s Wittgenstein is that any such occurrence can only be contingently, or “externally” related to subsequent performances, states, and affairs, etc. But an intention, say, is *identified* by its constitutive connections with such determinate performances, reactions, and so forth. Therefore, Wright concludes on Wittgenstein’s behalf, the identity, or content of an intention cannot *itself* be something that is available to consciousness: ‘Simply, there is nothing for an intention, conceived as determining subsequent conformity and non-conformity to it autonomously and independently of its authors’ judgements on the matter, to *be*’ (ibid: 317).

That appropriate performance is constitutively, and not contingently, connected to attributions of understanding is undeniable. But there are some fundamental problems with Wright’s reading of Wittgenstein which we need to consider. The first is that, whilst it is correct to say that understanding is in the nature of a practical ability—that my standing knowledge of the meaning of a word, for instance, is constituted by my being able to suitably deploy the word during occasions of my own linguistic performance, or to react appropriately on occasions of its utterance in the speech of another—as I have tried to indicate in the preceding discussion, an awful lot rides on how we understand the

notion of “ability”, and its cognate terms. Moreover, those previous discussions should raise our suspicions that Wright is, problematically, conceiving “ability” as something akin to “subrational propensity”, or “classificatory disposition”; and, insofar as that it is so, it makes it moot whether any such appeal to “abilities” can serve the purposes for which such an appeal is made—that is, to uphold the essential constitutive connectivity of intentional concepts. Wright is no doubt correct that Wittgenstein is clearly concerned to uphold essential normative notions such as accord, and of constitutive connections more generally. But McDowell has provided powerful objections to Wright’s reading, which objections begin by observing that Wright’s construal of the *a priori* argument ‘misses two directions in which [Wittgenstein’s discussion of meaning and understanding] generalizes. One is to occurrences, as opposed to states, that involve intentionality; the other is to non-intentional aspects of the inner life’ (McDowell 1998d: 300).

In order to substantiate his claim that Wright’s construal of Wittgenstein’s *a priori* argument overlooks the first direction over which those arguments generalise, McDowell refers to *PI* §386, where Wittgenstein considers the case of an occurrence, as opposed to a state, involving intentionality—the specific instance of which is having in mind the image of a particular colour. This passage, McDowell submits, ‘raises a parallel problem’ as that raised in relation to the constitutive connection between grasp of an arithmetical series and appropriate performance—‘again beset with parallel pitfalls’ (ibid: 301). The relevant portion of *PI* §386 reads:

...if you say that you see something red in imagination, then it will *be* red. ...I say without hesitation that I...have imagined this colour. The difficulty is not that I doubt whether I really imagined anything red. But it is *this*: that we should be able, just like that, to point out or describe the colour we have imagined, that the projection of the image into reality presents no difficulty at all. Are they then so alike that one might mix them up?

As McDowell observes, the issue raised by the passage clearly has to do with the *intentionality* of the image: or, in Wittgenstein's own words, "the projection of the image into reality". As McDowell suggests, 'focusing on the subject's ability to point or describe what he imagines is just a way of making this issue vivid' (McDowell 1998d: 301). This seems right, given that any such performance must point to or describe *this colour* (the one in the imagination). Hence *PI* §388: 'How can one *know* that one can show it [or point to, or describe, it] if..., in other words, that one can recognize it if one sees it? How do I know from my *image*, what the colour really looks like? How do I know that I shall be able to do something?...'. The image is of something determinately *red*—and *not* something blue, or purple, or mauve; it isn't an image of a colourless liquid, or of my old primary school headmaster, and so on *ad infinitum*. That we are indeed in familiar territory here is suggested by the implication that, when we ask after the possibility of such precise, determinate content, we can be tempted to imagine that the image and what it is an image of are "so alike that one might mix them up". What is being imagined here is the counterpart to the ultra-objective, superlative rule. Of course, the imagery of the rule-as-rail has no purchase here, as it does in the case of an arithmetical series, but that this is what Wittgenstein is gesturing at is confirmed during the following imagined exchange at *PI* §389: "'The image must be more like its object than any picture. For, however alike I make the picture to what it is supposed to represent, it may still be the picture of something else as well. But it is essential to the image that it is the image of *this* and of nothing else.'" Thus one might come to regard the image as a super-likeness'.⁴² Regardless, the significance of this passage, as it relates to Wright's reading, is summarised by McDowell in the following jab:

But if anything is an occurrent phenomenon of consciousness, surely having a visual image of a colour is. Wright's reading has Wittgenstein denying that intentionality is found among the occurrent phenomena of consciousness; but it is found there, and Wittgenstein cheerfully registers that it is' (1998d: 301).

⁴² McDowell correctly perceives that the idea of a "super-likeness" is the counterpart in this context of the "despotic demand", linked to "the hardness of the logical must" at *PI* §437, and of "the super-strong connexion...between the act of intending and the thing intended" at *PI* §197 (McDowell 1998d: 301, fn.6).

Earlier I allowed that Wright could submit that certain phenomenological events may occur in consciousness—a picture of a cube when one understands (“in a flash”) the word “cube”, for instance, or an algebraic formula enters one’s consciousness upon momentary understanding of a series, or one feels an experience of elation when one grasps the series thus—only so long as it was recognised that no such conscious phenomena could be constitutively connected with anything outside consciousness: most notably, subsequent, appropriate performance. Yet, the foregoing reflections put even this concession in doubt. For the picture is already presupposed as a constitutively connected occurrence—it is a picture of a *cube*—just as it is already being presupposed that it is an *algebraic formula* which is in mind. Of course, a feeling of elation may not in all instances be intentional. But even here, we may ask: what makes it a feeling of *elation* (as distinguished from, say, morbid dread, or trapped wind...?)

That Wright is deaf to such considerations is suggested by his response to McDowell’s discussion of an intentional occurrence in consciousness, such as an image before one’s mind:

When an image, or picture, comes before my mind, it presumably can’t constitute a more explicit or substantial presence than the coming of a real physical picture before my physical eye. And when the latter happens, it is of course consistent with my being in full command of all manifest features of the object that I remain ignorant precisely of its intentionality—of what it is a picture of. I want to say that, analogously, in the sense in which an image or mental picture can come before my mind, its intentionality cannot (2001j: 342).

But this is to miss entirely the import of Wittgenstein’s reflections: the fact that Wright clearly thinks that one can be “in full command of all the manifest features” of a picture, yet not know what it is a picture of, is precisely the kind of idea that Wittgenstein and

McDowell seek to attack⁴³. If I was so out of touch with the content of a picture—if I saw nothing more than a flat, square object with some coloured markings on it, say, then it is doubtful that I would even be seeing it as *a picture*. But even if, say, I can tell that it's a picture, but simply don't know what is being depicted (let us say it is a picture of an historical event, of which I am ignorant), then it would still not be true to say of me that I am "in full command of all the manifest features" (of course, I could overturn my present lack of command—by, say, engaging in relevant historical research).

Wright's *a priori* argument is that, if an intention were a matter *purely* of an occurrence in consciousness, then understanding could be only contingently connected to subsequent performance. But the connection between an intention and appropriate performance is essentially constitutive, not contingent. Therefore, an intention cannot be an occurrence in consciousness. Yet, given Wright's conception of an occurrence in consciousness—those which, in the view of he and his Wittgenstein, make no demands on the subject's conceptual capacities ("pure" or "genuine" occurrent states, in Wright's words, or 'mental events and processes strictly so termed' (2001j: 329)) —then it is doubtful, as McDowell argues, 'whether anything *could* be an occurrent phenomenon of consciousness' (McDowell 1998d: 302). No doubt, Wright would reply that, of course not—that is precisely Wittgenstein's point, on his reading. Yet, as I hope will become clear in §2c, the deeper point is that Wright can *only* conceive occurrent phenomena of consciousness in this problematic fashion, such that rejection of this picture is *eo ipso* rejection of occurrent mental phenomena being available to self-conscious awareness.

This brings us to the second direction in which the Wittgensteinian problems generalise. McDowell argues: 'If Wright's *a priori* argument worked, it would have to be because nothing introspectable could sustain internal relations to anything outer' (ibid.). It is indeed true that, *if* events in consciousness can only be conceived in the manner that Wright conceives "genuine" episodes in consciousness, then the conclusion does indeed

⁴³ Thus, the very next remark following the passage discussed by McDowell, reads: 'The *deep* aspect of this matter readily eludes us' (PI §387).

follow: for such putative mental phenomena could *only* be essentially unconnected to any publicly available events, states of affairs, performances, or behavioural manifestations. But, so far from instituting, as in Wright's reading of Wittgenstein, a class of "genuine" items for consciousness, McDowell urges that 'this conception of the introspectable, as a realm of facts constituted as what they are independently of internal relations to anything outside that realm', is a form of the picture of the introspectable which Wittgenstein is most fundamentally, and generally, concerned to undermine—for, 'to suppose such items are objects of consciousness', McDowell exhorts, 'would be precisely to make them into "private objects"' (ibid: 302, fn. 9; 312):

But it is precisely Wittgenstein's point, about the items Wright is willing to count as paradigms of occurrent phenomena of consciousness, that they do sustain such internal relations; they are not just externally related (by mere causation, for instance) to the public circumstances that constitute their expression. To parody what Wright says about coming to understand a word in a flash: if feeling a pain were a matter purely of the occurrence of certain events in consciousness, then it would be a point of contingency that people who feel pain were in normal circumstances disposed to engage in the sort of behaviour that we count as expressive of pain; but it is not a contingency. But if feeling a pain is never a matter purely of the occurrence of certain events in consciousness, then surely nothing is (ibid: 302-303).⁴⁴

Of course, this opens up a contentious topic, and a lot more needs to be said than this (more than is possible within the scope of the present work). Yet, enough has been said, surely, to cast serious doubt on the plausibility of Wright's *a priori* argument, considered both in itself (given that it rests upon a deeply suspect conception of "genuine" mental states and processes), and as a reading of Wittgenstein. Wright's claim, that intentional mental states cannot be made available to self-conscious awareness independently of our self-ascriptions, is based upon the idea that the relevant psychological concepts "hover

⁴⁴ This, of course, is precisely the parody I referenced earlier (p.122, fn. 34).

puzzlingly and unstably” between two epistemological paradigms. But I believe enough has been said—both by myself in relation to the behavioural-dispositional paradigm, and by McDowell in relation to the paradigm of occurrent mental states—to suggest that the problem is not so much with the antinomic tension between the two epistemological paradigms—that of knowledge by observation and knowledge by inference—as they are brought into relation to intentional mental concepts. But, rather, that the real source of the difficulties derives from the respective conceptions of those epistemological paradigms themselves, together with the complementary conceptions of the paradigms of mental concepts upon which they are based.

Summing up his manoeuvres against Wright’s problematic conception of the paradigm of “observation”, McDowell reflects:

Wittgenstein’s thought does not leave untouched the picture of the introspectable as a domain of self-containedly knowable states of affairs, only externally related to anything outside themselves, and expel the intentional from that domain. The key argument generalizes so as to undermine that picture of the introspectable. Once we understand that, we can see that there is no need to be suspicious of including intentionality among the occurrent phenomena of consciousness... [T]his is not an encouragement to category-mistakes; of course standing states of understanding are not occurrent phenomena of consciousness. But it should no longer seem problematic to countenance, say, being struck by a thought as an occurrent phenomenon of consciousness’ (ibid: 303).

Indeed, the very idea of being struck by a *thought* is, on Wright’s conception, incoherent⁴⁵. But the conclusion of his *a priori* argument only follows unless there is

⁴⁵ In this connection, Wright makes the absurd claim, on Wittgenstein’s behalf, that: ‘...X’s belief that P is false because there is *no such thing as* the belief that P...’ (2001h: 244).

some way of conceiving an event in consciousness which is *already*, in some sense, constitutively connected to publicly available events or states of affairs—that the determinate *identity* or *content* of an intention is, as McDowell puts it, ‘exactly something that can be, in some sense, all there in one’s mind before one acts on it’ (ibid: 315). Of course, more needs to be said about *how* we are supposed to make sense of such basic common sense notions of intentional mental concepts: ‘the idea that an intention, once identified, determines of itself, and in particular independently of its author’s subsequent judgements, what counts as conformity with it’ (ibid: 316). One of the principal aims of Chapter 3 will be to show how we can retrieve such basic common sense in face of the threat of paradox, or of antinomic tensions. But that we are able to give such a positive characterisation is imperative: for, as McDowell remarks, not only should it be ‘innocuous’, it is ‘indeed compulsory, on pain of losing our grip on our purported topic...to take it that intentions contain within themselves the distinction between conformity and non-conformity’ (ibid: 315).

§2c “A Gangway with Rigid Walls”

Giving expression to common sense notions, McDowell states that an intention is “exactly something that can be, in some sense, all there in one’s mind”; or, he states that we can, as he puts it, have ‘meanings in mind’ (ibid: 305). The trouble is, though, that when Wright hears such locutions, all he can hear on McDowell’s lips is Platonistic mythology, and which must, therefore, be resisted: ‘when we are mindful of the distinctive marks of avowals,’ Wright avers, ‘it appears that the states, and the mode of knowledge, must be something rather out of the ordinary’ (2001i: 367). In order to begin to understand how that which is common sense in the mouth of one man, is, to the ears of another, something extraordinary and mysterious, we must consider how Wright’s view of the paradigm of (“genuine”) occurrent mental states colours his conception of the corresponding epistemological paradigm—that of knowledge by observation.

Earlier, I quoted Wright referring to the denizens of the putative class of mental concepts as ‘pure states of consciousness’ (2001i: 298). Correspondingly, Wright identifies ‘an epistemology of observation’ with the picture of ‘pure introspection’ (ibid: 312). And what Wright means by “pure introspection” is revealed in the following passage, where he declares what he takes to be ‘the requirement...imposed by the observational model’ as such, which is that:

...intentional states can be fully determinate objects of inner contemplation before they issue in anything outward, before they are acted on—and hence that they can be fully identified without any consideration of events, reactions and performances lying in the future (2001i: 311).

What Wright is imagining here is the counterpart, although now applied to intentions, of the rule-as-rails imagery which comes naturally in the case of the rule for an arithmetical series. As we discussed in Chapter 1 (§4), Wittgenstein considers such potential imagery in the context of reflection on the order to develop the series “+2” (*PI* §§187-188). In the same context, Wright considers such a picture of ‘the autonomy of rules, the image of a

rule as rails laid to infinity, tracing out a proper course for practice quite independently of any judgement of the practitioners' (2001g: 178). Earlier, we will recall, Wright expressed the imagery in terms of 'ultra-objective standards' over against which our practical applications of the rule can be judged: such an imagined rule, as Wright puts it, 'conducts you like a gangway with rigid walls' (2001a: 11).

Remarking now on the possibility of our putative epistemological relation to any such 'rule-as rails, rules which somehow reach ahead of us and determine of themselves their every actual and counterfactual proper application', Wright considers: 'But if we have the capacity to keep track of rules when so conceived, we must be capable of somehow getting them "in mind"...The grasp of such a rule is thus the internalisation of an open-ended set of preordained requirements, an informational state accessible...only by a kind of guesswork' (2001f: 163-164). The latter mysterious feat of guesswork was earlier referred to, we will recall, as "'cottoning-on"—that is, a leap, an inspired guess at the pattern of application' demanded by the superlative, ultra-objective rule (2001a: 35). In such a situation, any attempted explanation of the dictates of the rule by another could only be, ultimately, 'illustrations' of mere finite portions of the application of the rule—'necessarily imperfect' gestures in the direction of the infinite expanse of the rule-as-rails (2001f: 164).

Yet as Wright had earlier remonstrated, once his initial crude Communitarian ambitions had been extinguished: 'there only seem to be these two options' (2001: 5). There is Platonist mythology, on the one hand, or the regress of interpretations paradox, on the other—the latter of which threatens to leave us 'spiralling into some kind of incoherent irrationalism about meaning' (2001f: 169). In a different sense to that which was originally meant by Wright, then, our thought really does, as a matter of fact, appear to be confined within a "gangway with rigid walls".

Yet the apparently irresolvable predicament with which our thought is seemingly fated is, as Wright once again recognises, 'perfectly general', and applies to the use of any word or

rule: 'Thoughts you may have had about how, quite generally, you would be prepared to use an expression will suffice to meet the sceptic's challenge only if you presuppose their *proper interpretation*' (2001d: 99). But Wright cannot see anything but mythology in such a response: 'Yet no category of mental item can be appropriate to the challenge except a general thought; only such a thought can *have enough to say*, can cover the indefinitely many potential situations (2001d: 99). There can be no such mysterious, infinite state contained within our finite minds. So, again, our predicament appears to dictate that there are only two, equally unpalatable options: 'It therefore appears that the only ploy that has any chance of accrediting your understanding of "green" with an appropriately general normative role...fails to meet the sceptic's challenge. And now "it seems the whole idea of meaning vanishes into thin air" [quoting Kripke 1982: 22]' (ibid: 99-100). And our thought is, yet again, revealed to be imprisoned within a "gangway with rigid walls..."

Returning to the issue of intentional states with which we began, it should be clear, if it wasn't clear before, that when Wright claims that we are required by the very idea of the introspective epistemological paradigm—or "pure introspection", as he calls it—that intentional states can be 'fully determinate objects of inner contemplation before they issue in anything outward, before they are acted on', or that they 'can be fully identified' without any consideration of subsequent sayings and doings, he can only conceive the intentional state in mythical, superlative form (2001i: 311). Wright cannot conceive how "having intentions in mind" can refer to anything but states 'which *seem* to have a content that can somehow transcend that of any accompanying thoughts in the subject's mind' (2001e: 114). Thus, in this context, Wright refers to *PI* §197:

There is no doubt that I now want to play chess, but chess is the game it is in virtue of all the rules (and so on). Don't I know, then, which game I want to play until I *have* played it? or are all the rules contained in my act of intending? Is it experience that tells me this sort of game is the usual consequence of such an act of intending? so is it impossible for me to be certain what I was intending to do? And if that is nonsense—what kind of

super-strong connexion exists between the act of intending and the thing intended?

Justifiably recoiling from any such Platonistic mythology, Wright erroneously concludes that the thrust of Wittgenstein's reflections is 'that the intention, construed as...an event in its own right, is evanescent' (2001e: 134). Wittgenstein, according to Wright's reading of him, rejects the common sense 'conception of avowals as *reports* and the associated conception of a self-standing subject matter which they serve to report' (2001k: 358). The *Investigations*, Wright contends, 'accomplishes a critique in depth of the Cartesian view after which we really have no alternative but to discard the inner observational model altogether' (2001k: 345). "There only seem to be these two options...!" On pain of having to accept the hopeless Platonist (or "Cartesian") mythology, we are seemingly forced—"a gangway with rigid walls..."—to conclude that 'particular psychological states are not items *for* consciousness in the way the Cartesian represents them' (2001i: 296). But notice that it does *not* follow that, therefore, intentional states are not items *for* consciousness. And, as McDowell has already warned, it better not follow, on pain of our losing our grip entirely on our purported topic...⁴⁶

⁴⁶ In response to McDowell's objection that Wright can see 'nothing but platonistic mythology in the idea that an intention determines what counts as conformity with it independently of its author's judgements' (1998d: 314), Wright, to my mind at least, simply goes on to prove the point:

If I form the intention to type a period, then sure, only typing a period will do. The anti-Platonist point is rather that there is nothing for my intention's having had *just that content* to consist in, if the fact has to be constitutively independent of anything which I may subsequently have to say about compliance or non-compliance with the intention, about what its content was. The Platonist mythology is a mythology of such constitutive independence (2001j: 339; emphasis in original).

Of course, my intention to type a period is not constitutively independent of my subsequent judgements, in the sense that only a judgement with a content along the lines of "I intend, or intended, to type a period" will constitute a correct judgement on the intention. But it does *not* follow that the intention is, of course, independent of any actual *act* of judgement. I may conceivably decline to make any judgement at all concerning my intention—yet, for all that, my intention is still *the intention to type a period*, and is constituted as such independently of any particular *act* of judgement (which I may, or may not, go on to make). Wright, it seems to me, conflates these two senses of "judgement-independence". (For a parallel distinction—between the *act* of thinking and *thinkable content*—see McDowell 1996: 28.)

But notice now that, despite our “long and involved journeyings” with Wright throughout this chapter, we haven’t moved any further beyond where our reflections began! Recall, *PI* §138-139:

...But we *understand* the meaning of a word when we hear or say it; we grasp it in a flash, and what we grasp in this way is surely something different from the “use” which is extended in time!

When someone says the word “cube” to me, for example, I know what it means. But can the whole *use* of the word come before my mind, when I *understand* it in this way? ...Can what is present to us in an instant, what comes before our mind in an instant, fit a *use*?

Such a sense of queerness or spookiness can be found strewn throughout Wright’s reflections, for instance: ‘How is the *subject* to know of the application of such a concept to himself if nothing of which he is conscious intimates the fact to him? And how, moreover, can it be reasonable to credit a subject—in cases where we do—with a *special* authority concerning such states to himself...?’ (2001e: 136). Or: ‘How is it possible to be, for the most part, effortlessly and reliably authoritative about, say, one’s intentions if the identity of an intention is fugitive when sought in occurrent consciousness...and the having of an intention is thought of as a disposition-like state?’ (2001f: 148). Compare Kripke’s own analogous expression of queerness: ‘It is not supposed to be an introspectable state, yet we are supposedly aware of it with some degree of certainty whenever it occurs. For how else can each of us be confident that he *does*, at present, mean addition by “plus”?’ (Kripke 1982: 51). ‘Sometimes,’ Kripke confides, ‘when I have contemplated the situation, I have had something of an eerie feeling’ (ibid.: 21).

But we are still grappling with these questions, albeit extended to include parallel problems about intentional mental concepts! Moreover, our attempted responses to this form of problem seem confined to movements of thought—a “gangway with rigid walls”—exactly parallel to those displayed at *PI* §§185-188. *Either*, events in

consciousness (“what is present to us in an instant”) can only be “externally”, or contingently, related to subsequent performance (“the ‘use’ which is extended in time”), in which case the essential constitutive connectivity is lost, and with it, the very possibility of semantic and mental content “seemingly vanishes into thin air” (Scylla: conceptual nihilism). *Or*: we are led to conceive the semantic or mental content as an ultra-objective, superlative infinite rule-as-rails (or its corresponding counterparts) which is, as Wittgenstein asserts at *PI* §221, clearly ‘a mythological description of the use of a rule’ (Charybdis: Platonist mythology).

