

Rule-Following, Normativity & Objectivity

An Analysis of McDowell's
'Wittgenstein on Following a Rule'

James Alexander Cross

UCL

Department of Philosophy

MPhil Stud

DECLARATION

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

ABSTRACT

This essay offers a close analysis and critique of the complex theoretical arguments in John McDowell's 1984 paper, *Wittgenstein on Following a Rule*, a seminal discussion of rule-following, normativity and objectivity that still stands in need of a thorough, clear analysis on its own terms. My aim is to clarify and assess the arguments McDowell makes against the views expressed by Crispin Wright in *Wittgenstein on the Foundations of Mathematics* and Saul Kripke in *Wittgenstein on Rules and Private Language*. I also offer analysis of relevant parts of these works as necessary. All three texts juxtapose attempts to expound the views of the historical Wittgenstein with attempts to argue for them on his behalf, but the philosophical issues under debate are entirely separated from that interpretive context in my discussion as they merit analysis in their own right. For simplicity I treat the views each commentator attributes to Wittgenstein as if they were the commentator's own.

Chapter 1 outlines and analyses the basic premises, terminology and some central arguments of the debate, focusing on analysis of the idea that understanding the meaning of an expression involves a kind of contract. Chapter 2 focuses on Wright's account of the 'contract' picture and McDowell's attempts to criticise it. In chapter 3 I analyse a notion of understanding that McDowell wants to reject, and which he believes is the source of Kripke's celebrated 'sceptical paradox'. The beginning of chapter 4 brings together all the lessons learned in WFR about rule-following, normativity and objectivity in order to suggest requirements on a plausible alternative contractual account of meaning and understanding. In §4.3 I sketch an account that McDowell might want to adopt, and in §4.4 I test it out on the problems identified.

CONTENTS

Declaration	2
Abstract.....	3
Introduction	5
1 Contractual Understanding.....	7
1.1 The Minimal Contractual Picture	7
1.2 Realism & Objectivity.....	11
1.3 Determinacy	14
1.4 Patterns of Application.....	17
1.5 Summary	21
2 Wright’s Anti-Realism	23
2.1 Anti-Realism & Objectivity	23
2.2 Communally Established Correctness	26
2.3 Some Problems for Wright’s Anti-Realism	37
2.4 The Basic Level.....	43
3 Understanding as Interpretation.....	49
3.1 Interpretations & Paradox.....	49
3.2 Super-Rigidity	56
4 Meaning-Perspectivism	60
4.1 McDowell’s Task	60
4.2 The Big Dilemma.....	63
4.3 Meaning-Perspectivism	66
4.4 Dodging the Big Dilemma.....	72
4.5 Conclusion	75
Bibliography.....	77
Quick-Reference Glossary	79

INTRODUCTION

The aim of this essay is to present a close analysis and critique of the complex theoretical arguments in John McDowell's 1984 paper, *Wittgenstein on Following a Rule* (referred to as 'WFR' hereafter)¹, a seminal discussion of rule-following, normativity and objectivity that still stands in need of a thorough, clear analysis on its own terms. My aim is to clarify and assess the arguments McDowell makes against the views expressed by Crispin Wright in *Wittgenstein on the Foundations of Mathematics* (referred to as 'W') and Saul Kripke in *Wittgenstein on Rules and Private Language* (labelled 'K').

The interest of WFR, for the present purposes, is that McDowell seems to pave the way there for an account of rule-following, normativity and objectivity in opposition to those offered by Wright and Kripke. The content of this account is far from explicit, however, so my aim is to examine those few clear commitments that McDowell does make and the assumptions underlying his criticisms of Wright and Kripke. After due consideration and analysis of the various issues McDowell confronts in the paper, I will offer a sketch of a positive account I call 'meaning-perspectivism' that is essentially an interpretation of McDowell's commitments.

WFR is extraordinary in its attempts to fight battles on several fronts simultaneously. First, it aims to show that McDowell's reading of the later Wittgenstein's reflections on rule-following (in the *Philosophical Investigations* and elsewhere) is correct, and a superior interpretation to those proposed by Wright and Kripke. Second, McDowell aims to show that an anti-realist conception of language and meaning is disastrous, both by attack on the programme of anti-realism itself and by exposing the supposed consequences of taking it seriously. Third, he aims to preserve an account of objectivity in the face of problems perceived in the readings Wright and Kripke present. Finally, McDowell aims to give some measure of positive explanation of how meaning relates normatively to understanding.

Here I will only analyse McDowell's approach to the last three of these four tasks, treating WFR as a piece of philosophical theory that embodies claims about rule-following, normativity and objectivity. The fact that the claims are argued on Wittgenstein's behalf is of no philosophical importance, and I will make almost no

¹ In *Synthese* 58, no. 3, pp.325-363. It has been reprinted in several collections, including McDowell's *Mind, Value and Reality* (London: Harvard UP, 2002) where it appears with minor cosmetic revisions, but I refer to the pagination of the original 1984 print as I presume the reader will be able to access an electronic copy of it on the publisher's website.