Yet, as Wright recognises: ‘We cannot leave matters like that, however, for now it is apt to seem utterly mysterious how the connection—between the prior intention [for instance] and the performance which implements or goes against it—is forged at all’ (2001e: 128). But what, then, are we to do? How are we meant to move ourselves forward—rather than, as at present, seemingly jogging endlessly on the spot? What, it seems, is required as a way of avoiding the ceaseless pendulum swings between Scylla and Charybdis—and therefore to dispel the mystery surrounding our ability to speak, think and act according to the requirements of meaning, mental content, conditions of correctness and constitutive connections—is to find some kind of *third way* of conceiving such ordinary phenomena—some way of steering, that is, a safe course between Scylla and Charybdis.

In Wright’s reading, as I have just endeavoured to make clear, the objectivity of meaning and of intentional mental content is identified with the superlative mythology, such that rejection of the latter is, for him, *eo ipso* rejection of the former. But Wright does not wish to bite the irrealist bullet, so Constructivism is his attempt to re-establish some revised notion of the objectivity of meaning and of intentional mental content. Wright’s Constructivism is an attempt to show in what sense an expectation, say, can be said to be available for us, but in such a way that we fall prey neither to Scylla (conceptual nihilism) nor to Charybdis (Platonistic mythology), via the idea that the intention is *determined*, or *constituted* by our self-ascriptions themselves.

Forging ahead towards his proposed “third way” between Scylla and Charybdis, Wright reflects that what must be rejected is the shared assumption which is common to both epistemological paradigms: ‘the assumption that there must *be* a substantial epistemology of intentional states, a mode of cognitive *access* to those states which is distinctively available to their subject and which is somehow able to measure up to the epistemic security with which sincere avowals of intentional states are standardly credited’ (ibid.) The “third way” which Wright proposes as a way of avoiding this dilemma is to attempt an account of intentional mental concepts according to a putative Constructivist epistemological paradigm: the special authority of the relevant class of mental concepts is, Wright contends, to be explained by the fact that our self-ascriptions of such concepts themselves *determine* the content of the intentional mental states: ‘the authority standardly granted to a subject’s own beliefs, or expressed avowals, about his intentional states is a *constitutive* principle’ (2001i: 312). Recall that Wright had urged that semantic and mental concepts must be seen ‘as plastic in response to speakers’ continuing performance’—that both forms of content must be seen as ‘ever open to ongoing linguistic behaviour’ (2001c: 72). ‘Such metaphors are,’ Wright acknowledges, ‘unsatisfactory’, and ‘are no substitute for a sharp account’ (ibid.). Thus, Wright is in need of a more substantial account of how the Constructivist theory is meant to work.

Beginning to fill in the details, Wright reflects that the metaphors of the “plasticity” and “openness” of semantic and mental content paint a picture of ‘ourselves as perennial creators of our concepts’; but, he suggests, ‘not in the style of conscious architects but just by doing what comes naturally’ (2001c: 78). ‘There is no essential inner epistemology of rule-following,’ Wright urges: ‘To express the matter dangerously, we have *nothing* “in mind” (2001f: 167). Rather, the ‘very existence of our concepts’, Wright asserts, depends on...*activity*’ (2001g: 156). Understanding involves the acquisition of ‘certain constitutive *abilities* which [the subject] did not possess just beforehand’ (2001e: 135). As would be expected, given our earlier discussions, however, Wright problematically conflates terms such as “ability”, “activity”, “practice” and “capacity” with dispositional inclinations, or subrational propensities. Thus, Wright conceives “capacity” as ‘a state of readiness for appropriate use’ (2001g: 155). It is undeniable that Wittgenstein does make some kind of appeal to abilities and practice in his discussion of our use of language. For instance, at *PI*

§180, Wittgenstein reminds us that we judge whether an exclamation such as “Now I understand!” ‘is rightly employed by what [the learner] goes on to do’. And, at *PI* §202, Wittgenstein affirms that “‘obeying a rule’ is a practice’, and, at *PI* §199, states that: ‘To understand a language means to be master of a technique’. But as I hoped to have shown, a whole lot hinges on how, precisely, we conceive such activity.

We will return to consider Wittgenstein’s appeal to activity in Chapter 3. For now, however, we will recall that Wright’s Constructivist conclusion is that the ‘authority which our self-ascriptions of meaning, intention, and decision assume is not based on any kind of cognitive advantage, expertise or achievement’ (ibid: 137-138). Upon what, then, is it based? What is Wright’s “third way”? He explains:

...it is as it were a *concession*, unofficially granted to anyone whom one takes seriously as a rational subject. It is so to speak, such a subject’s right to declare what he intends, what he intended and what satisfies his intentions; and his possession of this right consists in the conferral upon such declarations, other things being equal, of a *constitutive* rather than a descriptive role’ (ibid: 137-138).

It is the “subject’s right to declare what he intends, what he intended and what satisfies his intentions”... It is, Wright explains, ‘simply basic to the component ascription of the [psychological-intentional] attitudes that, absent good reason to the contrary, one must accord correctness to what a subject is willing to avow’: the subject’s ‘active self-conception, as what she is willing to avow, must be deferred to’ (2001k: 369). But, given that Wright equates the idea that an intention can, so to speak, determine those performances which accord with its content (and those which do not) with superlative mythology, then, Wright concludes:

...the authority of sincere self-ascription does not depend upon the presumption that it has somehow—in general, *per impossible*—taken account of such (relevant) sayings and doings. Still less is it based on some phenomenological occurrence which somehow guarantees their accord with the state ascribed (2001: 87).

So, what *does* supposedly serve to bring the essential constitutive connections back into the picture? At this juncture, the “first-personal” and “third-personal” prongs of Wright’s reading of Wittgenstein are fused: our practices of self-ascribing intentions ‘depend on certain deep contingencies’: ‘the contingency that we are, each of us, ceaselessly but...subcognitively moved to opinions concerning our own intentional states’ (2001i: 313). It is, Wright claims:

...a fundamental anthropological fact about us that our initiation into the language in which these concepts feature results in the capacity to be moved, who knows exactly how, to self-ascribe states of the relevant sorts—and to do so in ways which...tend to accord with the appraisals which others, similarly trained, can make of what we do (2001f: 140).

But it should be clear by now that such a picture does *not* serve to rehabilitate the required constitutive connections. For insofar as we are viewing the “subcognitive propensities” to be “moved, who knows exactly how”—somehow (“perhaps”): “I’ll get back to you...”—as merely brute, the constitutive connections vanish altogether, and we have fallen prey to Scylla. And when this is the case, the resulting overall picture seems little different from the basic picture which we earlier found to be operative in Kripke’s community appeal. Thus, ‘all we can say’ about following ‘unhesitatingly and automatically’ according to a determinate rule, according to Kripke’s picture, is ‘that the [subject] is licensed to follow the rule as it strikes him’ (Kripke 1982: 89, 15, 89). That Wright’s thought is sometimes captured by Scylla, is shown in his claim that Wittgenstein’s later philosophy presents a picture of ‘a child’s learning of a first language’

which is ‘assimilated not to the mapping of symbols onto concepts’—the view of training implied by the Platonist mythology (“there only seem to be these two options...”)—but, rather, to *training*, in the same sense in which an animal is trained in certain routines’ (2001i: 309).

Kripke describes ascriptions of meaning as ‘having no justification other than the speaker’s inclination on that occasion’ (ibid: 74, fn. 63); the subject, we are told, is ‘licensed to give, without further justification, the answer that strikes him as natural and inevitable’ (ibid: 88); the subject’s inclination to react in a particular way ‘is to be regarded as primitive’ (ibid: 91). Compare Wright on the “fundamental anthropological facts”, the “deep contingencies”: “It is so to speak, such a subject’s right to declare what he intends, what he intended and what satisfies his intentions...””; the subject’s “active self-conception, as what she is willing to avow, must be deferred to”. But Wright’s updated picture seems no less bound to the picture of the idiolectic rule-follower, which Wright, recall, had earlier characterised as a ‘personal semantic contract’: ‘I can be certain, we would ordinarily allow, of the rule I am following and of what it requires me to do’ (2001b: 31, 22; quoted in Chapter 1, §4). In the updated picture, too: ‘A subject’s sincere dispositions of avowal concerning his intentional states—or, better, his beliefs concerning them, so identified—...stand by default...’; ‘...truth is the default position, so to say. They count as true provided that we hold them’ (2001i: 313). Similarly, ‘the authority of our opinions concerning what would currently comply with [our previous intentions], may be viewed as likewise a matter of default correctness’ (ibid: 314-315). But recall that in Chapter 1 we heard Wright’s objection to the notion of any such “personal semantic contract”: “correctness” simply collapses into whatever application I am disposed (by my “subrational propensities”) to make of a concept: I simply find myself ‘with a sincere disposition’ to apply a concept or rule, or to self-ascribe an intention, in a particular way and, as Wright had earlier said, ‘that is the whole of the matter’ (2001b: 30).

What, then, underwrites the distinction in the putatively updated picture? Although the subject is, by default, allowed to call it as she sees it in regard to her intentions,

nevertheless, Wright claims, those self-ascriptions are ‘defeasible’ and ‘may on occasion *reasonably* be discounted by a third party’; ‘...and insofar as [an avowal of intention] may be criticisable, the basis for the criticism may only be constituted by states of affairs that were not salient, or even did not exist, at the time of avowal—par excellence, the subject’s subsequent behaviour’ (2001f: 138). The idea here, then, appears to be this: let’s say, for instance, I am suddenly hit (“in a flash”) by an occurrent phenomenon of consciousness—a “resolve to make a Caesar salad.” Now, according to Wright’s picture, I cannot know at the time of this occurrence that the identity of my intention is *to make a Caesar salad*. Such an occurrent phenomenon of consciousness cannot contain within itself, so to speak, such a constitutive connection with a determinate performance or state of affairs (my making a Caesar salad). If I explicitly judge, however, of the occurrent phenomenon of consciousness, “I intend to make a Caesar salad” then—hey presto!—that’s the identity of my intention. That it is the intention *to make a Caesar salad* is, it is claimed, *determined* or *constituted* by such explicit self-ascriptions. And that I can thus be taken to be authoritative on such matters is, according to Wright’s picture, a *concession*, unofficially granted to anyone whom one takes seriously as a *rational* subject’ (2001e: 137).

The subject’s ‘active self-conception, as what she is willing to avow, must be deferred to’ (2001k: 369). Although this avowal of mine will be granted default status by my fellows, however, the avowal is, nevertheless, defeasible in light of what I go on to do: thus, the avowal, Wright claims, “may on occasion *reasonably* be discounted by a third party”. And the kinds of *reason* that might cause a third-party to retract the default status of my avowal is based on what I go on to say and do. Thus, if I end up cooking a lentil curry, rather than preparing a Caesar salad, then a third-party may *reasonably*, it is supposed, retract the deference otherwise granted to my avowal. But this is to fall squarely into Charybdis. The default status of my avowal is overridden by a third-party if I don’t go on *to make a Caesar salad*. And Wright is here making liberal use of concepts—reason and reasonableness—which he elsewhere admits cannot be relied upon on pain of presupposition.

On the other hand, if Wright wishes to retract these problematic presuppositions, then the resulting picture of the subject being granted “concessions” and “deference”, it seems to me, will be little different from that of Kripke’s picture of idiolectic “rule-followers”, brutally mouthing off in anyway they see fit. Consequently, it is then difficult to see the retraction of the otherwise default “concession” or “deference” as any different, at bottom, to Kripke’s idea that, until such times as deviations are recognised, “self-ascribers” are taken ‘*provisionally* into the community’ (Kripke 1982: 112; emphasis added). It is but then a small step to a progressive fall from grace: the subject, but a moment ago granted supreme authority over her own avowals, has had that concession retracted and is no longer given the deference granted to “a friend of ours”.

That Wright’s picture *does* fall prey to Charybdis (when it’s not otherwise busy falling prey to Scylla) is demonstrated in Wright’s characterisation of the background which is supposed to frame the automatic deference, or concession granted to the authority of an individual’s avowals: ‘It is part of regarding human beings as *persons, rational reflective agents*, that we are prepared to *ascribe intentional states to them*, to try to *explain and anticipate their behaviour in terms of the concepts of desire, belief, decision and intention*’ (2001e: 140; all emphases added). But all this is precisely what was meant to be being explained to us in the first place! Elsewhere, Wright says: ‘Since the *telos*, in the most general terms, of the practice of ascribing intentional states to oneself and others is mutual *understanding*’—which, again, presupposes a concept which we began trying to explain (2001i: 313; first emphasis in original). And, again: ‘...the constraint of having one’s sincere self-ascriptions *make sense* in the light of one’s outward performance in effect supplies the *standard of correctness* for one’s impressions of self-knowledge of *meaning*...’ (2001: 87; all emphases added). Once more: ‘[the subject’s] views are to be accounted *correct* unless a more satisfying *rationalisation* of (*germane aspects*) of her deportment can be constructed by discarding them...’ (2001i: 314; all emphases added).

Of course, Wright would contend that he is *not* presupposing such concepts, but is revising them in the light of our underlying “subrational propensities”, which are what are being relied upon in order to explain such concepts. Thus, Wright tentatively refers to the

background here described as ‘a rationalising system of sorts’ (ibid: 313). Yet, as I have already argued, the appeal to “subrational propensities” is no less wedged within the gangway between the two rigid walls of Scylla and Charybdis.

Wright proposes that his Constructivist theory can be based on an analogy between semantic and mental content and Lockean secondary qualities, ‘with secondary qualities in turn viewed as distinguished by their being response—, or more specifically judgement—dependent’ (2001: 89). ‘One who thinks of colours as secondary,’ Wright explains, ‘thinks roughly this: that the relationship between what colour something has and how it visually seems to normally sighted humans is a constitutive relationship’ (2001: 86). Now, without delving too deeply into the detail of Wright’s analogy (see 2001g), there are a couple of fundamental problems with this strategy. Firstly, and most obviously, the Constructivist theory is here being presupposed in the case of secondary qualities. So, even assuming that Wright could successfully ground Constructivism on behalf of meaning and intention on the analogy with secondary qualities, it would then remain for Wright to substantiate Constructivism in the account of secondary qualities—a highly controversial matter in its own right.

Secondly, and more problematically, Wright aims to ground his theory on the suggestion that judgements of meaning and intention, like judgements of colour, but in contrast with Lockean primary qualities, fail what Wright calls the *order-of-determination* test (discussed at length in 2001g). Wright explains that “the order-of-determination test” concerns ‘the relation between *best* judgements—judgements made in what are, with respect to their particular subject matter, *cognitively ideal* conditions of both judge and circumstance—and truth’ (ibid: 192). Furthermore, Wright specifies that passing the order-of-determination test depends on the contrast between ‘judgements among which our best opinions *determine* the extension of the truth predicate, and those which they at most reflect an extension determined independently’ (ibid.). Wright explains:

Passing the test requires that there be some content in the idea of best judgements *tracking* the truth—the determinants of a judgement's being true and of its being best have to be somehow independent. Truth, for judgements which pass the test, is a standard constituted independently of any considerations concerning cognitive pedigree... (ibid.)

Wright's contention is, then, that our judgements of primary qualities—paradigmatically shape—track their extensions in the superlative, Platonistic manner. Compare Wright's characterisation of cognitively ideal conditions with his rejection of 'an absolute conception of truth: 'a conception absolved from all practical controls' (2001c: 77). The argument is then that judgements of secondary qualities (and then, in their turn, of meaning and of intentional mental states) fail to track any extensions in this manner, but, rather, are *extension-determining*:

...For judgements which fail the test, by contrast, there is no distance between being true and being best; truth, for such judgements, is constitutively what we judge to be true when we operate under cognitively ideal conditions' (2001g: 192).

But when we have probed this picture, what we found was little different from the stark, Kripkean picture of babbling brutes, granting each other the liberty to dispose towards their own private rules, but withdrawing such deference whenever one deviates too unacceptably from the brute inclinations of others (he is "no friend of ours"). But, as we have continually seen, this stark, dispositional waste land is haunted by superlative mythology, for, in order for the required distinction to work, a Platonic conception of the objectivity of our judgements of primary qualities is already being presupposed. Now that is problematic enough, but what, then, is supposed to supply the required contrast when we turn our focus on to the objectivity of meaning? For the Platonist mythology has already been rejected (in fact, it was this very rejection which prompted the need for a Constructivist theory of meaning in the first place). So, *either* Wright has to presuppose

a mythological conception of the objectivity of meaning in order to form the contrast required by the order-of-determination test; or, failing that, the question is being begged in favour of the claim that judgements of meaning are *extension-determining*—exactly the Constructivist claim which the order-of-determination test was being called upon to substantiate!

Conclusion

Despite the ingenious twists and turns on display in this chapter, we are no further than where we began, with the dilemma delineated by Wittgenstein at *PI* §§185-188: on the one hand, there are queer superlative rules and, on the other, the conceptual nihilism of babbling brutes—mere “jack-in-the-box” dispositions to react, and a community of collusion granting deference to whatever it is that each individual is merely disposed to do. On the one hand, Wright protests, ‘the price of objective meaning’ seems to be ‘an *absolute conception* of truth: a conception absolved from all practical controls’ (2001c: 77). But, on the other hand, as Kripke articulates, the vehicles of language ‘are visible (or audible or palpable), concrete phenomena—marks or diagrams on paper’ (Kripke 1982: 106). But conceived *merely* as such, then such concrete phenomena are seemingly open to infinite interpretation. As Kripke summarises, ‘no matter what is in my mind at a given time, I am free in the future to interpret it in different ways’ (ibid: 107). But now, on pain of relapsing once more into superlative fantasy, we are thrown again into the threat of a regress of interpretations, and its attendant conceptual nihilism.

Similarly, Wright begins his reading of Wittgenstein’s diagnosis of misunderstanding at *PI* §201 by assuming that Wittgenstein allows ‘apparently, that there are indeed an ineliminable multiplicity of conflicting possible interpretations’ of the expression of a rule (2001d: 124). Once this assumption is made, however, Wright is led to identify rationality with interpretation. Thus, when Wright turns to consider a positive account of the obverse of the misunderstanding—‘a way of grasping a rule which is *not* an *interpretation*’ (*PI* §201)—the substance he reads into this is that ‘something other than the exercise of interpretative (rational) faculties enters into the capacity to “read” another’s behaviour’ (ibid.). And this something is, according to Wright, “subrational propensities” or “classificatory dispositions”, which supposedly guide our linguistic and mental behaviour. But, as I argued in §2a, appeal to such notions merely leads us back into the clutches of the very dilemma we were hoping to escape.

The epistemological branch of Wright's reading, as we saw in §2b(i), begins in similar fashion to the metaphysical branch, described a moment ago. That is to say, we are led once more into a further instance of Wittgenstein's misunderstanding: 'an inclination to say: any action according to the rule is an interpretation' (*PI* §201). Despite justified protestations against the reductionist "prejudice" of Kripke's sceptic, Wright still assumes that any response to the sceptic, on pain of begging the question, must be given in terms of an *interpretation*, in the sense of 'the substitution of one expression of the rule for another' (*PI* §201). That is to say, Wright's strategy still assumes that one cannot supply a direct response to the sceptic, on pain of begging the question, by simply citing the fact that "I meant, or mean addition by 'plus.'" Rather, Wright submits, 'the challenge is to recall some *independently characterised* fact' (2001g: 176; emphasis in original).

Wright seeks in avowals of intention the independently characterised fact, and describes such practices of avowal in terms of *activity*. And use of language as activity is no doubt a common theme within the *Investigations*, as well as other later writings. But Wright conceives such activity in terms of dispositions, whose avatar, for Wright, are character traits—courage, for instance. As I argued in §2b(ii), whilst it is not wrong *per se* to class such qualities of character as dispositional, Wright's conception of the dispositional-behavioural-physicalist paradigm of the mental is deeply problematic. When such characteristics are called upon to play a role within an *explanatory* framework—for the purposes of *explaining* the putatively mysterious properties of intentional mental states and of linguistic behaviour— we are naturally drawn to conceive such qualities of character— those which are being marked as emblematic of the dispositional paradigm—as *mere* or *brute* dispositional propensities to react. Thus, Wright remarks, this paradigm of the mental is 'an option which Wittgenstein expects, writing when he did, will naturally take a behaviourist shape' (2001k: 359). In the current philosophical climate, however, Wright suggests that this paradigm will 'more likely [be] physicalist' (*ibid.*). But, as we have seen, insofar as we conceive such dispositional propensities as merely brute (or purely physical, if we are swept up in prevailing tides) then we fail to preserve the essential constitutive connection between, say, *courage* and courageous behaviour. The relation between such merely brute propensities to react and outward behaviour could only be contingent, not constitutive. But in a recoil from these,

ultimately nihilistic, implications we are led to conceive such dispositional states as mysterious *quasi-physical facts*. Thus we are once more assailed by the dilemma.

Yet Wright's contrasting putative paradigm of the mental is no less problematic, as we saw when we considered McDowell's objections to the paradigm whose avatar is occurrent states, but conceived as "pure", or "genuine": that is to say, without involving the subject's conceptual capacities—what Wright refers to as 'mental events and processes strictly so termed' (2001j: 329). But give this conception, McDowell argues, then it is doubtful 'whether anything could *be* an occurrent phenomenon of consciousness' (1998d: 302). McDowell parodies Wright's *a priori* argument for the case of pain, and which leads to the conclusion that the notion that pain can be an occurrence in consciousness is an incoherence. What needs to be rejected, however, is not the idea that an intentional mental occurrence (being struck by a thought, say) cannot be made available to introspectable awareness; what needs to be rejected, rather, is the picture of "pure", or "genuine" occurrences of consciousness, those which make no demands at all on our conceptual resources—"private objects".

Yet, as we saw in §2c, Wright can only conceive occurrent phenomena of consciousness in this problematic fashion, such that rejection of this picture is *eo ipso* rejection of the idea that occurrent mental phenomena can be made available to self-conscious awareness. Before we turn to the real import of Wittgenstein's argument in Chapter 3, however, consider the following remarks, which suggest that the real problem is not the reality of introspectable intentional states, as Wright contends, but of the superlative conception of their instantiation:

I want to say: "If someone could see the mental process of expectation, he would necessarily be seeing *what* was expected."—But that is the case: if you see the expression of an expectation, you see what is being expected. And in what other way, in what other sense would it be possible to see it? (*PI* §452).

“It’s as if we could grasp the whole use of a word in a flash.”—And that is just what we say we do. That is to say: we sometimes describe what we do in these words. But there is nothing astonishing, nothing queer, about what happens. It becomes queer when we are led to think that the future development must in some way be present in the act of grasping the use and yet isn’t present.—For we say that there isn’t any doubt that we understand the word, and on the other hand the meaning lies in its use (*PI* §197).

Chapter 3 “Turning Our Examination Around...”

...Something that we know when no one asks us, but no longer know when we are supposed to give an account of it, is something that we need to remind ourselves of. (And it is obviously something of which for some reason it is difficult to remind oneself.)

Philosophical Investigations: §89

...“Imaginary sticks can draw real blood”. What is vital is not to bleed to death, but to learn from the wounds of confusion. These wounds may go deep, but fruitful new understandings emerge through the process of healing them...

Dewi Phillips

Introduction

As we saw in the previous chapter, Kripke’s and Wright’s respective, but complementary, responses to Wittgenstein’s discussion of rules and of rule-following fail to escape the dilemma of *PI* §§185-188. One commentator who, I believe, has understood Wittgenstein’s insights, and whose thought has managed to grasp the way to escape the dilemma, is John McDowell. Unfortunately, however, McDowell’s early response to Wittgenstein’s paradox, published as “Wittgenstein on Following a Rule” is insufficiently distanced from the misunderstanding which Wittgenstein diagnoses at *PI* §201. Despite attracting justified criticism, what has gone unremarked within the literature is the fact that McDowell comes later to explicitly repudiate his early reading as based on a form of the very misunderstanding Wittgenstein is trying to warn us against. Thus the early paper tends to distract attention away from McDowell’s later insights into Wittgenstein’s diagnosis.

As we shall see in §1, (early) McDowell, not unlike Wright, fails to reflect deeply on Wittgenstein’s misunderstanding—‘an inclination to say: any action according to the rule

is an interpretation' (*PI* §201)—and, instead, moves immediately to import substance into the obverse of the misunderstanding—'there is a way of grasping a rule which is *not* an *interpretation*' (*PI* §201)—in the hope that it will *somehow* open up a "third way" between the two horns of the familiar dilemma. Yet this proposed "third way" (an appeal to communal institutions, or culture as that which will *somehow* bridge the perceived gap between the expression of a rule and appropriate performance) still retains an aura of mystery, and therefore seemingly necessitates further explanatory research. Thus McDowell's early reflections on the issues are insufficiently distanced from Kripke's and Wright's respective readings, and, therefore, his picture of meaning and of rule-following, not unlike those of Kripke and Wright, remains haunted by a vertiginous sense of eeriness related to the apparent radical contingency of the possibility of agreement in individual linguistic reactions.

As we shall consider in §2, McDowell's later reading begins, in contrast to the earlier, not by immediately looking for the substance of the obverse of the misunderstanding, but with greater reflection on the nature and implications of the misunderstanding itself. The fruit of this greater reflection reveals what I will call "the three-step insight" into the misunderstanding of *PI* §201. And this is that, in an attempt to explain the language-game from "outside"—that is, without presupposing meaning and understanding—we are led to conceive the expression of a rule as merely "dead signs," and behaviour made in response to the rule as merely brute. Once this step has been made then it seems that appropriate behaviour can be connected with the expression of the rule *only under an interpretation*. But, given our initial steps, then anything we care to bring forth as a candidate mediating interpretation, and which is supposed to bridge the gulf between the expression of a rule (merely "dead signs"), on the one hand, and (merely brute) behaviour which is in accord with that rule, on the other, will inevitably be conceived in such a way that it, too, requires interpretation. And so this is why we come to face the familiar interpretational regress.

As we shall consider in §2, there are two principal lessons of Wittgenstein's regress of interpretations argument. The first is that we must not allow any such conceptual gap to

open up between the expression of a rule and behaviour which accords or not with the rule. Once we allow such a gap to open ‘it will,’ McDowell warns, ‘be unbridgeable’ (McDowell 2009a: 101). The second principal lesson of Wittgenstein’s arguments brings us eventually to the obverse of the misunderstanding: ‘there is a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call “obeying the rule” and “going against it” in actual cases’ (*PI* §201). For those who are inductees into a particular normative practice, in response to the expression of a rule, they can act immediately on their *understanding* of that rule. Of course, there can be misunderstanding, and there can be failure to understand. But such notions are clearly parasitic on the concept *understanding*. A deviant rule is just that—*deviant*. It is not “the same rule under a different interpretation”, but, at best, a different rule, or a different concept.

But such direct appeals to unmediated understanding of a word or rule are not problematic for Wittgenstein, for the purpose of such appeals are not connected with any explanatory ambitions. Rather, their function in Wittgenstein’s dialectic are as “reminders” (*PI* §127). And the purpose of such reminders is precisely to ensure that we do not take the fateful first step (propelled by our explanatory ambitions) into a picture of normative inertness. Once behaviour is viewed as merely brute, and expressions of rules as merely “dead signs” then eventual regress is inevitable. And any attempt to squeeze normativity back into the picture inevitably deforms such normativity, such that the demandingness of rules (“the hardness of the logical must”: *PI* §437) can only show up as extraordinarily queer, spooky, and mysterious—as ultra-objective superlative rules (the rule as a “despotic demand”: *PI* §437). As Wittgenstein asserts at *PI* §654: ‘*Our mistake is to look for an explanation...*’

Consequently, as (later) McDowell comes to recognise, Wittgenstein’s appeal to *custom*, *forms of life*, and so forth, are *not*, as they are often conceived amongst commentators—including (early) McDowell—programmatic gestures towards a naturalistic *explanation* of our linguistic and mental lives. To think that we require further constructive philosophical explanations is still to be in the grip of the misunderstanding. We do not need any explanatory “third way” between Scylla and Charybdis. We only need to recognise that there is but *one way*: behaviour which displays unmediated

understanding of the meaning of a word or rule. If we remind ourselves of that, we do not need to step into Wittgenstein's misunderstanding. And, if we do not step into the misunderstanding, then we are no longer faced with the intolerable choice between Scylla and Charybdis. Therefore, we do not need a third way. What we require is to pay close attention to 'the spatial and temporal phenomenon of language'—unmediated understanding of a word or rule, and the circumstances which surround such norm-governed behaviour (*PI* §108).

§1 McDowell's Early Reading (or: More of the Same...)

Like Wright, but in contrast to Kripke, McDowell wishes to respect Wittgenstein's explicit denunciation of the regress of interpretations paradox at *PI* §201 as founded on misunderstanding. Again, in concert with Wright, McDowell perceives that the paradox is but one horn of a dilemma, the other horn of which is the Platonistic mythology. McDowell rightly distances himself from Kripke's reading, which takes the fact that Wittgenstein is clearly hostile to the mythology as an argument for the other horn—supposed acceptance of the implications of the paradox: 'the attack on the mythology is not support for the paradox, but rather constitutes, in conjunction with the fact that the paradox is intolerable, an argument against the misunderstanding' (McDowell 1998a: 231). The misunderstanding referred to here by McDowell is that which Wittgenstein diagnoses as the source of the paradox (and the accompanying Platonistic mythology): 'the inclination to say: any action according to the rule is an interpretation' (*PI* §201). 'The mythology is wrung from us', McDowell correctly perceives, in our need to avoid the paradox...only because we fall into the misunderstanding' (McDowell 1998a: 231). As we have seen, this certainly fits the movements of thought—"the gangway with rigid walls"—within which Kripke's and Wright's readings operate, constantly shifting between the two horns: *either* meaning and intentional mental states vanish in face of an indefinite number of interpretations, which can't be determined by any further interpretations without threat from the regress of interpretations paradox; *or* the determinacy of meaning can seemingly only show up in the guise of the conception of a superlative, ultra-objective rule (a special regress-stopping interpretation, or self-interpreting rule).

In order to bring a halt to these incessant movements of thought between Scylla (conceptual nihilism) and Charybdis (Platonistic mythology), McDowell, again not unlike Wright, perceives that we must find a way of discarding the assumption upon which the dilemma depends—"the assumption that understanding is always interpretation" (*ibid*: 238). We must, it seems, begin by trying to understand what Wittgenstein means by his

insistence that ‘there is a way of grasping a rule which is *not* an *interpretation*’ (*PI* §201)⁴⁷. And the clue to what Wittgenstein is getting at, McDowell suggests, is provided in the very next remark of the *Investigations*, where Wittgenstein states: ‘And hence also “obeying a rule” is a practice’ (§202). ‘That is’, McDowell recommends: ‘what mediates the inference (“hence also”) is this thought: we have to realise that obeying a rule is a practice if we are to find it intelligible that there is a way of grasping a rule which is not an interpretation’ (McDowell 1998a: 238). This seems eminently sensible, but we need to proceed with caution, given that this train of reasoning, thus far, merely echoes Wright’s reading which, as we saw in §2a of Chapter 2, did not in the end enable us to avoid the dilemma. Wright, recall, also reads “a way of grasping a rule which is not an interpretation” as a gesture towards activity, or practice: ‘Ascribing understanding is ascribing (something like) an ability’ (Wright 2001i: 296); ‘The very existence of our concepts depends on...*activity*’ (Wright 2001f: 156). Yet, everything hinges on how we conceive this appeal to practice, or activity.

In order to better discern what Wittgenstein is getting at with his appeal to practice, (early) McDowell reflects upon the exchange at *PI* §198, which begins with the interlocutor making reference to the regress of interpretations paradox:

“Then can whatever I do be brought into accord with the rule?”

—Let me ask this: what has the expression of a rule—say a sign-post—got to do with my actions? What sort of connexion is there here?

—Well, perhaps this one: I have been trained to react to this sign in a particular way, and now I do so react to it.

⁴⁷ Although a perfectly intelligible move to make, I will, in §2, be suggesting that this move eventually proves fateful for McDowell’s reading.

Teasing out the import of this dialogue, McDowell remarks: ‘When I follow a sign-post, the connection between it and my action is not mediated by an interpretation of sign-posts that I acquired when I was trained in their use. I simply act as I have been trained to’ (McDowell 1998a: 239). Again, this mirrors Wright, who is keen to avoid the picture of our having imbibed some kind of rule which, as Kripke puts it, is ‘engraved on my mind as on a slate’, and which comes readily as an image of training when our thought moves within the orbit of the Platonistic mythology (Kripke 1982: 15). Recall that the sense in which Wittgenstein conceives our initial induction, as children, into linguistic practice is, according to Wright, a matter of ‘*training*, in the sense in which an animal is trained in certain routines’ (Wright 2001i: 309).

McDowell next considers that Wittgenstein, in an ‘attempt to exorcize the insidious assumption that there must be an interpretation that mediates’ between the expression of a rule and subsequent performance, makes remarks along the lines of: “I obey the rule *blindly*” (*PI* §219) (McDowell 1998a: 239). McDowell also quotes RFM VI-35: ‘If I am drowning and I shout “Help!”, how do I know what the word Help means? Well, that’s how I react in this situation.—Now *that* is how I know what “green” means as a well and also know how I have to follow the rule in the particular case’; and he links the expression “Well, that’s how I react...” with *PI* §217: “...I am inclined to say: ‘This is simply what I do’” (McDowell 1998a: 240, fn. 29). McDowell further connects all these remarks with *PI* §241, where Wittgenstein famously proclaims: ‘It is what human beings *say* that is true and false; and they agree in the *language* they use. That is not agreement in opinions but in form of life.’ As McDowell reads him, Wittgenstein’s point in this passage ‘is that an opinion is something for which one may reasonably be asked for a justification; whereas what is at issue here is below that level—the “bedrock” where “I have exhausted the justifications” and “my spade is turned”’ (*PI* §217) (McDowell: 1998a: 240; *cf.* *OC* §204).

Now, most of these remarks form a similar assemblage to that which Kripke gathers as evidence that Wittgenstein is promoting a perspective on rules and of rule-following which recommends that: ‘Nothing justifies a brute inclination to answer one way rather than another’ (Kripke 1982: 15). The paradox has supposedly shown that one’s following

according to a rule is based on nothing over and above ‘simply following an unjustifiable impulse’; one’s confident grasp of a rule ‘a mere jack-in-the-box unjustified and arbitrary response’ (ibid: 18, 23). But, as was discussed earlier, in face of the threat of Platonistic mythology, Kripke’s Wittgenstein unashamedly grasps hold of Scylla—all is interpretation, further interpretation leads to an interpretational regression, and on from there to conceptual nihilism—and so we need to be careful how far we swing over in this direction, when we justifiably recoil from the image of imbibing, through training, a superlative rule: “engraved on my mind as a slate”. So, if we are going to be able to find a *third way* between Scylla and Charybdis, we cannot conceive “bedrock” in these outright bare terms, where all meaning and mind goes dark. So far, this again accords with Wright, who has no intention of biting the irrealist bullet if he can help it, and so wants to be able to *somehow* (“I’ll get back to you...”) keep correctness conditions or constitutive connections in the picture. With these thoughts in mind, McDowell points out that, although Wittgenstein ‘attributes the “bedrock” use of expressions as “without justification”, he nevertheless insists...[PI §289]: “To use an expression without a justification does not mean to use it without right”’ (McDowell 1998a: 241). And this remark is an expression of a resolve to avoid Scylla through ensuring that we, as McDowell says, ‘prevent the leaching out of norms from our picture of “bedrock”—from our picture, that is, of how things are at the deepest level at which we may sensibly contemplate the place of language in the world’ (ibid: 241-242). In this regard, McDowell quotes RFM VI-28: ‘Following according to a rule is FUNDAMENTAL to our language-game’.

As Wright acknowledges, in our attempts to explain such concepts as meaning and mind, objectivity and understanding, we must *somehow* make room in our account for normative notions of accord, for speech and deed undertaken in accordance with conditions of correctness, of the distinction between seems right and is right. In order to keep our conceptual and rational capacities in the picture—in face of the problem of underdetermination and the regress of interpretations paradox, on the one side, and superlative mythology on the other—Wright, we will recall, reads Wittgenstein’s injunction to recognise a “way of grasping a rule which is not an interpretation” as a gesture towards the idea of ‘a whole plethora of natural *classificatory* dispositions’ which

ground our linguistic behaviour, of 'sub-*rational* propensities towards conformity of response'; and he talks cagily of "disposition-like connections" (Wright 2001e: 124). This, according to Wright, points to the *third way* which does not succumb to the Scylla of merely brute, purely physical dispositions which, as he recognises, could only sustain external, or contingent connections, whereas semantic and mental concepts are constitutively connected to appropriate performance—between the rule expressed by the sign-post, on the one hand, and appropriate performance in response, on the other (going to the right, if the sign points to the right, for instance). Speaking of this predicament, the exchange at *PI* §198 continues:

"But that is only to give a causal connexion: to tell how it has come about that we go by the sign-post; not what this going-by-the-sign post really consists in."

—On the contrary; I have further indicated that a person goes by a sign-post only in so far as there exists a regular use of sign-posts, a custom.

In order to warn us off thinking that, when searching for "bedrock," 'we can dig down to a level at which we no longer have application for normative notions (like "following according to the rule")', McDowell reads Wittgenstein as gesturing towards a *third way* between Scylla and Charybdis through the image of 'training ...[as] initiation into a custom':

If it were not that, then the account of the connection between a sign-post and action would indeed look like an account of nothing more than brute movements and its causal explanation; our picture would not contain the materials to entitle us to speak of following (going by) a sign-post (McDowell 1998a: 239).

Summarising the situation, McDowell states: ‘Wittgenstein’s problem is to steer a course between Scylla and Charybdis. Scylla is the idea that understanding is always an interpretation. This idea is disastrous because embracing it confronts us with the dilemma...’ (ibid: 242). Now, we will recall from Wright’s picture that, having seemingly avoided the temptations of Scylla by invoking “sub-*rational* propensities” and “*classificatory* dispositions”, and in order to sustain the explanatory ambitions we have projected onto the notion of “grasping a rule which is not an interpretation,” at the exact same time we need to ensure that they are “sub-*rational propensities*” to react—“*classificatory dispositions*”. That is to say, according to Wright: ‘something other than the exercise of interpretative (rational) faculties enters into the capacity to [understand] linguistic behaviour’ (Wright 2001e: 124). Descriptions at the level of Wittgenstein’s “bedrock”—the putative *third way* between Scylla and Charybdis—according to Wright, must take account of the implication that ‘a *merely rational* methodology can indeed yield no determinate conclusions in regard to the disclosure of the meaning of a speaker’s words’ (ibid; emphasis in original). We must accept the ‘deep contingency’ at bedrock, ‘the contingency that we are, each of us, ceaselessly but...subcognitively moved’ (Wright 2001i: 313). Staking out a corresponding *third way* in his reading of Wittgenstein, McDowell remarks:

[T]he key to finding the indispensable middle course is the idea of a custom or practice. How can a performance both be nothing but a “blind” reaction to a situation, not an attempt to act on an interpretation (avoiding Scylla); and be a case of going by a rule (avoiding Charybdis)? The answer is: by belonging to a custom (*PI* §198), practice (*PI* §202), or institution (RFMVI-31)...Wittgenstein’s point is that we have to situate our conception of meaning and understanding within a framework of communal practices (McDowell 1998a: 242-243).

So, not unlike Kripke’s Wittgenstein, McDowell’s Wittgenstein makes a “community appeal” in face of the threat of paradox (or, more accurately, the threat of paradox together with the threat of Platonistic mythology, on McDowell’s early reading): ‘But,’

McDowell warns, 'it makes a difference how we conceive the requirement of publicity to emerge' (ibid: 243):

In my reading, it emerges as a condition for the intelligibility of rejecting a premise—the assimilation of understanding to interpretation—that would present us with an intolerable dilemma. So there are three positions in play: the two horns of the dilemma, and the community-oriented conception of meaning that enables us to decline the choice (ibid.).

Kripke, McDowell asserts, conflates the first horn of the dilemma (the paradox of *PI* §201) with the community-oriented conception (Wittgenstein's conclusion, on McDowell's reading), such that rejecting the Platonistic mythology (the ultra-objective, "superlative fact" of *PI* §192) entails acceptance of the paradox. McDowell also distances himself from Wright, for whom the appeal to community 'emerges as the only alternative left, after the notion of idiolectic understanding has been scotched' by the argument at *PI* §258 (that idiolectic understanding leaves no possibility for the required "is right/seems right" distinction): 'Wright makes nothing of Wittgenstein's concern—which figures at the centre of my reading—to attack the assimilation of understanding to interpretation' (ibid: 260, 261). But McDowell is here objecting to Wright's earlier, crude (constitutive) Communitarianism. Yet, as we have seen, Wright's *updated* response to the dilemma *does*, like that of McDowell, aim to respect Wittgenstein's rejection of the conflation of understanding with interpretation. As we discussed in Chapter 2 (§2a), the point of Wittgenstein's appeal, at *PI* §201, to a "way of grasping a rule which is *not* an *interpretation*" is, according to (updated) Wright:

...that something other than the exercise of interpretative (rational) faculties enters into the capacity to "read" another's linguistic behaviour. And the additional something is, crudely, human nature: certain subrational propensities towards conformity of response, towards "going on in the same way", which alone makes possible the formation of the common

conceptual scheme within which our rational capacities can be exercised
(Wright 2001e: 124).

Moreover, not unlike McDowell, Wright views meaning as ‘essentially social’ (Wright 2001g: 173). In order to avoid Scylla and Charybdis, (updated) Wright sees Wittgenstein’s *third way* as pointing to normative linguistic notions as grounded in institutional linguistic *activities*: ‘Language, and all rule-governed institutions...it seems, must be founded in primitive dispositions of agreement in judgement and action’ (Wright 2001f: 167). But these institutional practices, unlike in Wright’s earlier crude Communitarianism, do not *constitute* the standards by virtue of which the performance of individual rule-followers is judged. Rather, the institutional practices are best viewed, according to (updated) Wright, ‘as an expression of certain basic reactive propensities, primitive classificatory dispositions—a common human (or at least cultural) heritage without which our language would fail’ (Wright 2001c: 72). So, McDowell’s objections against (early) Wright miss their mark against (updated) Wright, who does, like McDowell, aim to respect Wittgenstein’s appeal to a “way of grasping a rule which is *not* an *interpretation*”. Moreover, both McDowell and (updated) Wright consider Wittgenstein’s *third way* beyond Scylla and Charybdis to involve *some* kind of gesture towards communal institutions. Thus, McDowell claims that ‘hearing a word in one sense rather than another is hearing it in one position rather than another in the network of possible patterns of making sense that we learn to find ourselves in when we acquire mastery of a language’ (McDowell 1998a: 260).

Yet, at least at this level of detail, this characterisation of the communal third way seems insufficiently distanced from that of (updated) Wright, who talks of ‘a rationalising system of sorts’ (Wright 2001i: 313). That is to say, McDowell’s characterisation of “a network of possible patterns of making sense” is insufficiently distanced from Wright’s contention that ‘it is part of regarding human beings as persons, rational reflective agents, that we are prepared to ascribe intentional states to them, and to try to explain and anticipate their behaviour in terms of the concepts of desire, belief, decision and intention’ (Wright 2001e: 140). But, of course, as I pointed out in Chapter 2 (§2a), this remark is riddled

with notions which we are in the process of trying to explain. Similarly, McDowell's proposed communal third way between Scylla and Charybdis makes use of the notion of "*making sense*" of each other—exactly that which is supposed to be being explained. So how, then, is this appeal to communal patterns of making sense, supposed to provide an *explanation* of our practices of rule-following? Wright, we will recall, claims that it is 'a fundamental anthropological fact about us that our initiation into the language in which these concepts feature results in the capacity to be moved, *who knows exactly how*', by the "subrational propensities" which— *somehow* ("perhaps": "I'll get back to you...")—serve to ground our communal, institutional practices from outside the realm of introspectable self-awareness (Wright 2001f: 140). Similarly, in regard to his communal third-way between Scylla and Charybdis, McDowell sates that: 'Until more can be said about how exactly the appeal to communal practice makes the middle course available, this is only a programme for a solution to Wittgenstein's problem' (McDowell 1998a: 242). In other words: "I'll get back to you..."

Moreover, McDowell's picture of shared agreement looks no less precarious than the 'brute empirical fact' that we are not 'reduced to a babble of disagreement' in Kripke's picture (Kripke 1982: 109, 91):

Wittgenstein warns us not to try to dig below "bedrock". But it is difficult, in reading him, to avoid acquiring a sense of what, as it were, lies down there: a web of facts about behaviour and "inner" episodes, describable without using the notion of meaning. One is likely to be struck by the sheer contingency of the resemblances between individuals on which, in this vision, the possibility of meaning seems to depend, and hence impressed by an apparent precariousness in our making sense of one another (McDowell 1998a: 249-250)⁴⁸.