Rule-Following, Normativity & Objectivity

mention of the McDowell's project of interpreting and defending his work, nor of Wright and Kripke's efforts to do the same. To this effect, claims and arguments I attribute to these three may all be claims and arguments they in turn attribute to Wittgenstein. The quality of such claims and arguments clearly does not depend on who endorses them.

1 CONTRACTUAL UNDERSTANDING

1.1 THE MINIMAL CONTRACTUAL PICTURE

'*Wittgenstein on Following a Rule*' opens with the introduction of a supposedly intuitive idea about the relation between meaning and understanding, called the 'contractual picture', expressly held to be intuitively appealing by both McDowell and Wright. The *minimal form* of the contractual picture amounts to something like the following:

Minimal Contractual Picture Understanding the meaning of an expression imposes some kind of constraints on one's future actions, amounting to a commitment to make only a certain kind of use of the expression.²

We should note that McDowell's own formulation involves not only understanding of expressions, but 'grasp of concepts' (WFR 325). I have opted to restrict my analysis to the expression case for clarity, as it is unclear precisely what McDowell thinks the relationship between concepts and expressions amounts to.³ Only a few of his conclusions pertain specifically to concepts, and restricting our analysis of the minimal contractual picture (and its more complex variants later on) to an examination of how it applies to expressions will not prevent us from understanding them, as it seems to apply to concepts in exactly the same way.

It should be noted at the outset that the discussion presented in WFR takes place in the heyday of the thesis that 'meaning is normative', broad acceptance of which I will call 'normativism'. The idea comes from Kripke's claim that "the relation of meaning and intention to future action is *normative*, not *descriptive*" (K, p.37). That some form of this idea was accepted by all parties is clearly the basis of the supposedly intuitive appeal of the contractual picture. However, recent discussions of the meaning-normativity thesis have raised doubts about its truth, so a call has been made for its justification. Answering this call is not my project here. The aim is to see what we can learn within the framework of this particular debate, given the shared assumptions common to the

² This is a combination of McDowell and Wright's slightly differing formulations of the contractual picture. Cf. WFR 325 & Wright 1980a (hereafter labelled 'W'), p.19.

³ For example, he talks about 'deploying' concepts without explaining what counts as a deployment (cf. WFR 326). We could speculate, but it would be unnecessary.

parties who were engaged in it. The most basic assumption is something like the following:

Normativity of Meaning	A condition for the meaningfulness of an expression is that some uses of it are correct and that the correct uses of it are normatively prescribed.
-------------------------------	---

The fact that there are correct and incorrect ways of using the expression is taken to entail that there are certain ways it *should* be used, thus introducing normative impositions and prescriptions. Note that in the formulation offered here, anyone who uses the expression is subject to the normative imposition to use it correctly whether they understand it or not. We might prefer a weaker thesis that the prescriptions in question only apply to those who understand the expression, but pending motivations for this we should opt for the stronger thesis—misunderstanding the meaning of an expression does not obviously detract from the correctness or incorrectness of your efforts to use it. The fact that correctness attaches to certain uses of the expression is enough to introduce some kind of universal prescription, for anyone who uses it, to use it correctly. Still, we might want to insist that in order to count as using an expression one has to have some minimal comprehension of what one is doing—perhaps we should not count animals mimicking the sound patterns of human speech as using expressions at all.

The contractual picture is supposed to grow from the general idea that meaning is normative because the latter involves the conception of correctness as *action-guiding*. Understanding an expression seems to involve some degree of knowledge of situations when it would be correct to use the expression, hence knowledge of how the expression *should* be used, and I suppose the idea is that knowing what one should do amounts to acknowledging that one should do it and allowing one's actions to be guided accordingly. In order to count as understanding the meaning of an expression, one needs to take on the relevant commitment to correct use of it. Provided we explain what the 'commitment' in the contractual picture amounts to, my understanding an expression could thus feasibly entail my making a commitment only to use that expression in a certain range of ways.

On the other hand, it is certainly a mistake to assume that the normativity of meaning has the minimal contractual picture as an immediate consequence, as the idea of committing to a range of usage is just one kind of normativism. McDowell and Wright both apparently accept it as an 'intuitive' consequence, though Kripke's silence on the

matter leaves it possible that McDowell's criticisms of him are off-target, despite not being directly reliant on acceptance of this 'intuition'. This issue will not have much impact on our concerns, as it turns out.