⁴⁸ Cf. *PI* §635: '...It is as if a snapshot of a scene had been taken, but only a few scattered details of it were to be seen: here a hand, there a bit of face, or a hat—the rest is dark. And now it is as if we knew quite certainly what the picture represented. As if I could read the darkness.' And also *PI* §101: 'the idea now absorbs us, that the ideal "*must*" be found in reality. Meanwhile we do not yet see *how* it occurs there, nor

In an earlier paper “Non-Cognitivism and Rule-Following”, McDowell refers to a passage from Stanley Cavell which talks of the ‘terrifying’ contingency that the practices that go to make up our ‘whirl of organism’ do manage to sustain agreement (McDowell 1998e: 206-207).⁴⁹ McDowell describes ‘a sort of vertigo, induced by the thought that there is nothing that keeps our practices in line except the reactions and responses we learn in learning them. The ground seems to have been removed from under our feet’ (ibid: 207)⁵⁰. McDowell admits that this picture looks like ‘a congruence of subjectivities not grounded as it would need to be to amount to the sort of objectivity we want if we are to be convinced that we are *really* going on in the same way’ (ibid.). Given that this picture, as McDowell admits, ‘seemingly threatens to dissolve the independent truth of arithmetic into a collection of mere contingencies about the natural history of man’, then it can very easily prompt a recoil to the Platonist mythology, such that objectivity of rules are constituted by standards which ‘transcend the reactions and responses of participants in our practices’, but which, via some occult process are able to guide those practices (ibid: 208, 207).

This, again, conforms closely to (updated) Wright’s recoil from the Platonistic mythology into the sheer contingency of shared agreement in our rule-following practices. Communal agreement cannot be explained, as it could be if we were prepared to entertain Platonism, through appeal to the ‘rules-as-rails mythology’; rather, ‘the truth is the other way round: it is the basic agreement which sustains all rules and rule-governed institutions’ (Wright 2001c: 72). It is, Wright asserts, ‘a *basic* fact about us that our ordinary forms of explanation and training do succeed in perpetuating practices of various kinds—that there is a shared uptake...’ (Wright 2001f: 167). And he talks of ‘the contingency that we are, each of us, ceaselessly but...subcognitively moved’ in similar fashion (Wright 2001i: 313). We will recall the passage that leads Wright to charge Wittgenstein with an unsupported quietism:

do we understand the nature of this “must”. We think it must be in reality; for we think we already see it there.’

⁴⁹ The passage referred to is from Cavell, *Must We Mean What We Say?*, p.52.

⁵⁰ Compare *PI* §108: ‘But what becomes of logic now? It’s rigour seems to be giving way here...’

The great difficulty is to stabilise [this aspect of Wittgenstein's thought] against a drift into a fatal simplification: the idea that the requirement of a rule, in any particular case, are simply *whatever we take them to be*. For, if the requirements of the rule are not constituted, as the Platonist thought, independently of our reactions to the case, what is there to constitute them *but* our reaction? (Wright 2001c: 72; emphases in original).

Again, McDowell's *third way* can be shown to be insufficiently distanced from this aspect of (updated) Wright's *tertium quid*, as emerges in the following remark:

...the fact is that it is only because of our own involvement in our "whirl of organism" that we can understand a form of words as conferring, on the judgement that some move is the correct one at a given point, the special compellingness possessed by the conclusion of a proof...We should accept that sometimes there may be nothing better to do than to appeal explicitly to a hoped-for community of human response' (McDowell 1998e: 209, 211).

'This is not an easy recipe', McDowell admits (ibid: 211, fn. 16). But however uncomfortable we find this vertiginous picture of our "whirl of organism", we must not allow it, McDowell warns us, to prompt a recoil into Platonist mythology ("But there only appear to be these two options..."):

At any rate, it is a bad move to allow oneself to conceive some area of thought from the extraneous perspective at which vertigo threatens, but then suppose one can make oneself safe from vertigo with the idea that rules mark out rails discernible from that external point of view...seeking to escape the threat by finding a solid, externally recognizable foundation (ibid: 211).

We must, therefore, ensure that we resist the recoil to superlative mythology. But how, then, are we supposed to live with the sense of vertigo? McDowell recommends:

Below “bedrock” there is nothing but contingency, so at any time in the future my interlocutor’s use of the expression in question may simply stop conforming to the pattern that I expect...[Now] what we make of the feeling of precariousness will be as follows. When I understand another person, I know the rules he is going by. My right to claim to understand him is precarious, in that nothing but a tissue of contingencies stands in the way of my losing it. But to envisage its loss is not necessarily to envisage its turning out that I never had the right at all (McDowell 1998a: 251-252).

And what secures this (admittedly precarious) shared agreement is, according to McDowell, the fact that one ‘can know another’s meaning without interpretation’ (ibid: 254). But how is this proposed *third way*—the gesture towards institutional practices, or customs—meant to secure this shared agreement? According to McDowell: this ‘response to Wittgenstein’s problem works because a linguistic community is conceived as bound together, not by a match in mere externals (facts accessible to just anyone), but by a capacity for a meeting of minds’ (ibid: 253). But how is pointing to the shared agreement (“a capacity for a meeting of minds”) supposed to *explain* the security of shared linguistic convention in face of the vertiginous sense that such agreement is precariously balanced on a knife edge, but without reaching for the rule-as-rails mythology and its cognates when we feel the ground beneath or feet giving way?

One can be forgiven for sympathising with Boghossian when he professes not to understand the import of these remarks (Boghossian 1989: 544, fn. 66). All McDowell has to say on the matter is to offer a promissory note for further elaboration of this proposed line of explanation (“I’ll get back to you...”), and to rest on Wittgenstein’s authority to ensure us that it promises to be a fruitful direction: ‘But even if we were at a loss as to how [Wittgenstein] might have thought the programme could be executed...this would be

no ground for ignoring the clear textual evidence that the programme is Wittgenstein's own' (McDowell 1998a: 242).

Yet, it is difficult to see how McDowell's proposal for an explanatory programme based on the idea of institutionalised practical activities, or culture can fly on the coat-tails of the philosopher who declared that: 'We must do away with all *explanation*, and description alone must take its place' (*PI* §109). And McDowell's gesture towards further explanatory exploration of the putative "tissue of contingencies" that lie below "bedrock"—no less that Wright's gesture towards supposed "subrational propensities"—sits very uneasily alongside Wittgenstein's insistence that philosophical practice:

...takes its rise, not from an interest in the facts of nature, nor from a need to grasp causal connexions... [I]t is, rather, of the essence of our investigation that we do not seek to learn anything *new* by it. We want to *understand* something that is already in plain view. For *this* is what we seem in some sense not to understand... (*PI* §89).

And, along similar lines:

Philosophy simply put everything before us, and neither explains nor deduces anything.—Since everything lies open to view there is nothing to explain. For what is hidden, for example, is of no interest to us... (*PI* §126).

McDowell's early response to Wittgenstein's dilemma has attracted, not undeservedly, some criticism. Martin Kusch complains that 'there is reason to doubt McDowell's claim according to which Kripke fails to recognize Wittgenstein's crucial third option: primitivism about meaning and rules that is centred around the ideas of training, acting blindly, agreement, custom, practice and institution' and goes on to provide a list of

passages in Kripke's text where such notions are discussed (Kusch 2006: 226). More perceptively, Anandi Hattiangadi objects that McDowell's third way does not enable us to escape the dilemma:

McDowell argues that Wittgenstein should not be seen as a quietist. Rather, he attempts to give a constructive answer to the question "how is meaning possible?" with the use of the notion of a custom or practice. By appealing, on the one hand to dispositions and on the other to customs, Wittgenstein hopes to steer a path between a "bald" naturalism, in which there are no norms, no correctness or incorrectness, and the regress of interpretations... Unfortunately, the appeal to custom here only works to provide a constructive response to the question "how is meaning possible?" if it is already assumed that the regress argument can be avoided (Hattiangadi 2007: 171).

Given that, as I have just pointed out, McDowell's appeal to customs already appears to presuppose shared agreement (which, in any case, itself appears to be haunted by the terror of vertiginous contingency), this judgement of Hattiangadi's appears to me to be sound (other than to point out that the dilemma McDowell wishes to steer a path between is a picture from which norms have been leached out, and which leads soon on to the regress of interpretations and its attendant conceptual nihilism, on the one hand, and ultra-objective, superlative rules, on the other). Moreover, once this criticism is granted, Hattiangadi also seems entirely fair to point out that McDowell's vague gestures towards a "meeting of minds" 'can only be a viable, constructive response to the sceptic if individuals have minds to meet' (ibid: 172). Furthermore, and with such considerations in mind, Boghossian's objections to McDowell's proposed third way also squarely meet their mark:

But if we are simply to be allowed to take the correctness for granted, unreduced and without any prospect of reconstruction in terms of, say,

actual and counterfactual truths about communal use, how is the necessity of an “orderly communal” practice to be defended? From what does the demand for orderliness flow? And from what the demand for community? McDowell’s paper contains no helpful answers (Boghossian 1989: 544).

Yet, although all these criticisms are right in their rejection of McDowell’s version of the “community appeal”, what appears to have been missed within the literature is that McDowell’s proposed third way between Scylla and Charybdis has been squarely rejected by one other prominent commentator on Wittgenstein—and that is John McDowell himself. Subsequent to the reading put forth in “Wittgenstein on Following a Rule”, published in 1984 (and which we have so far been considering in this section), McDowell published some further papers dealing with central Wittgensteinian issues: “One Strand in the Private Language Argument”, published in 1989, and “Intentionality and Interiority in Wittgenstein”, published in 1991 (aspects of which were discussed in Chapter 2). Then, McDowell returns to further sustained reflection on the specific topic of rule-following, and Wittgenstein’s regress of interpretations argument, published in 1993 as “Meaning and Intentionality in Wittgenstein’s Later Philosophy”. Now, admittedly, whilst he doesn’t exactly shout it from the rooftops, in this paper McDowell explicitly repudiates his earlier reading of Wittgenstein’s response to the dilemma. McDowell writes:

Commentators often suggest that the concept of *custom* and its cognates figure in Wittgenstein as elements in a constructive philosophical response to questions like “How is meaning possible?” According to some versions of the reading, Wittgenstein actually gives the response, according to others, he points towards it but does not give it, out of a quietism that must stand exposed as inappropriate by the sheer fact that the questions are (supposedly) good ones. *I am committed to regarding this as a misreading* (McDowell 1998b: 275; emphasis added).

And in a footnote to this passage of text, McDowell states that he considers his earlier paper, “Wittgenstein on Following a Rule” to be ‘too hospitable to this kind of reading’ (ibid: 275, fn. 6). Now, given that, as we witnessed during the previous section, McDowell’s earlier reading of Wittgenstein falls squarely within the first variety outlined here—a reading that explicitly attributes to Wittgenstein a constructive enterprise based on the concept of *custom*, and the like (Wright’s reading, of course, is an instance of the other variety: the reading which makes the accusation of quietism)—then one would perhaps be justified in considering “too hospitable” to be understating matters a little.⁵¹ Regardless, the crucial point is that McDowell retracts his earlier reading. Now, although Boghossian’s objections to McDowell’s earlier treatment of Wittgenstein on rules precede McDowell’s revised reading, the other commentators I referenced above—Kusch and Hattiangadi—include reference to the later paper during their critical reflections upon McDowell’s position. Yet, neither commentator—understandably by all means, given that, as I say, McDowell doesn’t exactly shout from the rooftops—take sufficient account of the fact that McDowell’s later paper embodies a totally distinct way of reading Wittgenstein’s response to the problems. Hattiangadi, for instance, reads McDowell’s early and later papers as weaving complementary strands in an overall interpretation of Wittgenstein which has a destructive aspect (contained in the later paper, “Meaning and Intentionality in Wittgenstein’s Later Philosophy”), and a constructive aspect (the custom-based explanatory proposal in the earlier paper, “Wittgenstein on Following a Rule”). But, as I hope to make clear, the insights contained in McDowell’s later paper cannot be understood without also understanding that these insights entail rejection of the whole explanatory (constructive) approach to the question, “How is meaning possible?” or “How is intentionality possible?” —an instance of which approach is embodied in McDowell’s earlier paper. There, he makes promissory gestures towards *custom* and its cognates as the third way that will somehow (“I’ll get back to you...”) enable us to steer safely between Scylla and Charybdis. But the earlier and the later reading cannot be coherently combined. As we shall see, the later reading supersedes, and makes vital critical objections to, that earlier reading.

⁵¹ Recall, for instance: ‘But even if we were at a loss as to how [Wittgenstein] might have thought the programme could be executed...this would be no ground for ignoring the clear textual evidence that the programme is Wittgenstein’s own’ (McDowell 1998a: 242).

§2 McDowell's Later Reading (or: A Radical New Departure...)

The best way to understand the reorientation of McDowell's reading of Wittgenstein is, I suggest, as a revision of the insight contained in Wittgenstein's assertion, at *PI* §201, that the misunderstanding upon which the seemingly unavoidable dilemma is based is embodied in 'the inclination to say: any action according to the rule is an interpretation'. As we witnessed in the previous section, McDowell earlier took this to imply that the obverse of this misunderstanding—"a way of grasping a rule which is *not* an *interpretation*"—should be identified with *customs*: institutional practices of linguistic rule-governed activities. *Somehow*, it is supposed, this will provide the key that will enable us to explain norm-governed behaviour, but in such a way that we avoid under-mentalizing such behaviour by picturing mere brute reactive propensities to react (avoiding Scylla), on the one hand, and over-mentalising when we readily recoil into superlative, ultra-objective mythology on the other (avoiding Charybdis). In his later approach, however, before he moves on to consider what may be meant by the obverse of the misunderstanding ("grasp of a rule which is *not* an *interpretation*"), McDowell stops to consider, first, what might be meant by Wittgenstein's diagnosis of the misunderstanding itself: "the inclination to say: any action according to the rule is an interpretation".

A good way to understand McDowell's new interpretation of Wittgenstein's diagnosis of the misunderstanding, and how it inevitably leads on to regress, is to see it as based on three broad steps that easily capture our thought when reflecting on meaning, understanding, intentional mental states, and like concepts. The misunderstanding—"the inclination to say: any action according to the rule is an interpretation"—is described in step 2. Step 3 describes how the second step, once made, leads inevitably, and quite generally, to a hopeless regress. Step 1 describes the source of the temptation to take step 2 (that is, to swallow whole Wittgenstein's misunderstanding).

Beginning, naturally, with the first step, McDowell reflects upon the bald statement with which Wittgenstein opens §85 of the *Investigations*: 'A rule stands there like a sign-

post'.⁵² McDowell now bases his whole new reading on a contemplation of this remark that reveals for him that which I am calling the three-step insight. The problem, according to McDowell's (updated) Wittgenstein, is that we begin our investigations with a conception of a rule which simply "stands there like a sign-post". Explaining Step 1, McDowell considers:

'For "a rule" here, we might substitute "an expression of a rule"... For example, we might say "A sign-post—an expression of a rule for following a trail—stands there like a sign-post". How could it be wrong to say that a sign-post stands there like a sign-post? Well, the formulation is dangerous, because "stands there" suggests a conception according to which the rule, or its expression, considered in itself, is normatively inert. It stands aloof from those who encounter it: a mere arrangement of matter, not something that as it were speaks to people, telling them which way to go' (McDowell 2009b: 100).⁵³

That is to say, the sign-post is *not* being conceived as, in itself, as it were, *already constitutively connected with behaviour which would be in accord with its instruction*. Thus the object is not, then, in actual fact being considered as a *sign-post* at all—but, rather, perhaps, as simply a wooden board with some shapes marked on it (compare Wittgenstein's reference, in the *Blue Book*, to mere 'dead signs' (Wittgenstein 1969: 4)).

⁵² In order to set out what I am calling McDowell's "three-step insight" into Wittgenstein's diagnosis of the misunderstanding which leads to regress and on to conceptual nihilism, I will in actual fact pay close attention to the three-step insight as it is described in a paper published by McDowell in 2002, "How Not to Read *Philosophical Investigations*: Brandom's Wittgenstein" (McDowell: 2009b). This is because, although this later paper uncovers the same train of thought as "Meaning and Intentionality..." (1993), it does so with the greater clarity afforded it by the passage of time and further reflection, and so it will better serve the contrast I am trying to make between McDowell's early and later readings of Wittgenstein.

⁵³ Reflecting on this imagery, McDowell observes: 'there is an extra layer of metaphor here [sign-posts telling one what to do], over and above what would be involved in speaking of a board inscribed "Go to the right" as telling one what to do (there would be no metaphor in speaking of someone who says "Go to the right" as telling one what to do)' (McDowell 2009b: 103, fn. 4).

Now describing Step 2—the step into Wittgenstein’s misunderstanding proper (“the inclination to say: any action according to the rule is an interpretation”)—made tempting by the move exemplified in the first step, McDowell writes:

‘[Step 1] threatens the idea that a sign-post points the way to go. One might hope to preserve the idea of a right way to go—even while conceiving a sign-post to be something that stands there, mute and aloof—by supposing that what tells people which way to go is not a sign-post considered in itself, but a sign-post *under an interpretation*. A sign-post under an interpretation sorts responses into those that are correct in the light of it and those that are not’ (McDowell 2009b: 100).⁵⁴

If all we are seemingly faced with is a mere arrangement of “dead signs”—not *a sign-post*, telling those versed in the specific practice of following particular instructions on sign-posts which way to go—then, in order to try to explain the constitutive connection between a sign-post, on the one hand, and behaviour which is in accord with what the sign-post tells one to do, on the other, we seemingly need to view the mere dead lump of wood as telling one which way to go only *under an interpretation*. Only then, it seems, can the required constitutive connection be brought back into the picture. Now, once we have fallen into Wittgenstein’s misunderstanding, described here, then the misunderstanding very quickly infects absolutely everything we may propose as a possible rule-preserving, mediating interpretation (which interpretation is seemingly required by the first step: conceiving the object of our investigations as “mute and aloof”, as normatively inert), and which very quickly shuffles us on to a vicious regress of interpretations, and then on to conceptual nihilism. Ordinary norm-governed practices, such as the practice ‘of erecting and following sign-posts, or perhaps more specifically the custom [or practice] of erecting and following sign-posts of just this style and configuration’ then show up as irredeemably spooky or queer (McDowell 1998b: 276).

⁵⁴ McDowell refers to what I am calling Step 2 as “the master thesis” in 1998b (“Meaning and Intentionality...”).

Of the third step, McDowell writes:

But now whatever made it seem right to say a sign-post stands there, mute and aloof, will equally make it seem right to say, of anything one might want to conceive as an expression of the interpretation that for a moment [PI 201] seems to get normativity back into the picture, that *it* stands there like a sign-post. (Consider for instance a pointing gesture, or an utterance of the words “Go to the right”.) The thing that was supposed to be an expression of an interpretation lapses into normative inertness just as the thing that was originally supposed to tell us the way to go did. And if we suppose that we can recover a normative sorting by considering the thing that was supposed to be an expression of an interpretation under an interpretation in its turn, we are obviously starting on an infinite regress’ (McDowell 2009b: 100).

The first thing to note about this three-step insight into Wittgenstein’s misunderstanding is that, unlike the earlier reading, which ran immediately to consider the possible substance of the obverse of the misunderstanding—“grasp of a rule which is *not* an *interpretation*”—McDowell reflects more deeply on the misunderstanding itself, and how we are led into it, and how it so easily generalises and spreads over anything we may call upon in an attempt to fund our explanatory purposes (that is to say, any possible candidate for the putative something that sorts behaviour into that which is in accord with the expression of a rule and behaviour that is not). No consideration has so far been given to the substance of the obverse of the misunderstanding. In order to accentuate this reflection on the misunderstanding, its source, and its contagious nature, I will likewise leave discussion of the substance of “a way of grasping a rule which is *not* an *interpretation*” until we have reflected first upon Wittgenstein’s misunderstanding, its source, and the virility of its conceptual nihilism.

McDowell's earlier reading began in its explication of the misunderstanding, not with reflection on the misunderstanding itself, but with an immediate turn to offer a substantive account of the obverse of the misunderstanding: 'When I follow a sign-post, the connection between it and my action is not mediated by an interpretation of sign-posts that I acquired when I was trained in their use. I simply act as I have been trained to' (McDowell 1998a: 239). But, at this stage, it is still presumed that the twin horns of the dilemma loom large over this substantial account: it is easy to view this action in accordance with the rule as if it were merely a brute disposition to react (and the relevant training into such dispositions conceived as without distinction from the sense in which dogs can be trained into certain routines). Therefore, it was earlier recommended, we must view the training as training into a *custom, practice, or institution*. But such concepts (pointing as they do for the early McDowell towards the communal *third way* which, he hopes, will enable us to avoid the dilemma) still retain an air of mystery. Wittgenstein's appeal to custom, practice, institution and the like are merely gestures in the direction of a satisfactory explanation of the constitution of the rule by virtue of which we follow a sign-post. Early McDowell agrees with Kripke and Wright that the Platonistic mythology is a non-starter, and so we have, seemingly, to live with the dizzy vertigo of "bedrock" in the face of the radical contingency that we are not reduced to "a babble of disagreement", as Kripke says, but that our communal practices do somehow manage to remain in step (and which communal practices also—*somehow*—point to the source of the normativity of the rule).

The fundamental difference between McDowell's early and late readings of Wittgenstein, then, is that the early reading *begins* by picturing the phenomenon under investigation (let's say, following a sign-post of a particular style and configuration) as essentially unconnected to the normative surroundings, or the circumstances in which it makes sense to conceive the inscribed board *as a sign-post, pointing the way to go*. Whereas, on the later reading, McDowell is careful not to make this move at all. Once this move has been made, we need some way of reinvesting the picture with the normative surroundings—which surroundings are required in order for us to be able to make sense of the practices *as practices of following a sign-post*. But, if we have already abstracted the inscribed board from the (normative) behaviour which surrounds it, *and also, at one*

and the same time then try and reinvest the picture with the normative surroundings—or the constitutive connections between the board and appropriate behaviour made in response to the board (going to the right if the sign points to the right)—then those normative surroundings are going to look warped and deformed as we try simultaneously to hold both pictures together (the normatively inert picture, together with a picture where norms have been squeezed back in). The situation is bound to remain hopeless for we have made our first approach to the (normative) phenomenon under investigation by taking out what it is we need to find! As Wittgenstein says, at *PI* §103, the normatively inert picture of the (normative) phenomenon under investigation ‘is like the pair of glasses on our nose through which we see whatever we look at. It never occurs to us to take them off’.⁵⁵

And so then we inevitably face a dilemma (“there only seem to be these two options!”): *either* we are returned to the fact that we have leached norms from our picture (and so behaviour made in response to a sign seems like a mere brute disposition to react; the connection between this behaviour and the meaning of the sign is then open to indefinite interpretation, and the threat of regress soon appears). Hence *PI* §430:

—It is as if we had imagined that the essential thing about a living man was the outward form. Then we made a lump of wood in that form, and were abashed to see the stupid block, which hadn’t even any similarity to a living being.