The envisaged constraints on action, the content of this commitment, seem to correspond to whether it would be correct or incorrect to apply the expression, such that if S understands the meaning of E then S's use of E is constrained by a commitment S undertakes as a condition of understanding the meaning of E. However, this is not just a commitment to use E correctly at the time when S acquires an understanding of E, which would not be action-guiding so much as definitional. If S understands E at t iff S uses E correctly at t, then S does not have time to 'commit' to anything before the action takes place. S's actions are not 'guided' by his understanding in this case, because his understanding consists in the fact that his actions are a certain way already. Rather, the commitment S undertakes is a commitment regarding his *subsequent* usage of E, and his understanding of E is not identical with this commitment. The minimal contractual picture thus seems to involve the following idea:

- (C) S understands the meaning of E \rightarrow S undertakes to use E only in a range of ways R whenever S uses E.

So when S acquires an understanding of E, S becomes subject to the prescription that he constrain *all* his use of E subsequent to the point of acquisition to a range of correct uses of E. However, C risks being over-strong unless further elaborated because the 'undertaking' here does not obviously allow for S's making occasional mistakes amidst his otherwise correct usage of E, which is something we clearly need to account for. I will return to questions about fallibility in §1.3.

It should also be noted that the minimal contractual picture merely proposes a necessary condition on understanding the meaning of E; undertaking a commitment of the kind represented by C is necessary but not sufficient for having an understanding of the meaning of E. Having an understanding plausibly amounts to more than making a particular kind of commitment and may involve other elements like mental states, knowledge of certain propositions and so on. What else 'having an understanding' amounts to is a further question, which we will return to in §4.3.

We need to understand precisely what the commitments in question amount to. McDowell claims that the intuitive contractual picture entails that if we fail to honour our commitments, we fail "to obey the dictates of the meaning we have grasped" (WFR

325). This seems to mean that someone who fails in this way ceases to understand what he understood before, but then we need to know what that amounts to in practice. An alternative interpretation of this might be much stronger, such that the understanding of E we ascribe to S would turn out to have been misplaced if S subsequently failed to honour the commitment to restrict his usage to range R. On this conception, if S does not uphold the commitments required of him by an understanding of the meaning of E then S never did understand E after all.

It seems unlikely that the stronger conception is what McDowell has in mind, because the arrangement described is not recognisable as something we would call a commitment, and the discussion here is focused on the specific component of S's understanding of E that is constituted by S's commitment to restrict his usage to range R. When a commitment is broken by failing to honour it, it is not retrospectively unmade, but rather ceases to apply after a period when it did apply: if S commits at t to doing A at t', it does not follow that if S does not do A at t', S did not commit at t to doing A at t'. Rather, a commitment is like a property that attaches to S and remains attached to S for as long as S does the right things, i.e. upholds the commitment—similar to how promises function.

A corollary of the minimal contractual picture seems to be that if no commitments are made to use an expression correctly, then no constraints on usage are in place, and one does not have an understanding of the expression in question. Similarly, if the wrong constraints are in place and S undertakes to use E in the wrong range of ways, then the understanding that sanctioned those constraints is *'faulty'*. This might be called a *'misunderstanding'*, but to avoid suggesting that it fails to be *any* kind of understanding I will stick to the term *'faulty understanding'*. This is because the minimal contractual picture seems to allow that S could attempt to learn the meaning of E, but come to acquire an understanding not of E but some similar expression E* used in a nearly identical way. Here it would be fair to say that S has an understanding not of E but of E*. However, S's intention was to learn the meaning of E, so there is an important sense in which S has an understanding of E, but one which is faulty. The need for this becomes apparent if we allow that E* could be a new expression of S's own invention, which corresponds to E except in cases where S would use E incorrectly, which would be correct uses of E*—in such cases, we would be more inclined to focus on S's commitment as it relates to E than to the new invented expression E*, to rule out the idea that every mistake in understanding is a proper understanding of something else. I will accordingly leave the option of calling this a *'faulty understanding of E'* open. I

will refer to the case where someone's understanding of an expression is not faulty in this way, so where the constraints their understanding imposes on their use of it correspond to the range of correct uses, as a case of *'proper'* understanding.

1.2 REALISM & OBJECTIVITY

The possibility of faulty understanding—the possibility that the constraints on use imposed by an individual's understanding of an expression may not reflect the range of correct ways of using it—involves some notion of objectivity with respect to correctness. As such, objectivity is a central theme in WFR. The question of how the correctness of a particular use of an expression is determined is important here, as finding this out will, it seems, provide us with an idea of how an understanding comes to qualify as proper rather than faulty. Is correctness a property assigned to actions by groups of people, or is it a property that attaches to them by virtue of some thinker-independent features of reality—and if neither of these, then what?

It might be tempting to think that language is a human creation, and that the facts about how to use it correctly are therefore up to humans. But this idea entails counter-intuitive consequences: if the correctness of a claim about mathematical truth depends ultimately upon human decisions about the circumstances under which such claims can be made, then it might seem to follow that the truth of a mathematical claim is a product of constructed language practices and not, as we often think, anything to do with facts about the universe as it exists independently of us. The problem does not go away if we decide that the claim is correct because it is true, because whether it is correct to say the claim is true falls to us as well...and so on. Motivated by this kind of worry, McDowell argues that allowing the reduction of all notions of objective correctness to something like 'communal assent' on correctness results in the eradication of any intuitive notion of objectivity.⁴ If statements of all kinds rely only on human attitudes to be correct, then the correctness of claims about the external world—their truth—seems to owe no allegiance to facts about the world as it exists independent of our reckoning. However, the fact that language is created by humans does not entail that the facts about how to use it are merely a matter of consensus unless further assumptions are adopted. Further motivation for entertaining this idea is required.

⁴ What a 'reduction' amounts to must be analysed: see §2.2.

McDowell's principle antagonist in WFR is Crispin Wright, who argues that we should dispense with 'realist' notions like mind-independent objectivity and objective standards on which the correctness of an application of a particular expression depends. Wright accordingly sets out to make revisionary 'anti-realist' claims about objectivity, particularly in mathematics, though the aim is apparently not to make revisions which are totally unpalatable.

The motivation for this call to revise our notion of objective correctness seems to draw on the following kind of reasoning. Whenever we agree that our most refined criteria for determining the correctness of some application of an expression are satisfied, the most we can know is that we agree that they are satisfied, not that they *are* satisfied in some way independent of our judgment that they are. For the assertion 'snow is white', we accept that it is correct just in case we agree that our most stringent criteria for the correctness of the statement 'snow is white' have been satisfied. If we all sincerely judge that some application is correct, then a fact of its correctness independent of that sincere communal judgement of correctness plays no role—for in such cases, if we sincerely judge that the application is correct relative to such facts then by definition it will not make a difference whether that is really true because no-one could tell. Since the most we can offer is our most sincere judgments about such facts of correctness, with no way of knowing that any judgment reflects such facts, it seems that such facts themselves play no correctness-determining role in our language practices.

The notion of objectivity attacked here is one we might describe as 'transcendent', which means that even our even the judgments we hold in the highest regard may still be false. It is a 'realist' notion, in that it takes facts to depend on reality and not on thinkers. Based on these ideas, Wright thinks that 'realism' entails the following thesis:

Logical Independence (LI) For particular true statements it is either unnecessary or insufficient, or both, to meet our most refined criteria of acceptability in order to be true. (W, p.199)

However, there is reason to doubt that many realists would accept this condition, and hence to reject the particular form of argument above as an argument against realism in general. The problem lies in the expression 'most refined criteria of acceptability,' which is slightly odd. It seems plausible that for any statement, our most refined criteria for its acceptability would demand the truth of that statement. So unless we grant that there is a distinction to be made between our notion of truth, as applied in

the case of statements of the form 'P is true', and some other kind of superior, transcendent 'truth proper', the offered thesis is rendered incoherent: meeting our most refined criteria of acceptability means being true. But if we acknowledged this distinction, surely we would add truth of the transcendent kind to our most refined criteria of acceptability! The realist's failing to demand this would seem like a concession to external world sceptics that we cannot know the real facts. If 'most refined criteria of acceptability' does not include things like truth, then it is mysterious why not.

Indeed, McDowell expressly rejects the following view, 'Problematic Realism':

Problematic Realism A genuine fact must be a matter of the way things are in themselves, utterly independently of us. So a genuinely true judgment must be, at least potentially, an exercise of pure thought; if human nature is necessarily implicated in the very formulation of the judgment, that precludes our thinking of the corresponding fact as properly independent of us, and hence as a proper fact at all. (WFR 351)

At first glance this might not seem so bad. It does not intuitively seem that worldly objects depend on us for their existence, and if facts are just configurations of such objects then the truth of judgments about such facts seems determined by the objects, and not by virtue of the way they relate to humans. However, this gloss is not quite what 'Problematic Realism' amounts to. The idea is rather that the facts and objects in question are totally alien to us, and that our status as human beings inhabiting the same world as them must be separable from our ability to make true judgments about them, if those judgments are to be true. This is what McDowell means by 'human nature' in opposition to 'pure thought'. The perspective adopted by human beings as a consequence of their interactions with the world is taken by Problematic Realism to be a bias that prevents them from getting to the truth. However, *realism* does not need to adopt this view. A realist could accept that human beings have a certain mode of existence and still make true judgments which are not merely true 'as far as we can tell'. This is the line I read McDowell as taking in WFR, and it explains his opposition to both 'Problematic Realism' and 'anti-realism'.