Or, on the other hand, the norms we try to squeeze back into the normatively inert picture of the scene inevitably look “spooky” or “queer”, as though such norms are, by some occult, magical process, able to reinvigorate the picture (that we are still

⁵⁵ And the remark following: ‘We predicate of the thing what lies in the method of representing it...’ (*PI* §104).

concurrently holding on to) of mere dead signs and merely brute behaviour. Hence *PI* §431:

“There is a gulf between an order and its execution. It has to be filled by the act of understanding.”

“Only the act of understanding can mean we are to do THIS. The *order*——why, that is nothing but sounds, ink-marks.—”

That (early) McDowell is still holding on to a version of the normatively inert picture is suggested by the following remark, where he states that unless we can point to customs, or communal practices, etc. as the hoped-for “third-way” that will reanimate our picture, then we would inevitably be left with the picture of rule-expressions as merely “dead signs” and the behaviour which surrounds these dead signs as *merely brute*. Training must be seen as ‘initiation into a custom’, (early) McDowell insists:

If it were not that, then the account of the connection between a sign-post and action would indeed look like an account of nothing more than brute movements and its causal explanation; our picture would not contain the materials to entitle us to speak of following (going by) a sign-post (McDowell 1998a: 239).

But *how*, exactly, is the notion of our being trained into a custom, or into institutions, or into communal practices supposed to reanimate the required normativity (or constitutive connections: going to the right when a sign points to the right, for instance) into the picture of following the sign-post, so that this seemingly inevitable slide to dead signs and brute behaviour (all “mute and aloof”) is halted? The answer to that question, for (early) McDowell is still a mystery. That is why further constructive examination of the proposed *third way* is required (“I’ll get back to you...”). All that is known at this stage is that we cannot remain squarely within the dead signs/brute behaviour picture because that

eventually leads to vicious regress (and soon on to conceptual nihilism). Neither can we appeal to the Platonist fantasy of ultra-objective, superlative rules as *somehow* standing judgement on our (merely brute) behaviour. Unless future examination of the *third way* (initiation into communal practice, customs, institutions) can *somehow* serve to reanimate the picture, then we seem destined to reverberate, without satisfaction, between these two problematic poles (“a gangway with rigid walls”). So, the early constructive promise (the gesture towards the third-way) is borne of desperation, rather than any particularly appropriate reason to think that the appeal to customs will be able to achieve this feat. Hence *PI* §203:

Language is a labyrinth of paths. You approach from *one* side and know your way about; you approach the same place from another side and no longer know your way about.

In McDowell’s later reading, however, the decisive step is *precisely not to begin* with the picture from which norms have been leached out (and then be required to reanimate this picture, *somehow*...)—but, rather, to *decisively reject taking this fateful step at all*. To return to Wittgenstein’s metaphor at *PI* §103, we must take off the glasses that it hadn’t previously occurred to us to take off. Once this step *has* been made, however, then our being faced by the Scylla-Charybdis dilemma is inevitable. Hence *PI* §308: ‘(The decisive movement in the conjuring trick has been made, and it was the very one that we thought quite innocent.)’ The right response to this situation is to reject the “dead signs”/merely brute behaviour picture altogether. McDowell thus points to two principal lessons of Wittgenstein’s regress of interpretations argument. This first lesson is that:

...it is disastrous to suppose that there is always a conceptual gap between an expression of a rule and performances that are up for assessment according to whether or not they conform to the rule, a gap that is made vivid by saying the expression of a rule stands there...[on the normatively inert, “mute and aloof” construal]. We must not acquiesce in the idea that

an expression of a rule, considered in itself, does not sort out behaviour into performances that follow the rule and performances that do not. ...If we let the gap open at all, it will be unbridgeable. This way, we lose our grip on the idea of an expression of a rule, or an expression of an interpretation. In the end we lose our grip on the idea of an expression of anything (McDowell 2009b: 100-101).

If we are in the grip of 'the confused equation of concepts with interpretations,' it is a confusion which, as the much neglected Wittgenstein commentator, the late Dewi Phillips points out, 'invites the question: "Interpretation of *what?*"' (Phillips 2001: 17). And this brings us to the second lesson of Wittgenstein's regress argument—and also finally brings us to the obverse of the misunderstanding: 'a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call "obeying the rule" and "going against it" in actual cases' (*PI* §201)—which is that, as McDowell observes:

...what is fundamental is the ability to act immediately on an understanding—to act in a way that is not mediated by an interpretation of what is understood...Wittgenstein's regress shows that acting on an understanding cannot in general be acting on an interpretation of what is understood (McDowell 2009b: 103).

Of course, McDowell acknowledges:

...not everyone who encounters a sign-post gets told which way to go. Sign-posts do not speak to those who are not party to the relevant conventions...To be capable of being told what to do by a sign-post, one needs to have been initiated into an appropriate practice (ibid: 101).

In order to try to make this point clear, consider, for instance, a wooden board at a crossroads upon which is painted the inscription: “Ar gyfer Farchnad Pontypridd, ewch i’r chwyth”. A non-Welsh speaker who apprehends the inscribed board will not know what it says: is it the expression of a rule? a warning? a profession of undying love? something else entirely? Assuming the person is proficient in broad practices and customs of following signs, however, she might reasonably interpret the inscribed board, placed at a crossroads, *as a sign-post* (although, lacking proficiency in the Welsh language, she nevertheless does not know what the sign-post tells one, specifically). Thus, if she is familiar with practices and customs of following signs; and she knows that, being in Wales, the inscription on the board is likely to be *something* written in the Welsh language; and moreover, if she knows that Pontypridd is a town in Wales: then she might reasonably interpret the board *as a sign-post, in Welsh, telling (Welsh-speakers) how to get to Pontypridd*. (Nevertheless, although a reasonable interpretation to make, *for all she knows*, her interpretation may be mistaken—*for all she knows*, the board might be mischievously daubed with professionally painted Welsh graffiti, perhaps, reading “Good grief! You wouldn’t want to go to Pontypridd...”, or some such).⁵⁶

Regardless, the crucial point is that a non-Welsh-speaker who apprehends the inscribed board—however she interprets it (or not): “it’s surely a sign-post...?”—is not going to be able to grasp (“in a flash”), as those who are proficient Welsh-speakers *will* upon reading the *sign-post* (for that is what it most definitely is, for those who have been inducted into practices of following signs, and who are Welsh speakers) *telling one that for Pontypridd Market they must take a left turn*. And the Welsh-speaker who, desiring to get to Pontypridd Market, immediately takes a left turn at the behest of the sign-post is manifesting *in* (not behind) *her behaviour* that which Wittgenstein calls “‘obeying the rule’ in actual cases’: that is to say, directly acting upon an understanding ‘which is *not* an *interpretation*’ (PI §201). It is, as McDowell says, ‘behaviour [which] is animated by grasp of a rule’ (McDowell 2009b: 105). Her behaviour ‘would reflect the fact that the subject is such that the sign-post itself, not the sign-post under an interpretation, tells her which way to go’ (ibid: 101). And, although the behaviour is immediate, and not mediated by

⁵⁶ In which case, the sign would read: “Jiw Jiw! Byddet ti ddim moen fynd i Bontypridd...”

any interpretation, it is nevertheless *behaviour which manifests understanding* of the normative practice (a performative demonstration of her competency in the practice: a demonstration of her grasp of the norms involved in the practice)—and *precisely not*, as McDowell says, ‘a mere uncomprehending disposition to react [to the rule-expression]’ (ibid: 101, fn. 3). For to conceive her behaviour in these terms would be to remain in the grip of the misunderstanding (it would be, to borrow Wittgenstein’s ocular metaphor once more, to have failed to have taken off the spectacles).

Notice, however, that even for the non-Welsh-speaker, her interpretations presuppose *understanding* (of sign-posts, of graffiti, etc.), and include an *understanding* of the context of her surroundings (she is at a crossroads; in Wales; they have their own language; Pontypridd is a town there; and so on). Summarising these kinds of considerations, Phillips, writes:

Interpretations, like theory-laden perceptions, are parasitic on concepts which are not interpretations, and on perceptions which are not theory-laden. For example, there are situations where we need to interpret whether someone is angry or sorrowful. But were there not situations in which what we mean by “anger” and “sorrow” does not call for interpretation, the call for interpretation, in other contexts, would be unintelligible. If “anger” and “sorrow” were not appreciated by someone in the situations where no interpretation is called for, we would conclude that they did not understand what is meant by “anger” and “sorrow” (Phillips 2001: 9).

Of course, somebody may misunderstand something, or fail to understand. But mis- or non-understanding is clearly parasitic on the concept of *understanding*. Thus, a pupil might well write, as imagined at *PI* §185, “...1000, 1004, 1008, 1012...” upon being asked to extend the series “Add 2” beyond 1000. But such a deviant rule is exactly that—*deviant*. It is not, as Kripke’s sceptic would have us believe, somehow the same

rule, but also at the same time a different rule (“the particular rule *under a different interpretation*”).⁵⁷ The “quus” function is not “the plus function under a different interpretation”. It is a different function entirely. And someone who answered “5” upon being presented with the sum “68 + 57” would be thereby demonstrating—*manifesting in her behaviour* (not suggesting anything lying “behind” it)—her lack of understanding of the plus-function. Thus, immediately after Wittgenstein has introduced the hypothetical student’s hypothetical deviant response, hypothetically: ‘We say to him: “Look what you’ve done!”—He doesn’t understand’ (*PI* §185). (*Not*: he understands “the same rule under a different interpretation,” or some such. But: *he doesn’t understand the rule*).⁵⁸ To quote Phillips once more:

The logical point to which we return is that reflective interpretations are dependent on concepts which are not further interpretations, and which are invoked in the course of the interpretations we are offered. To reject concepts in these non-interpretative contexts would not be to reject an interpretation, but to show that one did not understand the concepts in question (Phillips 2001: 11).

Notice that Phillips describes the crucial point as a *logical*, or *conceptual* one. For someone might object that, in directly citing the notion of understanding which is “*not an interpretation*”, we are, like the earlier commentators we discussed (including early McDowell), problematically *presupposing* what is supposed to be being explained. But this would betray a misunderstanding. For explanation was only needed when we had already fallen into the misunderstanding. In order to *explain* the linguistic or mental phenomenon under investigation, we must not, we feel, presuppose the meaning or the understanding (for that is what we hope to explain). So, we take out the meaning and the understanding and are left with behaviour considered as merely brute, or the expression

⁵⁷ I think that the “transtemporal” form with which Kripke often discusses the constitutive problem (and which I attempted to clear away in Chapter 1) serves to hide this obvious incoherence.

⁵⁸ Compare *PI* §120: ‘You say: the point isn’t the word, but it’s meaning, and you think of the word as a thing of the same kind as the word, though also different from the word. Here the word, there the meaning. The money, and the cow that you can buy with it.’

of the rule as merely “dead signs” (“mute and aloof”). And then, as a consequence, we really do seemingly need to *explain the meaning and the understanding back into the picture* (by virtue of something else—a “third way”). But, as we have already seen, such explanatory ambitions are destined to remain forever thwarted, for we have already removed precisely that which we seek in our picture (*PI* §308: ‘The decisive movement in the conjuring trick has been made...’): behaviour manifesting understanding of the meaning of a word, or of the significance of a rule. Wittgenstein observes:

The fundamental fact here is that we lay down rules, a technique, for a game, and that then when we follow the rules, things do not turn out as we had assumed. That we are therefore as it were entangled in our own rules (*PI* §125).

The fundamental significance of pointing to instances of manifest and unmediated understanding of the meaning of a word, or of the significance of a rule, is decidedly not, in Wittgenstein’s method, being brought forth as any kind of explanation. Desiring explanation is precisely what gets us into the difficulties in the first place. *PI* §125 continues: ‘This entanglement in our rules is what we want to understand (i.e., get a clear view of).’ The reason for which direct understanding (unmediated by any interpretation) is being called forth is *not* in the service of any kind of explanation of the concepts which puzzle us. And so, it is not problematic that the meaning or the understanding is being presupposed. Rather, the reason for which direct understanding (unmediated by any interpretation) is being called forth *is to try to stop us from making the problematic move—i.e., by leaching norms out of the picture* in the first instance (*PI* §308: ‘The first step is the one that altogether escapes notice’). The function of the appeal to unmediated understanding (‘which is exhibited in what we call “obeying the rule” and “going against it” in actual cases’: *PI* §201) is not in the service of any kind of explanation (so that it would then be, problematically, being presupposed); rather, its function is for ‘clearing misunderstandings away’ (*PI* §90).

As I mentioned in Chapter 1, Kripke detects a “Kantian flavor” in Wittgenstein’s investigations: ‘our investigation, however, is directed...towards the “possibilities” of phenomena’ (*PI* §90). That is to say, the issues discussed are *constitutive*: they concern the possibility of using a word according to a determinate meaning, or of forming a thought or an intention with a specific content. But we need not ‘penetrate phenomena’ or dig out ‘something that lies *beneath* the surface’ (*PI* §§90, 92)—speculating about what lies beneath “bedrock”, for instance. The apparent need to do this only comes from being enmired within the misunderstanding (when we still have on the spectacles). Rather: ‘We remind ourselves...of the *kind of statement* that we make about phenomena’ (*PI* §90). *PI* §127: ‘The work of the philosopher consists in assembling reminders for a particular purpose’. And the purpose for which we call forth such reminders (the reminder that people—those who are competent practitioners, at least—can act immediately on the understanding of a sign-post, for instance) is to clear away misunderstanding: ‘The problems are solved, not by reporting new experience, but by arranging what we have always known’: ‘something that already lies open to view and that becomes surveyable by a rearrangement’ (*PI* §§109, 92). To repeat: instances of unmediated understanding are called forth *not* as vague programmatic gestures toward future constructive philosophical explanation—but in the service of dispelling misunderstanding. Hence at *PI* §§654-655, Wittgenstein remarks:

Our mistake is to look for an explanation where we ought to look at what happens as a “proto-phenomenon.” That is, where we ought to have said:
this language-game is played.

The question is not one of explaining a language-game by means of our experiences, but of noting a language-game.

As McDowell points out, *PI* §654: ‘*this language-game is played*’ echoes *PI* §180: ‘*This is how these words are used*’ (2009a: 88). At *PI* §180, Wittgenstein makes the point that it would betray misunderstanding to speak of those avowals of instantaneous understanding (“Now I can go on!” “Now I understand!”, and so forth) as ‘a “description of a mental state”’, as if such exclamations were a mere *symptom* of the real *meaning* or

understanding which lie hidden beneath such fulgurations. Rather: ‘—One might call them a “signal”’: that is to say, a signal that the person has understood (or at least thinks she has understood); ‘and we judge whether it was rightly employed by what she goes on to do’: that is to say, they would be incorrectly used if she doesn’t then go on to manifest the appropriate understanding in her behaviour (*PI* §180).⁵⁹ ⁶⁰

Removing the spectacles, as it were, involves *not* trying to explain the language-game in terms of something else, but by tending to the ‘*particular circumstances*’ in our lives where certain concepts are used (*PI* §§154-155). Wittgenstein states: ‘We are talking about the spatial and temporal phenomenon of language, not about some non-spatial, non-temporal chimera’ (*PI* §108). The pictures of non-spatial, non-temporal chimeras (Platonistic fantasies of rules-as-rails, for instance) are a consequence of our thought orbiting within the gravitational space of the misunderstanding. Once we have made the step into viewing the mental and linguistic phenomena of our investigation in a way that is normatively inert (“mute and aloof”) then the norms bound up with such mental and linguistic phenomena inevitably become warped and deformed.

To try and make this point clearer, let us consider the train of reasoning which is set on course at *PI* §138, where Wittgenstein raises a distinction between an occurrent phenomenon of consciousness (understanding—“in a flash”—the meaning of the word “cube” on an occasion of its utterance, for instance), on the one hand, and appropriate application of the word, extended in time, on the other. At *PI* §139, Wittgenstein raises a question as to the nature of the connection between an occurrent phenomenon of consciousness, on the one hand, and subsequent linguistic performance, on the other: ‘And how can what is present to us in an instant, what comes before our mind in an

⁵⁹ Of course, we can imagine cases where, say, the person has a heart attack at the moment sudden understanding dawns, so then is prohibited from performing the correct application—even though she could have done so before the advent of the imagined medical emergency, and so the exclamation (“Now I understand!”) would still have been correctly used here, although the person is unable to demonstrate that understanding (but now for contingent, medical reasons—not because she didn’t actually understand): cf. *PI* §§181, 183.

⁶⁰ Compare, for instance: *On Certainty* §12: ‘—For “I know” seems to describe a state of affairs which guarantees what is known, guarantees it as a fact. One always forgets the expression “I thought I knew”.’

instant, fit a *use*?' Immediately, Wittgenstein begins canvassing plausible candidates for what may be said to enter into consciousness at the instantaneous dawning of one's understanding of the meaning of a word, and which may be able to mediate, for us, we suppose, the connection between such instantaneous understanding, on the one hand, and the propriety of applications of that word, on the other: 'What really comes before my mind when we *understand* a word?—Isn't it something like a picture? Can't it *be* a picture?' But a picture of a cube, say, for it to be a picture *of a cube*, must itself be constitutively connected with appropriate applications of the word "cube" (that is to say, if it really is a picture of *a cube*, then the word "cube" can be correctly applied to the picture—but if not, then the word cannot, in this instance, be correctly so applied). And so the same question about constitutive connectivity can be asked of it: 'Well, suppose that the picture does come before your mind when you hear the word "cube", say, the drawing of a cube. In what sense can this fit or fail to fit a use of the word "cube"?'

Over the next couple of remarks, Wittgenstein argues that the appearance of any such intentional mental occurrence within introspectable awareness is neither necessary nor sufficient for the manifestation of understanding (that is to say, for appropriate subsequent performance, made in accordance with the word's meaning). The occurrence of an image before one's conscious mind isn't necessary for appropriate performance, for we can perfectly conceive of instances where no mental picture appears before a person's mind, or even different mental pictures on different occasions, and yet he still goes on to apply the word correctly. Wittgenstein makes this point vivid by noting that 'it is absolutely inessential for the picture to exist in his imagination rather than as a drawing or model in front of him', and we surely do not wish to insist that we are required to draw a physical diagram of a cube each time we apply the concept (*PI* §141).

The mental image isn't sufficient to dictate a determinate use because the same picture may flash before the mind, and yet subsequent applications of the word be different (*PI* §139). For instance, imagine that a teacher orders two pupils to each fetch a cubic object. And let us further imagine that an identical schematic diagram of a cube flashes before the mind of each pupil. It is still possible to suppose that one pupil returns with a cube,

whereas the other pupil who has, let us suppose, understood the very same picture as ‘a plane figure consisting of a square and two rhombi’ (*PI* §74), rather than as a three-dimensional rendering, accordingly returns with a planar object of just that shape. Thus, any such conscious imagery appearing before one’s mind can be of no explanatory use in our search to make sense of the source of normative authority. There still needs to be a *constitutive connection* between a picture of a *cube* that comes before my mind, and appropriate performance of the word “cube”. ‘What is essential,’ Wittgenstein urges, ‘is to see that the same thing may come before our minds when we hear that word and the application still be different. Has it the *same* meaning both times? I think we shall say not’ (*PI* §140). And we cannot secure a connection between such an occurrent mental image and appropriate subsequent performance through claiming that ‘not merely the picture of the cube, but also the method of projection comes before my mind: ...say a picture of two cubes connected by lines of projection’ (*PI* §141). The same questions could now be asked of this further interpretation: what serves to secure it as the *correct* interpretation of the first interpretation? And now we are clearly sliding towards a familiar regress.

At the end of this train of thought, at *PI* §141, Wittgenstein asserts: a criterion of understanding is ‘the application which—in the course of time—[the subject] makes of what he imagines’; then he further asserts that there is appropriate performance and there is inappropriate performance, according to the standards laid down by the rule: ‘I want to say: we have here a *normal* case, and abnormal cases’. This clearly links with *PI* §201: ‘...there is a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call “obeying the rule” and “going against it” in actual cases’. That is to say, if the person has been successfully trained into competence in the practice, or language-game, of making avowals of instantaneous understanding (“Now I understand”), and also *has* understood the meaning of the word (which she will then be able to demonstrate in her performance), then all that has happened is that the person has manifested her linguistic competence. If such conditions are met, then what has happened during the instantiation of the occurrent phenomenon of consciousness is, as McDowell says, ‘the onset of understanding itself’: ‘It is no more than an exploitation of the language-game to say that what makes its appearance in the moment, on these

occasions, is the understanding itself. What happens is that the understanding sets in' (McDowell 2009a: 90). Or, to borrow Aaron Ridley's metaphor: 'Meaning has been earthed' (Ridley 2004: 30).

But Wittgenstein soon imagines a possible rejection of this simple appeal to the meaning having been earthed, the understanding having set in—the understanding which is embodied (or not) in appropriate performance ('exhibited in what we call "obeying the rule" and "going against it" in actual cases'), and is made possible by prior training into such competence ('"Mastery" of a technique': *PI* §150). In the context of a discussion of understanding the principle of an algebraic series, at *PI* §146, Wittgenstein remarks:

—Perhaps you will say here: to have got the system (or, again, to understand it) can't consist in continuing the series up to *this* or *that* number: *that* is only applying one's understanding. The understanding itself is a state which is the *source* of the correct use.

What is one really thinking of here? Isn't one thinking of the derivation of a series from its algebraic formula? Or at least something analogous?—But this is where we were before. The point is, we may think of more than *one* application of an algebraic formula, and any type of application may in turn be formulated algebraically; but naturally this does not get us any further.—The application is still a criterion of understanding.