1.3 DETERMINACY

The commitments described in the contractual picture are supposed to have *determinate* consequences for future language use, so we need to understand what kind of determinacy is at stake.⁵ It apparently involves facts about whether any application of an expression to a particular situation would be correct, which I shall sometimes call the ‘correctness-value’ of the application.⁶ For McDowell (as I read him), statements about whether a particular application of an expression would be correct are timelessly true or false. This is part of what the relevant notion of determinacy amounts to, and I will formulate it as follows:

- (M1)** For any possible application *a* of an expression E, *a* is either correct or incorrect.⁷

How does M1 relate to a particular understanding of E on the contractual picture? ‘C’ (from §1.1) seems to explain how commitments are supposed to relate to understanding here, so we can tie correctness into the picture of commitments by adapting ‘C’ to produce the following account of ‘proper’ understanding:

- (M2)** S’s understanding of the meaning of E is proper \leftrightarrow S undertakes to apply E only in a range of ways R, where R is the range of correct ways of applying E.

M2 leaves it open whether S needs to know the contents of range R, and so whether the undertaking in question is a conscious commitment to certain things or something more tacit (as in most people’s commitment to avoid shoplifting), but I suspect it makes better sense as something tacit and will assume so when referring back to M2. If we accept M1 and M2 together, we have a version of the contractual picture where every possible application of an expression E has a determinate correctness-value, and where a ‘proper’ understanding of E means committing to applying E only in the correct ways. M3 below gives a corresponding definition of ‘faulty’ understanding:

⁵ Cf. WFR 325 (quoted above, §1.1).

⁶ I will use expressions of the form ‘the correctness value of *x*’ in a non-substantive way to replace those of the form ‘whether *x* is correct or incorrect’, simply for convenience. Anything that can be correct or incorrect thus has a correctness-value, just as anything that can be true or false has a truth-value, before we make any claims about what the notions of correctness and truth amount to.

⁷ Questions of vagueness may arise here, but they can be safely suspended as their relevance is minimal.

- (M3)** S's understanding of the meaning of E is faulty \leftrightarrow S undertakes to apply E only in a range of ways R*, where R* is not the range of correct ways of applying E.

Accepting M1-3 seems to give us a *realist* contractual picture. This is because, by M1, range R is a range of applications with determinate truth values. Replacing M1 with a different principle will eventually allow us (§1.4) to produce an *anti-realist* contractual picture.

A word on fallibility is appropriate at this point. We may be forgiven for finding the normative claims about understanding puzzling, since it is perfectly consistent for someone to understand an expression properly and then go on to use it incorrectly sometimes; errors in performing additions are not necessarily symptomatic of a misunderstanding of addition, even if it is the case that one ought not to make errors. But if it is possible to make mistakes without stepping outside of the constraints the contractual picture says are imposed by a proper understanding, then those constraints must pertain to something other than our ability to err in this way.

The kind of constraint in question does not ensure the individual's adherence to it, but it does ensure that actions of a certain kind, those that fail to adhere, will qualify as incorrect—and this amounts to a negative prescription, such that the individuals should not do it. But M2 does not seem to specify whether occasionally failing to act within range R entails breaking one's commitment, though it does seem to say that breaking one's commitment entails ceasing to understand the use of E. Based on considerations so far, the following might seem apt:

- (M4)** S undertakes to apply E only in a range of ways R \rightarrow S can accidentally apply E in a way falling outside R.

This says that as long as the incorrect application is a result of an accident, perhaps a cognitive slip of some kind, S can be said to maintain his undertaking (commitment) so this alone does not negate his understanding of the meaning of E. This allows us to reconcile ordinary fallibility with the notion of determinate commitments with respect to future action. However, M4 leaves us no specification of the difference between an 'accident' and a momentary switch to a faulty understanding, both of which could be indicated by the same incorrect application. We need to add that what makes something an 'accident' is a failure to uphold the commitment, and not a total break or alteration of the commitment.

On the contractual picture, a mistake of the kind where someone with a proper (non-faulty) understanding of addition answers “5” to the question “What is 68+57?” is a mistake in two respects: a) by virtue of its failure to use the expression “5” within the range of ways specified by that individual’s understanding, and b) by virtue of its being objectively incorrect to use the expression “5” in this way.⁸ The distinction is important because the range specified by an understanding may not coincide with the range of objectively correct ways to use the expression, as described by the M3 account of faulty understanding. Only if the understanding is ‘proper’ will the two ranges coincide, as per M2.

Accordingly, it is possible for S to uphold the commitment made as a condition of his acquiring and understanding of the meaning of E, and apply it ‘correctly relative to his understanding’—that is, in a way from among the range of ways specified as correct by S’s understanding—but simultaneously apply it incorrectly in objective terms. If S has a faulty understanding of the meaning of E, then S may apply E correctly on most occasions but apparently have a ‘blind spot’ for a particular range of situations, consistently applying E to them when it is incorrect to do so. Assuming S upholds the commitments required by his understanding, the explanation for the ‘blind spot’ is that on S’s faulty understanding of the meaning of E, his systematic errors are not errors but possible applications of E that fall within the range R* but not, it turns out, within the range R (see M2 and M3 above). The converse would apply if S upheld his commitment but consistently treated certain correct applications of E as incorrect. S’s commitment is upheld, so S does what his understanding prescribes, but not what is objectively correct, so objectively prescribed. This is the difference between ‘*relative correctness*’ and ‘*objective correctness*’. The word ‘correct’ as it appears in M2-4 means ‘objectively correct’, although for reasons that will become apparent⁹, we should allow that in M1 it means both objectively (in)correct and (in)correct relative to any given understanding. Thus M1 should be revised as follows :

- (M1)** For any possible application *a* of an expression E, *a* is either objectively correct or objectively incorrect, and for any understanding of the meaning of E, *a* is either correct or incorrect relative to that understanding.

Hopefully this distinction makes it apparent that properness in M2 can be understood in terms of a perfect overlap between relative and objective correctness, such that the

⁸ What ‘objective incorrectness’ amounts to will depend on whether we adopt M1 or some alternative principle.

⁹ Cf. §2.1.

range R of objectively correct ways of using E is precisely the range of usage that S's understanding prescribes for S. M3 makes it possible for someone to have an understanding of the meaning of E such that correctness relative to that understanding does not entail objective correctness, but what if S makes an objectively correct application of E that is incorrect relative to his understanding? This would not mean that S properly understands E, because in this case S is not committed to making this kind of application, and being so committed is a condition on proper understanding.

1.4 PATTERNS OF APPLICATION

The 'minimal contractual picture' seems to be well characterised by the conjunction of C, M2, M3 & M4, provided we can overlook the absence of any appeal to *concepts* from these theses, as I suggested we should for now (§1.1). Sadly, the picture is not formulated quite so explicitly in WFR. The prominent gloss is provided by a quotation from W:

...we are 'committed to certain patterns of linguistic usage by the meanings we attach to expressions'. (WFR 325)¹⁰

This 'pattern' idea is unhelpfully metaphorical. 'Pattern' seems to mean an observable regularity in language usage, such as using E in a certain range of ways on different occasions, and the patterns to which we are committed are the objectively correct patterns of usage of the expressions whose meanings we understand, corresponding to the range R in M2. 'Conforming to a pattern of application for E' just seems to mean 'restricting one's usage to a certain range of possible ways of using E'.

The pattern metaphor arises by analogy with the practice of continuing a series in mathematics. If someone (properly) understands the rule that generates the series, then if he does try to continue it he *should* produce terms that do, in fact, correctly continue the series. McDowell reluctantly¹¹ accepts the idea as a reasonable gloss on the contractual picture and adopts the terminology, apparently for the sake of better connecting with Wright's explicit critique of it as an intuitive notion which is almost entirely *wrong* and in need of reinterpretation. According to McDowell, the reinterpretation Wright goes on to offer is unacceptable (WFR 326).

Wright claims that "there is in our understanding of a concept no rigid, advance determination of what is to count as its correct application" (W, p.21), and for present

¹⁰ The embedded quotation is from W, p.21.

¹¹ Cf. WFR 359n3, also 361n41.

purposes we can treat this as a claim about understanding the meaning of an expression. This is simply a denial of M1. Instead, Wright holds that the correctness-value of some postulated future application is indeterminate until that application is actually made. This is because he claims the objective correctness-value of any particular application is determined by communal evaluation, and communal evaluations are not committed to internal consistency across time—what everyone thinks is correct today may easily be regarded as incorrect at some future time for no more reason than that our attitudes are open to change. So preserving the pattern idea means making sense of how one can restrict one's applications of E to an indeterminate range, as whether any particular way of using E falls within R is in some sense indeterminate.

McDowell's gloss on Wright's offering is roughly that the commitment in question is a commitment to using the expression in a way that achieves the approval of the community, a.k.a. 'objective correctness'. I suppose the role of a personal understanding here would be reduced to offering some kind of inductive guidance about how I should use an expression. Perhaps I have enough in common with my peers that what I think would be correct is the same thing that the community would think correct. M2 & M3 thus amount to the following: a proper understanding is an understanding that commits S to using E in a way that tracks communal consensus, in the manner of a voter who always manages to vote for the winning candidate in elections; a faulty understanding is an understanding that commits S to using E in ways that *mostly* track communal consensus with occasional divergence. Wright might replace M1 with the following:

(W1) For any application *a* of an expression E, *a* is either correct or incorrect \leftrightarrow *a* is an actual (not merely possible) application of E.