Actual behavioural application of the series is here being conceived as indirect evidence, a mere *manifestation* of the *real* meaning, or the *real* principle of the series. What we need, it seems, is to examine the *real source* of these correct behavioural applications, and which will, we suppose, serve to *explain* what makes those applications correct (as distinguished from performance which would not be in accord with the principle of the series). But, as was the case with the possibility of the picture of the word "cube" coming to mind during sudden understanding of the word, Wittgenstein, for the duration of a couple of subsequent remarks (*PI* §§151-152), patiently argues that the appearance of an

algebraic formula coming to mind, say, is neither necessary nor sufficient for appropriate application of the series during subsequent linguistic performance. Again, as was the case with the introspectable picture of a cube, there is a *constitutive connection* between an algebraic formula for the series, which may (or may not) come to mind, and correct application. Any algebraic formula for the series which may (or may not) enter introspectable awareness cannot serve as an explanatory bridge between a sudden onset of understanding, say, and subsequent performance: in other words, it cannot serve to explain the required constitutive connection. For, any algebraic formula for the series which may (or may not) enter introspectable awareness is—if it is indeed a *formula for the series*—itself constitutively connected to appropriate application of the series. And, once again, it is of no use for our explanatory purposes to appeal to a further algebraic formula in order to fix the constitutive connectivity of the first formula, for now we are on the high road to regress.

So, Wittgenstein affirms yet again: ‘—*The application is still a criterion of understanding*’ (PI §146; emphasis added). Once more: there is appropriate performance and there is inappropriate performance, according to the standards laid down by the rule: ‘I want to say: we have here a *normal* case, and abnormal cases’ (PI §141). That is to say: ‘...there is a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call “obeying the rule” and “going against it” in actual cases’ (PI §201). Meaning has been earthed, understanding is manifest (or not).

And yet, once more, Wittgenstein’s alter-ego cannot rest satisfied with such assertions. The interlocutor cannot but see such behavioural applications as mere, finite *manifestations* of the *real meaning*, or the *real principle of the series*. Such manifestations are then being viewed as behavioural data, and which supposedly serve as empirical evidence (an “external”, or contingent connection) of instances where appropriate application of the series has been performed. The actual series is indefinite, so how can one’s mere finite, behavioural applications constitute the meaning, or the principle?

“But how can it be? When I say I understand the rule of a series, I am surely not saying so because I have *found out* that up to now I have applied the algebraic formula in such-and-such a way! In my own case at all events I surely know that I mean such-and-such a series; it doesn’t matter how far I have actually developed it.” —

Your idea, then, is that you know the application of the rule of the series quite apart from remembering actual applications to particular numbers. And you will perhaps say: “Of course! For the series is infinite and the bit of it that I have developed finite.” (*PI* §147).

So, the infinite expanse of the rule—its meaning—is one thing, it is supposed, and the actual, finite behavioural applications of the rule are another. Thus, *PI* §152: “‘He understands’ *must have more in it* than: the formula occurs to him. And equally, more than any of those more or less characteristic *accompaniments* or manifestations of understanding’ (first emphasis added). What, then, provides the *must have more in it*? What is the *source* of the mere, finite behavioural manifestations of the understanding? If the state or process that can serve to explain the mere manifestations, or accompaniments of the understanding cannot be found within the domain of introspectable, conscious awareness then, surely, we must look elsewhere for such a state:

If one says knowing the ABC is a state of the mind, one is thinking of a state of a mental apparatus (perhaps of the brain) by means of which we explain the *manifestations* of that knowledge. Such a state is called a disposition (*PI* §149).

Application of the ABC is mere behaviour, we suppose, and so we need to look for the real source of the understanding which can explain this behaviour as being in accord (or not) with the principle of the extension of the series of letters of the alphabet. But nothing introspectable could be found which could serve to pin down behaviour which is

in accord with the series (and behaviour which is not). In the case of an algebraic formula coming to mind, for instance, we can, in turn, ‘think of more than one application of an algebraic formula; and any type of application may in turn be formulated algebraically’ (*PI* §146). And so we turn to look for the *meaning* of the formula beyond introspectable awareness: a hidden mental state (something like “subrational propensities”, perhaps?) which is most likely to be found located somewhere in the brain, or in the central nervous system: ‘And if there is no difference in what [we] happen to be conscious of there must be one in the unconscious workings of [our] minds, or, again, in the brain’ (*PI* §156).

But, as we have continually seen throughout the discussion in earlier chapters, if the state or process of the brain is going to be able to secure the condition of correctness in relation to appropriate behaviour, then it cannot be viewed as *brutely* dispositional, or *merely* causal—just purely physical bits of brain matter, say. For as we have already seen, such brain states could, at best, be only “externally”, or contingently connected with appropriate performance, whereas what we require is to uphold a *constitutive* connection. Wittgenstein makes this point plain when he points out that we could ‘have knowledge of the construction of the apparatus, quite apart from what it does’ (*PI* §149). And so, again, we appear to need a further interpretation in order to secure the right kind of connection—the *constitutive* connection—between the states of the brain or nervous system, or what have you, and appropriate application of the series, or of the formula. Such purely physical bits of matter would stand no less “mute and aloof”, as McDowell says, as the formula in the mind—considered apart from the application of the series for which it is putatively a formula—was conceived to be (and which could, then, be given a seeming indefinite variety of interpretations, linking it to some pattern of behaviour or another). Summing up this line of thinking, Wittgenstein remarks:

We are trying to get hold of the mental process of understanding which seems to be hidden behind those coarser and therefore more readily visible accompaniments. But we do not succeed; or, rather, it does not get as far as a real attempt. For even supposing I had found something that had happened in all those cases of understanding,—why should *it* be the

understanding? And how can the process of understanding have been hidden, when I said “Now I understand” *because* I understood?! And if I say it is hidden—then how do I know what I have to look for? I am in a muddle (*PI* §153).

Returning to the stage at which Wittgenstein has argued that appeal to a mental image of a cube is neither necessary nor sufficient to guarantee appropriate performance of the word “cube”, Wittgenstein asks: ‘Is there such a thing as a picture, or something like a picture, that forces a particular application on us...?’ (*PI* §140). As we saw, appealing to an intentional mental occurrence (the image of a *cube*, say) cannot serve as the explanatory bridge between instantaneous understanding of the word “cube,” on the one hand, and appropriate performance, on the other. This is because such a picture could be, on some interpretation, brought into relation with all and any pattern of subsequent behaviour. And it is of no use to appeal to a further interpretation as fixing only one interpretation of the picture as correct, for any such further interpretation would itself be open to further interpretation, and so on to an intolerable regress... But what if there were such a thing as a picture which *forces a particular application on us*? This would be like an ordinary interpretation—in that it serves to connect momentary grasp of the word, on the one hand, with appropriate performance, on the other—‘except that it is’, as McDowell says, ‘somehow not itself susceptible to interpretation’ (McDowell 2009a: 83). Recall Wittgenstein’s reference to such a special, regress-stopping interpretation, or self-interpreting rule in the *Blue Book*: ‘What one wants to say is: “Every sign is capable of interpretation, but the *meaning* musn’t be capable of interpretation. It is the last interpretation”’ (Wittgenstein 1969: 34).

This is, again, the superlative mythology, but now applied to an intentional occurrence (an image in the mind). Recall *PI* §388: “‘...But it is essential to the image that it is the image of *this* and of nothing else.” Thus one might come to regard the image as a superlikeness’. And a similar temptation might assail us in the context of posited mental mechanisms. The mental states or process can’t, of course, be viewed in *merely* brute, or *purely* physical terms, for such phenomena, so conceived, cannot serve to ground the

required *constitutive* connection with appropriate performance. So, as we saw in Chapter 2: §2b (ii), we are led to view such mental mechanisms as mysterious kinds of *quasi-physical fact*: something which broadly resembles an ordinary physical fact except, unlike an ordinary physical fact, is a special *content-bearing* or *constitutive connection-maintaining* physical fact. But now to extend that earlier discussion further in light of recent insights, what we are in fact imagining when we imagine special regress-stopping interpretational states, are not so much quasi-physical facts; they are, rather, as McDowell points out, *quasi-magical*:

So we picture a criterion that still relates to the [physical] construction of the [mental] apparatus, but mysteriously contrives to incorporate standards for correctness and incorrectness in performances. This is a way into picturing what one acquires, when one grasps a sense, as something that “determines the future use” not causally, at any rate not by way of any ordinary causation, but in a quasi-magical way (McDowell 2009a: 84-85).

Imagining the source of normativity as located in the mental apparatus (the brain, or the nervous system, or whatever) is thus a counterpart to the special regress-stopping picture of the “rule-as-rails”, and, despite its pseudo-scientific imagery, is no less mythical. This temptation towards a mythical regress-stopping, superlative interpretation (once we have started on the conception that leads all understanding to appear as interpretation) thus reveals the source of the “queerness” of which Wittgenstein speaks, or the “eeriness” described by Kripke: ‘Can we conceive of a finite state which *could* not be interpreted in a quus-like way? How could that be? ...[T]he nature of the proposed state is left mysterious’ (Kripke 1982: 52-53). Or Wright: ‘How is the *subject* to know of the application of such a concept...if nothing of which he is conscious intimates the fact to him?’ (Wright 2001e: 136). Recall *PI* §§195: “But I don’t mean that what I do now (in grasping a sense) determines the future use *causally* and as a matter of experience, but that in a *queer* way, the use itself is in some sense present” ...’; and *PI* §196: ‘In our failure to understand the use of a word we take it as the expression of a queer *process*. (As we think of time as a queer medium, of the mind as a queer kind of being.)’ In this

connection, McDowell raises also Wittgenstein's reference at *PI* §454 to: '...a hocus-pocus which can be performed by the soul'.

We will recall also that, at *PI* §158, Wittgenstein considers a possible objection here, along the following lines: But isn't the mysteriousness of such states and processes 'only because of our too slight acquaintance with what goes on in the brain and nervous system? If we had more accurate knowledge of these things we should see what connections were established by [our] training [into rule-governed practices, etc.]'. In Chapter 2: §2b (ii), I argued that such a question would be point-missing as it would neglect to see that in positing quasi-physical facts we are: *either* not providing a scientific explanation at all; *or*, we are implicitly recommending a broadening of what we are willing to class as a scientific fact. But, in light of the foregoing discussion, this point can now be made even more forcefully, as McDowell does in the following remark:

Once it is clear that nothing but a configuration in an occult medium could seem to meet the requirements we are tempted to place on a state of the understanding, the idea that the brain might serve as the locus of the required configurations must surely emerge as unsatisfactory. The brain is no doubt very remarkable, but not in a way that amounts to its being capable of quasi-magical feats ("hocus-pocus")... [And] once it is evident that only something occult could even seem to meet the supposed need, it is clear that it would be merely point-missing to suppose neuroscience, say, might help with the difficulty (McDowell 2009a: 85-86).^{61 62}

⁶¹ For an analogous discussion along the lines presented here, albeit briefer, see Genia Schönbaumsfeld's discussion of 'the model of "brute fact plus magic ingredient"' (Schönbaumsfeld 2016: 75). Of significance also for the reading being offered here is that, during that discussion, Schönbaumsfeld refers to 'the view that signs by themselves are "dead" and that what, as it were, breathes life into them is some sort of mental act (be it the "mental act" of "meaning", "interpretation", or "intention")' as 'the thought that, arguably, the whole of Wittgenstein's later philosophy is meant to be an antidote to' (ibid: 74).

⁶² Although these objections are, of course (and not unlike the interlocutor who resists Wittgenstein's repeated refrains), not going to be decisive for the proponents of theories of mental mechanisms—highly prevalent under various guises (broadly functionalist theories, the "Language of Thought" hypothesis, and the like)—it is worth noting the relevance of the preceding remarks in regard to such programmes. Although functional theorists, broadly construed, claim that internal mental states can be individuated by

At *PI* §154, the temptation to say that expressions of a fulguration of understanding (“Now I can go on!”) are employed as ‘a description of a process occurring behind or side by side with that of saying the formula’, Wittgenstein, famously, retorts:

If there has to be anything “behind the utterance of the formula” it is *particular circumstances*, which justify me in saying I can go on—when the formula occurs to me.

Try not to think of understanding as a “mental process” at all.—For *that* is the expression which confuses you. But ask yourself: in what sort of case, in what kind of circumstances, do we say, “Now I know how to go on,” when, that is, the formula *has* occurred to me?—

In the sense in which there are processes (including mental processes) which are characteristic of understanding, understanding is not a mental process...

describing their causal roles in the production of types of meaningful or intentional behaviour, they nevertheless affirm that such behaviour is, as Kim Sterelny asserts, ‘largely the effect of organic processes...i.e., that they are physiological processes located, presumably, in the nervous systems of organisms’, so that, ultimately, ‘mental events have true descriptions in the vocabulary of an ideally completed physiology’ (Sterelny 1990: 2, 9). When the functionalist talks of mental functions, then, ‘the notion of function is biological; it is the same sense in which the function of the kidney is excretion and water regulation’ (ibid: 11). Yet, as we have seen, such purely physical matter cannot uphold semantic and mental content, and so, alongside such a physical description, the functionalist needs to provide a ‘complete functional description [which] specifies the causal roles of the full range of human psychological states’ (ibid: 3). Nevertheless, our mental lives, it is claimed, must ultimately be fully describable in purely physical terms. Functionalists do not want to claim to be, as Sterelny puts it, ‘in the miracle business’ (ibid: 22):

[E]ach functional role must have a physical occupier. Functions must be performed; the entities carrying through the performance are physical entities with physical properties. So a theory of the functional role of intentional states must avoid attributing magical powers to mental states. The functions allegedly essential to mental states must be functions actually performable by physical stuff (ibid).

Yet, as I have suggested, such theorists are destined to remain forever stuck betwixt and between these two poles: purely physical, or merely causal states, from which all meaning and mind goes dark, on the one hand, and ‘spooky causal processes, or devices with impossible causal powers’, on the other (Sterelny 1990: 81). Consider, Wittgenstein’s remark in the *Blue Book*: ‘It was in fact just the occult character of the mental process which you needed for your purposes’ (Wittgenstein 1969: 5).

And, once more, at *PI* §155 Wittgenstein, in response to the temptation to suppose that, during those fulgurations of understanding, ‘possibly he had a special experience’, again asserts: ‘it is *the circumstances* under which he had such an experience that justify him in saying in such a case that he understands, that he knows how to go on.’ When the discussion returns after a digression on “reading”, at *PI* §179 Wittgenstein fills out the kind of circumstances he has in mind, which surround, for instance, such exclamations:

The words “Now I know how to go on” were correctly used when he thought of the formula: that is, given such circumstances that he had learnt algebra, had used such formulae before.—But that does not mean that his statement is only short for a description of all the circumstances which are a stage setting for the language-game.—Think of how we learn to use the expressions “Now I know how to go on”, “Now I can go on” and others; in what family of language-games we learn their use.

We can also imagine the case where nothing at all occurred in [the subject’s] mind except that he suddenly said “Now I can go on”—perhaps [or perhaps not] with a feeling of relief; and that he did in fact go on working out the series without using the formula. And in this case too we should say—in certain circumstances—that he did know how to go on.

Wittgenstein repeatedly advises us to consider the *circumstances* surrounding instances of linguistic performance. Yet there is a constant temptation to avert our gaze from such circumstances and to look elsewhere for something that will serve to *explain* the applications of, say, an arithmetic series. Rather than simply consider the familiar contexts in which, for instance, arithmetical competence is displayed, we keep turning away to look for some introspectable inner event, or to search for some hidden mental mechanism, or whatever. But, then, whatever phenomenon we turn up for inspection cannot be the understanding, because we have already averted our gaze from the circumstances in which the actual understanding is displayed *in the behavioural applications* of the series, the word, or the rule (and the contextual circumstances which surround such irreducibly norm-governed behaviour). Whatever it is we reach for as an

explanation of the understanding—of, say, the principle of a series—it is inevitably going to be *something other than the understanding itself* (or, a “third way”): that is to say, something other than the behavioural (but irreducibly normative) applications of the steps of the series. If we have such explanatory ambitions, then we cannot simply cite the onset of understanding itself. As McDowell says: ‘Invoking such things would just be making moves in the language-game—not stepping outside it, as one would need to do in order to explain it’ (McDowell 2009a: 89).

If it is explanation we are after, then invoking the steps of the language-game (the appropriate steps in application of the series) would be to pre-suppose the meaning and the understanding we are hoping to explain. And so our explanatory ambitions are thwarted. But if it is the *understanding of the meaning of the series* that we are trying to explain, then our explanatory ambitions are anyway going to remain forever thwarted for so long as we restrict ourselves to not being able to invoke the appropriate steps in application of the series—for making the *appropriate* steps at any particular stage of the development of the series *is precisely what constitutes* the understanding of the series! A criterion of saying of someone that they understand the series is that they are able to apply the appropriate steps at the appropriate stage. If they cannot do this, then that is a criterion for saying that they do *not* understand the series. Competence in the behavioural applications of the relevant language-game constitutes understanding the language-game: ‘—*The application is still a criterion of understanding*’ (PI §146); ‘...*there is a way of grasping a rule...which is exhibited in what we call “obeying the rule” and “going against it” in actual cases*’ (PI §201).

As McDowell remarks: ‘Learning a language-game that includes “Now I understand!”...gives us a new possibility of saying, straight off...how it [is] with us...on suitable present occasions’ (ibid: 90): that is to say, on occasions where we have understood—or, at least, *think* we have understood. Hence PI §180 again:

This is how these words are used. It would be quite misleading, in this last case, for instance, to call the words a “description of a mental state”.—One might rather call them a “signal”; and we judge whether it was rightly employed by what he goes on to do.

“Now I know how to go on” is not a description of a feeling; nor some other kind of introspectable occurrence; nor of a hidden mental state—but an expression we have learned how to use in contexts such as, for instance, extending an algebraic series, and is correctly used, in most cases, when I am then able to apply the series correctly.

Once we have abstracted the linguistic or mental phenomenon we are investigating from its normative surroundings —the familiar circumstances which surround employment of the relevant concepts—then the only way we seem able to squeeze all the normative surroundings back into the picture is by imagining that *all the steps* of the extension of an algebraic series, say, or *all the rules* of the game of chess must—*somehow*—be *literally* present in the “act of understanding.” Of course, no such “infinite” state is available to introspectable awareness, and so we are led to posit some hidden, quasi-magical process, which is conceived as the *source* of all the mere behavioural, finite manifestations of the relevant concept applied over time. *PI* §93: ‘a misunderstanding of the logic of our language seduces us into thinking that something extraordinary, something unique, must be achieved by propositions.—A *misunderstanding* makes it look to us as if a proposition *did* something queer’. ‘The forms of our expression’, Wittgenstein asserts, ‘prevent us in all sorts of ways from seeing that nothing out of the ordinary is involved, by sending us in pursuit of chimeras’ (*PI* §94).

Wright, for instance, sees that an intention to play a game of chess, say, cannot *literally* contain all the rules of the game. But, given that Wright’s thought operates within the orbit of the misunderstanding, he cannot but identify the idea that one’s intention to play chess is constitutively connected with all the rules of chess (for chess is the game one intends to play) with such mythology. So when Wright justifiably rejects such superlative

fantasy, he thereby rejects the idea that one can be consciously aware that one intends to play a game of chess, or that one can grasp the whole use of a word in a flash. But, if we take care not to enter into the misunderstanding in the first place, then there is no need to deny such common sense:

“It’s as if we could grasp the whole use of a word in a flash.”

—And that is just what we say we do. That is to say: we sometimes describe what we do in these words. But there is nothing astonishing, nothing queer about what happens. It becomes queer when we are led to think that the future development must in some way be present in the act of grasping the sense and yet isn’t present.—For we say that there isn’t any doubt that we understand the word, and on the other hand its meaning lies in its use (*PI* §197).

Linguistic meaning and the content of intentional mental states only seem queer or extraordinary when our thought is moving within the orbit of the misunderstanding. The meaning of a word, or the principle of a series only seem queer when we try to step outside the language-game, as it were, to try to explain it from “outside”.⁶³ In other words, to attempt to explain the meaning of a word, say, without presupposing the determinate content of that meaning. But once the misunderstanding has been cleared away, there should be no problem speaking of the “ratification-independence”, to borrow Wright’s expression, of linguistic meaning and of intentional mental phenomena. Thus, as McDowell says: ‘Putting the idea picturesquely, we can say that the meaning reaches forward in the series ahead of anyone who actually works the series out, and is so to speak already there waiting for such a person, ready to stand judgement over her performance, at any point she reaches in the series’ (McDowell 1998b: 274). Or there should be no problem with the idea that a sign-post is, in itself, so to speak, constitutively

⁶³ Thus the relationship between language and the world come to seem as if it were ‘a *super-order* between—so to speak—*super* concepts. Whereas, of course, if the words “language”, “experience”, “world”, have a use it must be as humble a one as that of the words “table”, “lamp”, “door”’ (*PI* §97)

connected to behaviour which is in accord with its instructions (that the sign-post can tell competent practitioners the way to go). Similarly: 'When we say, and *mean*, that such-and-such is the case, we—and our meaning—do not stop anywhere short of the fact; but we mean: *this—is—so*' (*PI* §95). And, if such ordinary linguistic feats seem queer, or spooky—that, for instance, there seems to be a 'super-strong connexion...between the act of intending and the thing intended' (*PI* §197) —then we need only remind ourselves of the very ordinary circumstances of our employment of such concepts:

—Where is the connexion effected between the sense of the expression
"Let's play a game of chess" and all the rules of the game?

—Well, in the list of rules of the game, in the teaching of it, in the day-to-day practice of playing (*PI* §197).

Similarly:

It may now be said: "The way the formula is meant determines which steps are to be taken". What is the criterion for the way the formula is meant? It is, for example, the kind of way we always use it, the way we are taught to use it (*PI* §190).

Thus, the purpose of the *reminders* of "the spatial and temporal phenomenon of language" (the circumstances which surround the institution of chess, for instance, or the practice of going to the right when one encounters a sign saying "Go to the right!") are there to stop us from falling into the misunderstanding in the first place: 'But we talk about it as we do about the pieces in chess when we are stating the rules of the game, not describing their physical properties' (*PI* §108). We do not need some kind of explanatory description (or, more accurately: a proposed quasi-description) of "subrational propensities", "classificatory dispositions", "communal practices", "customs", "institutions", and so forth (the 'tendency to assume a pure intermediary', as

Wittgenstein says at *PI* §94, between, say, a word and its applications). The need for a *third way* between Scylla and Charybdis only seems to be required because our thought keeps reverberating between these two hopelessly problematic poles (“a gangway with rigid walls”). Hence, (early) McDowell states: ‘So there are three positions in play...’ (1998a: 243). But we do not need a third way. What we need to recognise is that there is but *one way*: behaviour which is directly animated by an understanding of a rule (or not). If we can remind ourselves of that, then we do not take the step into the misunderstanding. And if we do not step into the misunderstanding, then we are not faced with the seemingly unavoidable choice between Scylla and Charybdis. And, hence, we do not need any third way. What we need to do is to renounce our explanatory ambitions, which end up deforming the (irreducibly normative) phenomena we wish to understand, and simply allow the phenomena to be what they are: *This language-game is played. This is how these words are used...*

Conclusion

As I hope to have shown in this chapter—although not, to my knowledge, explicitly discussed within the published commentary on Wittgenstein on rules and of rule-following—McDowell's thought on this topic undergoes a radical reorientation, and this reorientation reflects the very change in our thinking which Wittgenstein aims to persuade us to undertake (on pain of our reflections on meaning and mind remaining forever imprisoned within the "gangway with rigid walls"). To borrow Wittgenstein's metaphor, McDowell takes off the spectacles that it hadn't earlier occurred to him to take off. And we are invited to take them off too.