Objective correctness-values for possible applications are indeterminate, but determinate for actual applications that have been made, because the matter of whether a future application of E is objectively correct is only determinate once the application has been made and the community has deemed it correct or incorrect. This last clause highlights the important of *consideration* of the case for Wright. Note that 'actual application' here seems to have some impact on 'mentions' of the expression as well as 'use' of it: correctness-values are determinate precisely for cases of expression use that have been considered, even hypothetically. However, the account of indeterminacy of subsequent correctness applies in such cases as well, such that the application of an expression to a considered hypothetical context *c* may have a

different correctness-value from the same application if it is subsequently made in the corresponding actual context *c*. The fact that we have made the determining judgment of correctness-value of a case prior to its arising does not mean that our judgment as it arises will respect the agreed correctness-value; requirements of internal consistency in our usage do not arise in this way.¹² What all this highlights, in any case, is that merely possible applications, whose correctness-values are determinate according to M1, are *unconsidered* or *uninvestigated* applications.

McDowell's characterisation of Wright's view, which I have aimed to follow here, is a characterisation of a counter-intuitive and revisionary view—although we need to trace the arguments in order to be sure that it really is Wright's own (see Chapter 2 below). According to McDowell, Wright's view is totally unacceptable because it entails consequences that are simply *too* counter-intuitive, on the plausible assumption that there is a limit to how counter-intuitive a respectable philosophical thesis can be.

McDowell charges Wright with ultimately failing to preserve the normativity of meaning, hence meaning itself. He also argues that Wright fatally undermines a "familiar, intuitive notion of objectivity" (WFR 325), implying that it too must be preserved. Indeed, Wright joins McDowell in rejecting 'Problematic Realism' (cf. §1.2), a notion of objectivity that perhaps has some *prima facie* appeal when it takes the form of 'platonism' in mathematics. However, McDowell's claim is that Wright's approach to rejecting Problematic Realism jeopardises a more plausible intuitive idea that should be preserved:

The idea at risk is the idea of things being thus and so anyway, whether or not we choose to investigate the matter in question, and whatever the outcome of any such investigation. That idea requires the conception of how things could correctly be said to be anyway—whatever, if anything, we in fact go on to say about the matter; and this notion of correctness can only be the notion of how the pattern of application that we grasp, when we come to understand the concept [or expression] in question, extends, independently of the actual outcome of any investigation, to the relevant case. (WFR 325, my brackets)

It is one thing to deny that applications of an expression are correct by virtue of facts about the world as the Problematic Realist view conceives them, but it is quite another to deny that what determines correctness is in any way independent of the judgments

¹² It might seem that this view requires us the rejection of any kind of causal determinism that would allow (in principle) the prediction of what views would be held, as this would apparently make objective correctness-values determinate. However, the fact that the predictions would need to be interpreted means this is not the case (cf. my discussion of mechanical proof-checking in §2.1).

and attitudes of thinkers. The point McDowell wishes to preserve is the idea that an application has a correctness value independently of thinkers coming to investigate its correctness-value. The idea is that the world is a certain way independently of what people think and say about it, and accordingly that judgments about the world are correct or incorrect by virtue of the way the world is and not by virtue of the judgments and attitudes of thinkers. If 'this cup is red' is a correct deployment of the concept 'red' then that is because the cup is red, and not just because most speakers of English would agree that the cup is red.¹³ By contrast, Wright seems simply to identify the outcome of an investigation into how things are with how the community judges things to be at the time of investigation, such that 'our best judgment as to how things are' is just how things are.

The final sentence of the passage just cited shows how poorly suited the pattern idea is to expressing the notion of objectivity McDowell wants to preserve. How does a pattern 'extend independently' in this way? The thought seems to be that there is an independently existing 'potential pattern of application' that would be the right way to apply E to the situations S encounters and has yet to encounter, but that sounds rather like something metaphysical and thinker-transcendent in line with Problematic Realism.

All the idea actually amounts to is that correctness-values for all possible applications of an expression are determinate as per M1 (questions of vagueness aside). Thus, when we apply E to a new situation, the question of whether the application is correct is 'settled in advance', so to speak. The pattern imagery suggests that as we encounter new situations we trace a course of possible relatively correct actions that was laid out beforehand by our understanding, which is also a course of possible objectively correct actions in the case where our understanding is proper. But this is unnecessarily misleading: the idea is simply that relative correctness-values are assigned to all possible cases by an understanding, and that at least M1 and M2 are true.¹⁴ In 'pattern' language, this idea is the 'investigation-independence' of objectively correct patterns of application; what would be correct is a fact we can discover by investigation, which obtains independently of our doing so. A related notion is 'ratification-independence',

¹³ See the discussion of 'calling something yellow' at WFR 335, which is analysed later in the present paper (§2.4).