Although McDowell does not express himself in these terms, the fundamental difference between his earlier and later readings of Wittgenstein, I have been urging, is that in the earlier reflections McDowell, not unlike Wright, moves immediately to consider how we might *explain* the positive substance of the obverse of Wittgenstein's misunderstanding: 'there is a way of grasping a rule which is *not* an *interpretation*' (PI §201). Unfortunately, however, McDowell's reflections on the obverse of the misunderstanding thus take place with the spectacles very much still on. As a result, his thought is still haunted by the threat from the regress of interpretations, and the obliteration of norms and correctness conditions (Scylla), on the one hand, and ultra-objective superlative fantasy (Charybdis), on the other. Thus, the substance which McDowell imports into the obverse of the misunderstanding is characterised by vague gestures towards a putative "third way" between Scylla and Charybdis: an appeal to communal practices, customs, or social institutions as a way of explaining phenomena involving meaning and understanding. Yet these gestures are borne of desperation in face of the intolerable dilemma, rather than based on any sound reasons to think this "communal third way" will enable us to answer questions of the form, "How is meaning possible?" With the spectacles of misunderstanding still on his nose, and with Scylla in one lens and Charybdis in the other, McDowell's putative third way remains enigmatic and mysterious. All McDowell can say about it at this early stage is: "I'll get back to you..."

At this early stage, then, McDowell's picture of practices of following according to a rule are insufficiently distanced from that of Kripke and Wright, and his proposed "third way" beyond Scylla and Charybdis has justly attracted criticism. Yet, in his own understated manner, McDowell eventually repudiates his earlier reading as based on the very misunderstanding Wittgenstein is trying to warn us against. Hattiangadi reads McDowell's early and late readings of Wittgenstein as complementary aspects of a two-pronged strategy: a destructive element based on a rejection of the "master thesis" (that which I have been calling Wittgenstein's misunderstanding, or Step 2 of the three-step insight) and a constructive element (the "third way" appeal to customs, communal practices, institutions, and so forth). But the destructive element (proffered in the later reading) is destructive *precisely of* the earlier constructive programme. McDowell's later reading expresses an explicit repudiation of the earlier reading, and so they cannot intelligibly be combined.

The fruits of McDowell's greater reflection on Wittgenstein's diagnosis of the source of misunderstanding—"the inclination to say: any action according to a rule is an interpretation"—are contained in what I have been calling the three-step insight into the source of the constitutive paradox. McDowell later comes to realise that, in an attempt to satiate the craving for explanation in face of the seemingly mysterious properties of meaning and mind, we are inclined to picture an expression of a rule—a sign-post, for instance—as "mute and aloof" and behaviour made in response to the rule-expression as a merely brute propensive reaction. Once such a picture is accepted, however, we are very swiftly led, in our attempts to make sense of following according to a rule, to irresolvable paradox. If we allow any such 'conceptual gap' to open up between the expression of a rule, on the one hand, and behaviour which is open to assessment according to the normative standards of the rule, on the other, then the gap, as McDowell warns, 'will be unbridgeable. This way, we lose our grip on the idea of an expression of a rule, or an expression of an interpretation. In the end we lose our grip on the idea of an expression of anything' (McDowell 2009a: 100-101). Any attempt to reinvest normativity into such a picture, without retracting the fateful first step, leads inevitably to norms, or constitutive connections showing up in extraordinary, queer, or superlative manner (on pain of not showing up at all).

This is the principal lesson of Wittgenstein's rule-following considerations, according to McDowell. Yet it is a lesson not learned in Wright's reading of Wittgenstein, as it is precisely such a conceptual gap which constitutes Wright's putative mental paradigms: "pure" or "genuine" occurrences in consciousness, constitutively autonomous and entirely unconnected with the world "outside", on the one hand, and merely brute ("dispositional-like") behaviour, on the other—and, of necessity, never the twain shall meet. As we saw in Chapter 2, the more closely we examine each face of the two putative paradigms of the mental, in both instances what we find staring back at us are the two horns of the irresolvable dilemma.⁶⁴

The antidote to falling into this hopeless conceptual bind, is not to take the fateful step into the picture where an unbridgeable conceptual gap opens up between rule-expressions standing "mute and aloof" from merely brute propulsive reactions (mysteriously made in accordance with the demands of the "rule"). And the way to do this, McDowell recommends, is to—eventually—pay heed to the obverse of Wittgenstein's misunderstanding: 'there is a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call "obeying the rule" and "going against it" in actual cases' (*PI* §201). For those who are inductees into a relevant practice, in response to the expression of a rule they can act immediately on their understanding of that rule. Whilst not mediated by interpretation, such action is *decidedly not*, however, 'a mere uncomprehending disposition to react', to borrow McDowell's phrase (McDowell 2009b: 101, fn. 3). Such action is, rather, and to borrow another expression of McDowell's, 'behaviour which is animated by grasp of a rule' (*ibid*: 105). On pain of

⁶⁴ The following passage from Kripke's text reveals that, in addition to the possibility of meaning-constituting facts ("Jones means addition by 'plus'", and so forth)—rejected as "desperately mysterious" and queer—the only other options he can consider conform to the two paradigms of the mental delineated by Wright—"pure" occurrences, on the one hand, and dispositional behaviour, on the other:

Perhaps we may try to recoup, by arguing that meaning addition by "plus" is a state even more *sui generis* than new have argued before. Perhaps it is simply a primitive state, *not to be assimilated to sensations or headaches or any "qualitative" states, nor to be assimilated to dispositions*, but a state of a unique kind of its own (Kripke 1982: 51; final emphasis added).

inevitable misunderstanding, then, we must view linguistic behaviour as already *constitutively connected* with the meaning of a word (which is embodied in the appropriate linguistic behaviour); or view behaviour made in response to a rule as already *constitutively connected* to the significance of the expression of a rule (going to the right at the behest of a sign-post pointing to the right, for instance).

In summary, then, McDowell's later reading involves the decisive step of rejecting falling into a picture where norms have been leached out of irreducibly norm-governed behaviour ("taking off the glasses"): 'What is fundamental,' McDowell urges, 'is the ability to act immediately on an understanding—to act in a way that is not mediated by an interpretation of what is understood' (McDowell 2009b: 103)⁶⁵. This insight is the fruit, as

⁶⁵ In stating that linguistic behaviour is irreducibly norm-governed, I do not mean to be defending a position which in any way imputes to Wittgenstein a thesis about language as rule-governed (that all instances of language-use, say, are grounded, or underpinned by rules), or that rule-governedness is a necessary condition of linguistic meaning. This would raise serious suspicions that the spectacles of misunderstanding are firmly in place, and that any such proposed theory is merely another guise of the hoped-for "third way", or "something else" (in this instance, "a rule" or, perhaps, "a norm") which putatively underpins, say, the connection between the meaning of a word and linguistic behaviour which is made in accord with that meaning. All that I wish to convey by saying that linguistic behaviour, say, or action made in accordance with a rule is irreducibly norm-governed is to say no more, but no less, than that there are correct or incorrect applications (no matter how precise or vague) of a word; or that there is behaviour which is correct or incorrect in light of the significance of a rule (as I endeavoured to make clear in the first chapter). Thus, for instance, the word "red"—in its "primary sense" (I will return to this in a moment)—is correctly used only of red things; the intention to go on holiday is satisfied only in the event that one goes on holiday; the order to "stand roughly there" is fulfilled only in cases where the person does remain in the general vicinity being pointed to; "125" is the appropriate response on being asked to calculate the sum of "68+57"; just as going to the right is the appropriate response to a sign which points to the right. I do not wish in any way to propose that, in stating such facts, we are committed to any claim about what all such cases must be. Clearly all these cases are very different, and there is no general standard of correctness which underpins each case. Nevertheless, each instance involves *some kind of correctness-condition*. (If this were denied, there would be no problem, or paradox—for trying to make sense of such behaviour made in accordance with a correctness-condition was our very starting position: if we denied such "norm-governed" behaviour from the very get go, there would be nothing to be puzzled about!) At *On Certainty* §28, Wittgenstein remarks: 'What is "learning a rule"—*This*. What is "making a mistake in applying it"?—*This*.' Yet whatever "*This*" behaviour that is in accord with the particular rule might be, or whatever "*This*" behaviour which does not accord with the rule may be, being able to discern which behaviour does, or does not, accord with the relevant rule presupposes an understanding of that rule. *PI* §146: '—*The application is still a criterion of understanding...*' Failure to discern that *This* behaviour accords with the significance of the rule, or that *This* behaviour does not, would be a criterion of saying of someone that they do not understand, or have misunderstood the relevant rule.

Furthermore, given that I only wish to point out this basic, non-theory-laden level of behaviour manifesting understanding of some kind of correctness-condition, or constitutive connection (whatever it may be, from instance to instance), it makes no difference to me (or my Wittgenstein) that words can be used according to a secondary sense—saying of someone that they "saw red" to mean that they were enraged, for

I have argued, of McDowell's greater reflection upon the implications of Wittgenstein's diagnosis of the misunderstanding (rather than immediately searching for explanatory substance in the obverse of the misunderstanding). Grasping this insight is indeed fundamental, as McDowell rightly states, for the reorientation in our thinking which Wittgenstein is trying to instil in us. Yet this insight does not, I contend against McDowell, reach as far as the full import of the reorientation in our thinking that Wittgenstein invites us to undertake. More needs to be said here for us to plumb the depths of Wittgenstein's insight. In order to understand the full import of the Austrian philosopher's methodology we must become self-consciously aware that what is called for—what is being recommended—is *precisely a reorientation in our thinking*.

In order to begin to spell out what I mean by this, I will refer to Wittgenstein himself, who talks, in *PI* §125 (quoted on p.189), of: '[t]he fundamental fact here.' And the fundamental fact to which he refers is 'that *we lay down rules*, a technique, for a game...' That is to say, through a desire to respond to puzzlement in face of, say, behaviour which manifests understanding of the meaning of a word in the form of giving an *explanation* of such behaviour (from the "outside", as it were), then *we implicitly stipulate* that we cannot cite the behaviour which manifests understanding ("...*we lay down rules*, a technique, for a game..."). Yet, at the very same time, the behaviour which manifests understanding must remain in the picture—for that is precisely what we are hoping to explain. But, on the other hand, it cannot remain in the picture—for then we would not have provided an *explanation* (which would require citing something other than the behaviour which manifests understanding...!) So we end up in a paradoxical bind. *PI* §125 thus continues: '...and then when we follow the rules, things do not turn out as we had assumed. That we are therefore as it were entangled in our own rules. This entanglement in our rules is what we want to understand (i.e., to get a clear view of).'

instance. For, even in this case, the expression would be incorrectly, or inappropriately used of someone who had all the while remained in a state of unperturbed and calm tranquillity.

In order to facilitate this understanding of our entanglement in our own rules, then, Wittgenstein calls upon reminders (such as the reminder that, for instance, going to the right is the appropriate response to a sign-post pointing to the right): ‘The work of the philosopher consists in assembling reminders for a particular purpose’ (*PI* §127). And the purpose of such reminders—their whole point—is ‘for clearing misunderstandings away’ (*PI* §90). Such as, for instance, the misunderstanding embodied in the entanglement in our own rules, occasioned by the rules of the game *laid down by ourselves* (in the service of our explanatory ambitions). Thus, trying to *explain* away the puzzlement in face of the understanding of meaning manifest in linguistic behaviour, say, is an entirely futile and hopeless endeavour. For, trying to explain (the rules for the game that *we* lay down) is *the source of puzzlement* in the first instance! This, then, is the fundamental distinction between an attempted explanation of linguistic and mental concepts (which, for the reasons given, inevitably leads to hopeless paradox) and “reminders”: which serve to alert us to the source of this hopeless conceptual bind, *in the very form of our approach* to matters of meaning and mind (propelled by our explanatory ambitions).

Recognising *the fundamental fact* of *PI* §125, then, involves an underlying and requisite shift of focus: a fundamental reorientation in our approach. That is to say, removing the spectacles of misunderstanding essentially and necessarily involves shifting our focus *from* reflection which is primarily trained upon linguistic meaning, or of mental concepts *to reflecting on our reflection* (on linguistic meaning and mental concepts). Thus, at *PI* §108, Wittgenstein states: ‘—The preconceived idea...can only be removed by turning our whole examination round. (One might say: the axis of reference of our examination must be rotated, but about the fixed point of our real need.)’ In other words, our examination must now be focussed *on our reflection* on meaning and mind, and no longer *primarily* directed at matters of meaning and mind. Thus, when Wittgenstein offers his diagnosis of the source of the paradox of meaning and of intention at *PI* §201, he states that the misunderstanding is based ‘*an inclination to say...*’ (that ‘every action according to the rule is an interpretation’). Thus, the primary focus—‘...*the fundamental fact...*’—is now,

once our examination has been turned around, what we are inclined to say (about say, understanding the meaning of a word, or of following according to a rule).⁶⁶

‘What is fundamental,’ to repeat McDowell’s recommendation, ‘is the ability to act immediately on an understanding—to act in a way that is not mediated by an interpretation of what is understood’ (McDowell 2009b: 103). But what is even more fundamental—Wittgenstein’s ‘...*fundamental fact*...’—is that the source of our puzzlement concerning such instances of unmediated understanding lies within ourselves, as it were. The source of puzzlement is contained within, so to speak, the form of our investigation—the way we approach the linguistic and mental phenomena under scrutiny. Thus *PI* §203 (quoted on p. 184; second emphasis in original): ‘*You approach from one side and know your way about; you approach the same place from the other side and no longer know your way about...*’ This fact concerning our form of approach to the problems is more fundamental than the fact of our ability to act immediately on an understanding (going to the right in response to a sign pointing to the right, for instance)—that which McDowell takes to be fundamental. That it is more fundamental is owing to the fact that, if we fail to recognise that the source of our puzzlement lies in ourselves, as it were—in the way in which we approach matters of meaning and mind—then we will be unable to take the point about the ability to act immediately on an understanding (“to act in a way that is not mediated by an interpretation of what is understood”) in the required spirit. For we will still be inclined to say: “Yes—but *how?*! How is that possible...?” But any such response would indicate that the appeal to action which manifests understanding without mediation by interpretation has not worked as a “reminder” at all—it would betray that our thought is still propelled, rather, by

⁶⁶ Thus, in *PI* §430 (quoted on p.182), Wittgenstein states: ‘It is as if *we had imagined...*’ (‘...that the essential thing about a living man was the outward form’). And immediately continues: ‘Then *we made...*’ (‘a lump of wood in that form, and were abashed to see the stupid block, which hadn’t any similarity to a living being’). And, at *PI* §120 (quoted on p. 188, fn. 58), Wittgenstein quips: ‘*You say...*’ (‘...the point isn’t the word, but it’s meaning, and you think of the word as a thing of the same kind as the word, though also different from the word. Here the word, there the meaning. The money, and the cow that you can buy with it.’) Many more examples could be given of instances in the *Investigations* where Wittgenstein turns the focus on to what *we say in our reflections* (on meaning and mind). At *PI* §308, Wittgenstein remarks: ‘The decisive movement in the conjuring trick has been made...’ to which we might add: *by us* (by virtue of what we are inclined to say).

explanatory ambitions (and thus have not grasped that those explanatory ambitions are, ultimately, the source of puzzlement).⁶⁷

Wittgenstein is primarily teasing out the characteristic (but problematic) steps in our practices of reflecting on, say, linguistic phenomena, in order that the nature of those reflective practices are brought to our self-conscious awareness, and their problematic nature revealed. This is the sole reason for which facts about linguistic phenomena, or about rule-following are asserted (after all, is it really news to us that people are able to immediately manifest understanding of a sign-post by, for instance, turning right at the behest of a sign-post pointing to the right?! Cf. *PI* §128). And where Wittgenstein *does* draw attention to familiar features of such practices, the *purpose/function* of such reminders is *not* to tell us something about our linguistic lives (our language-use is, after all, everyday and ordinary), but, rather, to alert us to the deformities in our thinking: to “clear misunderstandings away).

In other words, the principal topic of Wittgenstein’s discussion has been, not so much meaning, mind, and intentionality as we might have thought—but, rather, the movements of thought which are characteristic of our *reflections on* such phenomena. *Our reflections*, then, constitute the principal topic of conversation, contrary to what we may have thought. This is not to say that considerations of meaning, mind, and intentionality are absent from those reflections. The point, though, is that *those reflections* (on meaning, mind, and intentionality) *are often the principal focus of*

⁶⁷ Perhaps we might say that McDowell’s appeal to action which manifests understanding without mediation by interpretation will then be taken as a thesis. There is a sense in which Wittgenstein may accept such an appeal as a thesis. But, if so, it is a thesis which, on its own, is next to vacuous: ‘If one tried to advance *theses* in philosophy, it would never be possible to debate them, because everyone would agree to them’ (*PI* §128). The real danger is, however, that without grasping Wittgenstein’s “fundamental fact”, then such an appeal to unmediated understanding will likely be taken as an attempt to import substance into the obverse of Wittgenstein’s misunderstanding—as an *explanatory* thesis: as a proposed “third way” between Scylla and Charybdis. McDowell invites such a response by not clarifying that understanding the full import of such an appeal to unmediated understanding can only be grasped as requiring a fundamental reorientation in our approach to matters of meaning and mind (that we need to turn our investigation round)—which is another way of saying that Wittgenstein’s “fundamental fact” is indeed more fundamental.

Wittgenstein's reflections. Thus, Wright recommends that we reorient our intuitive conceptions of meaning and mind. Wittgenstein, on the contrary, urges us to see that what is required is a reorientation, rather, of *our reflections on such mental and linguistic phenomena*. Our intuitive conceptions of meaning and mind—that we can understand the determinate meaning of a word, say, or that we can form an intention to do something specific—are perfectly fine. What we took to be queer, eerie, or extraordinary features of our (ordinary and everyday) linguistic and minded lives are, in truth, features of *the approach we take in our reflection* upon such phenomena.

As we have seen, when we are enmired in misunderstanding, ordinary mental and linguistic phenomena can appear queer and mysterious. Thus, for instance, unmediated understanding of a rule—manifest in behaviour (turning right at the behest of a sign-post pointing to the right, for instance)—can come to seem spooky or extraordinary. But an understanding of the rule under discussion is presupposed *in the very same reflection where understanding seems queer and mysterious*. As McDowell says, ‘sign-posts do not speak to those who are not party to the relevant conventions’ (McDowell 2009b: 101). And were we not party to conventions of responding to sign-posts of the relevant configurations, then our questions could get no purchase—consider, for instance, the difference between “But how is it possible for us to follow according to a rule?” or “How is it possible to understand a sign-post?”, over against “But how is it possible to wongdoodle a vomchok?” *We presuppose understanding of the very same understanding which puzzles us, in our reflections on that understanding*. Yet we like to imagine that, at the same time, we can look upon behaviour which manifests understanding (of some rule, say) from a perspective from which understanding is absent, or inoperative (as we appear to require, in order to *explain* the meaning that is understood). Even as we demonstrate—make manifest—the understanding which, we imagine, has been bracketed from our reflections!

As I said earlier (p. 187), failures of understanding are parasitic upon understanding, and that remains the case in this context. The notion of a particular, specific rule which is one at the same time a different rule is obviously incoherent (thus, *quus* and *plus* are different

functions: the *quus* function cannot be, at one and the same time, the plus function too). Yet, the form in which we approach our investigation (propelled by our explanatory ambitions) serves to mask such obvious incoherence. We import mystery into our investigation—the mystery does not inhere in the phenomena under investigation. At *PI* §147 Wittgenstein remarks: ‘*Your idea*, then, is that you know the application of the rule of the series quite apart from remembering actual applications to particular numbers’ (emphasis added). As Wittgenstein states at *PI* §104: ‘*We predicate* of the thing what lies in the method of representing it...’ (emphasis added). Given that we import the sense of mystery through the form of our investigation, unless we take off the spectacles of misunderstanding, we are forever doomed to run around in endless circles, chasing our own tails trying to dispel the mystery which, we erroneously think, inheres in the phenomena under investigation—as if a novel fact (a “third way”, a “something else”) would silence our confusion.

We understand what it is to follow a sign-post—to understand its instructions. What we fail to understand is the source of our puzzlement, during our reflections on such ordinary and everyday feats of understanding. We are inclined to think that the source of our puzzlement lies in some mysterious features of practices of following according to a rule, such that, therefore, we need to investigate such phenomena more closely. It does not occur to us that the real source of our puzzlement is contained in *our approach* to the phenomena under investigation. It is the very *form of our investigation*—not the phenomena under investigation—which is the source of our puzzlement (and which lends an aura of mystery to what is being investigated). *We are entangled in our own rules.*

If we return once more to the misunderstanding of *PI* §201—“*there is an inclination to say: every action according to the rule is an interpretation*”—then in light of the discussion which begins at *PI* §§138 (cf. pp. 191-203), one of the messages imparted by Wittgenstein is that this inclination is relentlessly stubborn and resistant. In other words, taking off the spectacles of misunderstanding and turning our investigation round is no easy matter and requires equally relentless and stubborn counter-resistance. Thus, the inclination of Wittgenstein’s interlocutor/alter-ego to incessantly search for a “third way”

which will finally serve as the bridging interpretation between the meaning of a word, or significance of a rule, on the one hand, and appropriate applications of the word or rule, on the other. Thus, we will recall, in an attempt to silence the temptation to reach for an introspectable mental image as the “something else” which will guarantee the required constitutive connection, Wittgenstein argues that such introspectable mental occurrences are neither necessary nor sufficient to explain the required connection. And so, Wittgenstein submits that a criterion of understanding is ‘the application which—in the course of time— [the subject] makes of what he imagines’ (*PI* §141). That is to say, understanding manifest in behaviour, unmediated by interpretation. Yet, as we saw, Wittgenstein’s alter-ego cannot rest satisfied with this, and immediately moves to search for an alternative “third way”, or “something else” which will, the interlocutor hopes, serve as the putative bridging interpretation—the source of meaning which supposedly lies hidden in the mind-brain. Yet, in amongst a rich battery of argumentation against such a picture, Wittgenstein once more asserts: ‘*The application is still a criterion of understanding*’ (*PI* §146). And yet, once more, the interlocutor cannot rest satisfied with this, and moves to search for yet another candidate “third way, or “something else”—surely, Wittgenstein’s alter-ego is inclined to say, the source of meaning (supposedly hidden in the mind-brain) is indicated by a ‘special experience’ which appears when understanding dawns instantaneously. In face of this inclination, Wittgenstein asserts once more: ‘it is *the circumstances* under which [the subject] had such an experience that justify him in saying in such a case that he understands, that he knows how to go on’ (*PI* §155).

‘We are trying to get hold of the mental process of understanding which seems to be hidden...’ (*PI* §153). This is, first and foremost, a remark about our habits of reflection. It is not principally a remark concerning matters of meaning and mind (for we are already have a hold, in our reflection, on an instance of the relevant phenomena of understanding under investigation—applying the appropriate steps in the development of an arithmetic series, for instance). ‘But,’ Wittgenstein’s remark continues, ‘we do not succeed.’ But our lack of success is *not* because we have yet to find the mysterious, hidden fact (the hoped-for “third way” beyond Scylla and Charybdis, hidden somewhere behind the mere behavioural application of the series). But, rather, we do not succeed

because such an attempt, in our reflection on matters of meaning and mind, *PI* §153 continues, 'does not get as far as a real attempt.' That is to say, in searching for a hidden source of meaning we are chasing after a phantom. An illusion occasioned by our entanglement in our own rules.

Conclusion

Throughout the first two chapters of the present work we were confronted by different instances of the same general form of dilemma. On the one hand, as Wright protests, the ‘price of objective meaning’ seems to be ‘an *absolute conception* of truth: a conception absolved from all practical controls’ (Wright 2001c: 77). But, on the other hand, as Kripke articulates, the vehicles of language ‘are visible (or audible or palpable), concrete phenomena—marks or diagrams on paper’ (Kripke 1982: 106). But conceived *merely* as such, then such concrete phenomena are seemingly open to infinite interpretation. And so, on pain of lapsing into superlative fantasy, we are thrown once more into the threat of a regress of interpretations, and its attendant conceptual nihilism.

At the beginning of Chapter 3, we saw that McDowell’s early reading of Wittgenstein is similarly captured by a form of this dilemma: brute dispositions to react and threat from regress, on the one hand, and superlative mythology, on the other. Not unlike Wright, (early) McDowell fails to reflect with sufficient depth on Wittgenstein’s misunderstanding—‘an inclination to say: any action according to the rule is an interpretation’ (*PI* §201)—and, instead, moves immediately to import substance into the obverse of the misunderstanding—‘there is a way of grasping a rule which is *not* an *interpretation*’ (*PI* §201)—in the hope that it will *somehow* (“I’ll get back to you...”) open up the “third way” between Scylla and Charybdis. Yet this proposed third way (communal practices, institutions, culture) still retains an aura of mystery. Thus McDowell’s early reflections on the issues are insufficiently distanced from Kripke’s and Wright’s respective readings, and, like theirs, his picture of meaning and of rule-following remains haunted by vertiginous terror at the sheer contingency of agreement in individual reactions. Below “bedrock” ‘there is nothing but contingency’ and ‘the ground seems to have been removed from our feet’ (McDowell 1998a: 251; 1998e: 207).

Although this has gone unremarked amongst commentators, McDowell’s reflections on Wittgenstein’s discussion of rules and of rule-following undergoes a radical reorientation, which is based upon a greater reflection upon Wittgenstein’s diagnosis of the source of

the misunderstanding at *PI* §201: ‘an inclination to say: any action according to a rule is an interpretation.’ And the fruits of this greater reflection are that which I have termed McDowell’s “three-step insight” into the source of the constitutive paradox, whereby in an attempt to satiate the craving for explanation in face of the seemingly mysterious properties of meaning and mind, we are inclined to conceive an expression of a rule—a sign-post for instance—as normatively inert (“mute and aloof”) and behaviour made in response to the putative rule-expression as a merely brute propensive reaction. Once such a picture is accepted, however, we are inevitably led—in our reflections upon following according to a rule, or behaviour manifesting understanding of the meaning of a word—to irresolvable paradox.

The way to avoid succumbing to such a hopeless conceptual bind is, according to McDowell, to refuse to take the fateful step (propelled by our explanatory ambitions) into a picture according to which behaviour made in accordance with a rule isn’t already constitutively connected to the expression of a rule (as going to the right, for instance, is constitutively connected to a sign-post pointing to the right). And the way to avoid succumbing to such a picture is, according to McDowell’s later (reoriented) reading of Wittgenstein, to hold fast to the idea that—for those who have been inducted into the relevant practices—they have ‘the ability to act immediately on an understanding: to act in a way that is not mediated by an interpretation of what is understood’ (McDowell 2009b: 103). The fact of such abilities is, McDowell urges, ‘fundamental’ (*ibid.*).

Yet, as I argued at the close of the final chapter, McDowell is correct that we must bring a halt to any ultimately quixotic search for a hoped-for “something else” that will supposedly bridge the unbridgeable gap between a (normatively inert) “rule-expression” and mere brute propensive behaviour by virtue of holding fast to the idea of action which manifests understanding without mediation by interpretation. Nevertheless, something further is required. McDowell does certainly point to a fundamental insight of Wittgenstein’s diagnosis of the source of misunderstanding at *PI* §201—‘...there is an inclination to say: every action according to the rule is an interpretation’—and holding fast to the idea of behaviour animated by understanding, unmediated by interpretation,

is certainly a crucial feature of bringing a halt to this inclination. Yet, if our insight goes no further than this, then there is a real danger that the appeal to behaviour animated by unmediated understanding will be heard in the wrong key, and will lead our thought to remain enmired in confusion: “Yes, but *how* is such unmediated understanding possible?!” And then the familiar reverberations of thought between conceptual nihilism and superlative fantasy are free to continue as before, imprisoned still within the “gangway with rigid walls” (with, perhaps, the only novelty being an altered conception of the hoped-for “third-way” which, we suppose, will enable us to sail beyond Scylla and Charybdis).

In order to finally bring such incessant habits of thought to a standstill, however, we are required to view the appeal to unmediated understanding as the hinge, around which our entire approach to the putative problems must be rotated. That is to say, in order to fully grasp Wittgenstein’s dissolution of the paradox we must become reflexively aware that what is being urged is precisely *a reorientation in our thinking*. ‘The fundamental fact here,’ Wittgenstein asserts at *PI* §125, is ‘that *we lay down rules*, a technique for a game’ and then become hopelessly entangled within these rules: ‘This entanglement in our rules,’ the remark continues, ‘is what we want to understand.’ Thus, the sense in which we become hopelessly confused, in our reflection, now becomes the primary topic of discussion. We must become self-consciously aware that the source of our problems is precisely that which, during our reflections on meaning and mind, *we are inclined to say*.

To this end—and despite what we may, naturally, have thought all along—the principal topic of Wittgenstein’s discussion is not so much meaning, mind, and intentionality but, rather, *the habitual movements of thought which are characteristic of our reflections upon such phenomena*. It are these habits of thought which Wittgenstein aims to bring to our self-conscious recognition. And I have attempted to mirror this practice during the first two chapters of the present work. That is to say, I have attempted there to highlight the same problematic habits of thought (the incessant, and hopeless, oscillations between Scylla and Charybdis) as they manifest themselves in Kripke’s and Wright’s respective responses to Wittgenstein’s discussion—which habits of thought, ironically,

embody, in crucial respects, forms of very same deficiencies of thought towards which Wittgenstein is trying to draw our attention, and to warn us against. And if those chapters have seemed repetitive, then this is because, in actual fact, there are deep-seated tendencies whereby our reflection on mental and linguistic concepts repeat the same problematic habits of thought (albeit clothed in slightly different guises).

A highly instructive way to think of what Wittgenstein is doing within the rule-following sections of the *Investigations*, I contend, is to train us out of such deeply-rooted, but problematic, habits of thought. Yet, as I argued in the previous chapter, one of the messages of the discussion which begins at *PI* §138 is that such inclinations and habits are relentlessly stubborn and resistant. Thus, the interlocutor will not rest content with Wittgenstein's repeated reminder that '*The application is still a criterion of understanding*' (*PI* §146). In face of this restlessness, Wittgenstein engages in painstaking and equally stubborn efforts to counter the deeply engrained inclination to say, instead, that every action according to a rule is an interpretation. And such deeply-engrained habits of thought, as I have tried to show throughout the present work, carries over into some of the most important published responses to Wittgenstein's discussion. For the respective readings of Kripke, Wright, and McDowell (before his thought undergoes the required reorientation, at least) display the same inclinations, and habits of thought as Wittgenstein's alter ego. Propelled by explanatory ambitions, they each succumb to a picture excised of norms, which picture irresistibly inclines them to say: every action according to a rule must be an interpretation. And without the required reorientation in the approach to matters of meaning and mind, their thought is condemned to reverberate incessantly, but hopelessly, within the "gangway with rigid walls".

At *PI* §143, Wittgenstein reflects that perhaps 'it is possible to wean a pupil' who makes systematic mistakes during the teacher's attempts to train the pupil into arithmetical practices, 'as from a bad habit': 'Or perhaps one accepts his way of copying and tries to teach him ours as an offshoot, variant of his.' Yet not unlike the reactions to his own

discussion of rule-following embodied , however, Wittgenstein reflects that: ‘—And here too our pupil’s capacity to learn may come to an end.’

In order to self-consciously enact the required reorientation, then, we must renounce our explanatory ambitions. Yet this is no easy matter within a culture which assumes that all and any problem or puzzle requires resolution through scientific explanation (or putative scientific explanation). Aaron Ridley refers to the error of scientism as ‘not so much an intellectual position, as an intellectual condition, or syndrome’—an ‘endemic’ but baseless habit of thought (Ridley 2004: 11). Reflecting on his own former peonage to this intellectual syndrome in a remembrance of philosophical writings past, Ridley recalls the source of the problem as habitually, but unreflectively, importing into philosophical investigation a conception of objectivity borrowed from the natural sciences: ‘*A picture held us captive*’ (*PI* §115).

Given the constant oscillations betwixt and between normative inertness (and therefore conceptual nihilism), on the one hand, and the queer mythology of superlative rules, on the other, we can best describe this intellectual syndrome as a form of cyclothymia. Trying to bring such oscillations to a standstill by marching forward in the pursuit of some kind of pseudo-scientific explanation—some “third way” beyond Scylla and Charybdis—is the road to nowhere (we merely suspend ourselves within a constantly frozen, never obtainable “I’ll get back to you...”). What we need to do, instead, is to turn around, go back, and to retrace the steps which led us into confusion. And this necessarily implicates us in meta-philosophical self-reflection. That is to say, unlike those endeavours which are still enmired within the misunderstanding— and which are constantly fixated on the mystery of meaning, understanding, and the like—“turning the whole examination round”, with Wittgenstein, necessarily implicates us in reflection on our own

philosophical thought: 'This entanglement in our rules is what we want to understand (i.e., get a clear view of)' (*PI* §125).⁶⁸

At *PI* §114, Wittgenstein remarks that 'One thinks one is tracing the outline of the thing's nature over and over again', when, in actual fact, 'one is merely tracing round the frame through which we look at it.' Similarly, at *PI* §401, he remarks: 'You interpret a grammatical movement made by yourself as a quasi-physical phenomenon which you are observing' (cf. *PI* §436).

It is an utterly sobering thought that the implications of these remarks on behalf of what often passes as philosophical activity in the modern period (but which has also infected, to an incalculable degree, modern thought more broadly) are merely the fruit of *illusion*—a self-imposed prison occasioned by our own explanatory ambitions, itself occasioned by the envious glance at the immense success of the modern scientist: 'You think that after all you must be weaving a piece of cloth: because you are sitting at a loom—even if it is empty—and going through the motions of weaving' (*PI* §414; cf. *PI* §118)

⁶⁸ Thus, *contra* Wright, there are intentional mental states which do have independence of our judgements upon them. Moreover, as I have argued, we are urged to bring to self-conscious awareness a subset of our own intentional mental states—our thought-processes during philosophical reflection—and to judge their deeply misguided nature.

Bibliography

Bloor, D. (1997) *Wittgenstein, Rules and Institutions*, London: Routledge.

Baker, G. (1981) "Following Wittgenstein: Some Signposts for Philosophical Investigations 143-242" in *Wittgenstein: To Follow a Rule*, London: Routledge & Kegan Paul.

Baker, G.P. & Hacker, P.M.S. (1984) *Scepticism, Rules & Language*, Oxford: Blackwell.

Baker, G.P. and Hacker, P.M.S. (2009a) *Wittgenstein Understanding and Meaning: Second Edition Volume 1 Part 1 - Essays*, Chichester: Wiley-Blackwell.

Baker, G.P. and Hacker, P.M.S. (2009b) *Wittgenstein Understanding and Meaning: Second Edition Volume 1 Part 11 – Exegesis 1–184*, Chichester: Wiley-Blackwell.

Bakhurst, D. (2011) *The Formation of Reason*, Chichester: Wiley-Blackwell.

Bennett, M., Dennett, D., Hacker, P., Searle, J. (2007) *Neuroscience & Philosophy: Brain, Mind, & Language*, New York: Columbia UP.

Bennet, M.R. and Hacker, P.M.S. (2003) *Philosophical Foundations of Neuroscience*, Oxford: Blackwell.

Blackburn, S. (2009a) "The Individual Strikes Back" in Miller, A. & Wright, C. (eds.) *Rule-Following and Meaning*, Chesham: Acumen.

Boghossian, P.A. (1989) "The Rule-Following Considerations", *Mind* Vol. 98, 392.

Braddon-Mitchell, D. and Jackson, F. (2007) *Philosophy of Mind and Cognition: Second Edition*, Oxford: Blackwell.

De Caro, M. & McArthur, D. (eds.) *Naturalism and Normativity*, New York: Columbia UP.

De Gaynesford, M. (2004) *John McDowell*, Cambridge: Polity Press.

Dennett, D.C. (1987) *The Intentional Stance*, London: MIT.

Diamond, C. (ed.) (1976) *Wittgenstein's Lectures on the Foundations of Mathematics*, London: The University of Chicago Press.

Fodor, J.A. (1975) *The Language of Thought*, Cambridge, Massachusetts: Harvard UP.

- Fodor, J.A. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, London: MIT.
- Glock, H-J. (2003) *Quine and Davidson on Language, Thought and Reality*, Cambridge: CUP.
- Goldfarb, W. (2009b) "Kripke on Wittgenstein on Rules" in Miller, A. & Wright, C. (eds.) *Rule-Following and Meaning*, Chesham: Acumen.
- Hacker, P.M.S. (1993a) *Wittgenstein: Meaning and Mind Volume 3 Part 1: Essays*, Oxford: Blackwell.
- Hacker, P.M.S. (1993b) *Wittgenstein: Meaning and Mind Volume 3 Part 11: Exegesis 243-427*, Oxford: Blackwell.
- Hattiangadi, A. (2007) *Oughts and Thoughts Rule-Following and the Normativity of Content*, Oxford: OUP.
- Haugeland, J. (1985) *Artificial Intelligence: The Very Idea*, London: MIT.
- Haugeland, J. (1997) *Mind Design 11*, London: MIT.
- Haugeland, J. (1998) *Having Thought*, London: Harvard UP.
- Horwich, P. (2009e) "Meaning, Use and Truth" in Miller, A. & Wright, C. (eds.) *Rule-Following and Meaning*, Chesham: Acumen.
- Kenny, A. (2006) *Wittgenstein: Revised Edition*, Oxford: Blackwell.
- Kripke, S. (1982) *Wittgenstein on Rules and Private Language*, Oxford: Blackwell.
- Kusch, M. (2006) *A Sceptical Guide to Meaning and Rules*, Chesham: Acumen.
- Malcolm, N. (1993) *Wittgenstein A Religious Point of View?*, Abingdon: Routledge.
- McCulloch, G. (1995) *The Mind and its World*, London: Routledge.
- McDowell, J. (1996) *Mind and World*, London: Harvard UP.
- McDowell, J. (1998a) "Wittgenstein on Following a Rule" in *Mind, Value, & Reality*, London: Harvard UP.
- McDowell, J. (1998b) "Meaning and Intentionality in Wittgenstein's Later Philosophy" in *Mind, Value, & Reality*, London: Harvard UP.

- McDowell, J. (1998c) "One Strand in the Private Language Argument" in *Mind, Value, & Reality*, London: Harvard UP.
- McDowell, J. (1998d) "Intentionality and Interiority in Wittgenstein" in *Mind, Value, & Reality*, London: Harvard UP.
- McDowell, J. (1998e) "Non-Cognitivism and Rule-Following" in *Mind, Value, & Reality*, London: Harvard UP.
- McDowell, J. (2009a) "Are Meaning, Understanding, etc., Definite States?" in *The Engaged Intellect*, London: Harvard UP.
- McDowell, J. (2009b) "How Not to Read *Philosophical Investigations*: Brandom's Wittgenstein in *The Engaged Intellect*, London: Harvard UP.
- McDowell, J. (2010) "Autonomy and Its Burdens", *The Harvard Review of Philosophy*, vol. XVII
- McGinn, C. (1987) *Wittgenstein On Meaning*, Oxford: Blackwell.
- Millikan, R.G. (2009d) "Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox" in Miller, A. & Wright, C. (eds.) *Rule-Following and Meaning*, Chesham: Acumen.
- Morris, M. (2007) *An Introduction to the Philosophy of Language*, Cambridge: CUP.
- Papineau, D. (1993) *Philosophical Naturalism*, Oxford: Blackwell.
- Pears, D. (1997) *Wittgenstein: Second Edition*, London: Fontana Press.
- Pettit, P. (2009c) "The Reality of Rule-Following" in Miller, A. & Wright, C. (eds.) *Rule-Following and Meaning*, Chesham: Acumen.
- Phillips, D.Z. (2001) *Religion and the Hermeneutics of Contemplation*, Cambridge: CUP.
- Puhl, K. (Ed.) (1991) *Meaning Scepticism*, Berlin: Walter de Gruyter.
- Ridley, A. (2004) *The Philosophy of Music*, Edinburgh: Edinburgh UP.
- Ritchie, J. (2008) *Understanding Naturalism*, Stocksfield: Acumen.
- Schönbaumsfeld, G. (2007) *A Confusion of the Spheres*, Oxford: OUP.
- Schönbaumsfeld, G. (2016) *The Illusion of Doubt*, Oxford: OUP.
- Sterelny, K. (1990) *The Representational Theory of Mind: An Introduction*, Oxford: Blackwell.
- Stern, D.G. (1995) *Wittgenstein on Mind and Language*, Oxford: OUP.

- Stern, D.G. (2004) *Wittgenstein's Philosophical Investigations: An Introduction*, Cambridge: CUP.
- Thornton, T. (1998) *Wittgenstein on Language and Thought: The Philosophy of Content*, Edinburgh: Edinburgh UP.
- Thornton, T. (2004) *John McDowell*, Chesham: Acumen.
- Winch, P. (1990) *The Idea of a Social Science and its Relation to Philosophy: Second Edition*, Abingdon: Routledge.
- Wittgenstein, L. (1953) *Philosophical Investigations*, Anscombe, G.E.M. (trans.), Oxford: Blackwell.
- Wittgenstein, L. (1969) *The Blue and Brown Books: Second Edition*, Oxford: Blackwell.
- Wittgenstein, L. (1972) *On Certainty*, Anscombe, G.E.M. & von Wright, G.H. (eds.), Paul, D. & Anscombe, G.E.M. (trans.), London: Harper & Rowe.
- Wittgenstein, L. (1974) *Tractatus Logico-Philosophicus*, Pears, D.F. and McGuinness, B.F. (trans), London: Routledge.
- Wittgenstein, L. (1978) *Remarks on the Foundations of Mathematics*, von Wright, G.H. & Rhees, R. & Anscombe, G.E.M. (eds.), Anscombe, G.E.M. (trans.), Oxford: Blackwell.
- Wittgenstein, L. (2006) *Culture & Value: Revised Edition*, von Wright, G.H. & Nyman H. (eds.), Winch, P. (trans.), Oxford: Blackwell.
- Wright, C. (1980) *Wittgenstein on the Foundations of Mathematics*, London: Duckworth.
- Wright, C. (1993) *Realism, Meaning & Truth: Second Edition*, Oxford: Blackwell.
- Wright, C. (2001) *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*, London: Harvard UP.
- Wright, C. (2001a) "Following a Rule" in *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*, London: Harvard UP.
- Wright, C. (2001b) "Rule-Following, Objectivity and the Theory of Meaning" in *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*, London: Harvard UP.
- Wright, C. (2001c) "Rule-Following, Meaning and Constructivism" in *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*, London: Harvard UP.
- Wright, C. (2001d) "Kripke's Account of the Argument against Private Language" in *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*, London: Harvard UP.

- Wright, C. (2001e) "On Making Up One's Mind: Wittgenstein on Intention" in *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*, London: Harvard UP.
- Wright, C. (2001f) "Excerpts from a Critical Study of Colin McGinn's *Wittgenstein on Meaning*" in *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*, London: Harvard UP.
- Wright, C. (2001g) "Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics" in *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*, London: Harvard UP.
- Wright, C. (2001h) "Does *Philosophical Investigations* 258-60 Suggest a Cogent Argument against Private Language?" in *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*, London: Harvard UP.
- Wright, C. (2001i) "Wittgenstein's Later Philosophy of Mind: Sensation, Privacy and Intention" in *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*, London: Harvard UP.
- Wright, C. (2001j) "The Problem of Self-Knowledge (I)" in *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*, London: Harvard UP.
- Wright, C. (2001k) "The Problem of Self-Knowledge (II)" in *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*, London: Harvard UP.
- Wright, C. (2001l) "Skolem and the Sceptic" in *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*, London: Harvard UP.
- Zalabardo, J.L. (2009) "Kripke's Normativity Argument" in Miller, A. & Wright, C. (eds.) *Rule-Following and Meaning*, Chesham: Acumen.