¹⁴ I assume that someone holding M1 and M2 would also want to hold M4 to make sense of fallibility, and would hold M3 unless aiming to define 'an understanding' more strongly as necessarily involving the right range of commitments. I think McDowell accepts all four, simply because his denying M2 would make it difficult to understand his gloss on Kripke's discussion of the 'sceptical paradox' (Cf. my chapter 3).

which is simply the same idea of objectivity as it relates specifically to independence from communal assent to a given judgment about the facts.

1.5 SUMMARY

McDowell writes:

...if the notion of investigation-independent patterns of application is to be discarded, then so is the idea that things are, at least sometimes, thus and so anyway, independently of our ratifying the judgement that that is how they are. (WFR 325)

His worry is that discarding this notion leads us to a kind of ‘idealism’, sacrificing the notion that a judgment to which everyone assents at a given time may be false. Notice that this is different from rejecting ‘Logical Independence’ (LI) from §1.2 above: we may all assent to a false judgment because we falsely believe it to be true, meaning that it did not in fact meet our ‘most refined criteria of acceptability’. But in rejecting both Problematic Realism and LI, Wright also denies that this kind of situation is possible: communal assent to a judgment is identified with its truth.

Imagine we want to account for the correctness of the sincere claim that a chair is blue, as characterised by a certain application of the expression “this chair is blue”. According to the kind of objectivity in question, accounting for the correctness of that claim involves checking that it tallies with the way the world is—whether it reflects reality. In terms of relativity and objectivity, we want to know whether the application of the expression is objectively correct—but with the additional caveat that objective correctness is more than just a matter of agreement among people, and is primarily determined by the way the world is and how correctness-values are assigned to interactions with the world by a proper understanding. If this particular application is a correct application of the expression “this chair is blue”, then that is not just because it is judged to be correct by everyone who matters, but because this chair is blue. Intuitively, it is possible for a chair to be blue even if no-one thinks it is. This is ultimately what Wright must deny in replacing M1 with W1.¹⁵

In summary, McDowell suggests that Wright offers a ‘reinterpretation’ of the pattern idea, which is itself a gloss on the minimal contractual picture that corresponds to the conjunction of principles C, M2, M3 and M4. The reinterpretation involves “discarding the thought that the patterns are independent of our ratification” (WFR 326). This

¹⁵ Note that other ‘anti-realist’ accounts of the contractual picture could be produced by substituting W1 for other principles. Wright’s is not the only option.

Rule-Following, Normativity & Objectivity

simply means discarding the idea that the objective correctness of a given application is a fact independent of communally agreed-upon judgment to that effect, which involves a denial of M1; the thesis that objective correctness-values for possible applications are determinate. Wright's intention is to undermine Problematic Realism, and with good cause. But as McDowell observes, the collateral damage of his efforts is the unacceptable loss of the idea that universally accepted judgments and truth are separable.

2 WRIGHT'S ANTI-REALISM

2.1 ANTI-REALISM & OBJECTIVITY

We have seen (§1.4) that Wright wishes to reinterpret the idea that understanding meaning commits us to patterns of application, which is a particular gloss on the minimal contractual picture whose essential components are principles C, M2, M3 & M4. Wright denies the realist principle M1 in favour of something like W1, which amounts to denying determinacy of objective correctness-value to merely possible applications of an expression while accepting that actual applications have a determinate objective correctness-value. As we saw, this is neatly summarised by the following:

There is in our understanding of a concept no rigid, advance determination of what is to count as its correct application. (W, p. 21)

Wright is not aiming to deny that, for example, “we may programme a machine effectively to check any putative proof” (W, p.20)—that a conclusion can be derived from certain premises does not seem to be ‘up to us’, so not indeterminate in the way that motivated the denial of M1, despite the supposition here that a judgment about correctness is indeterminate in objective correctness-value until it is made. Rather, Wright’s concern seems to be that the *interpretation* of mathematical rules and inference rules is not mechanically determinate in this way, and that this is really what matters in the case of human language users.

In principle, we could always run a program that would output the correctness-value of any possible application of an expression if we could specify the range R of correct ways of applying it. But Wright’s claim is that the range R itself is indeterminate for merely possible applications because objective standards may be different when they come to be actualised, since objective standards are flexible, *human* standards. So while in principle we can always specify a range and have something mechanically test whether an application falls within the range, that turns out to be irrelevant if, as Wright believes, we cannot guarantee that the range will apply to non-actual applications because we cannot guarantee that our judgments about the output of the machine at that time will coincide with our judgments now. If we accept this idea of flexibility, then it is quite possible that we will eventually interpret the machine as having made an error when by present standards its output would have been judged

veridical. In fact, Wright ultimately wants to challenge the idea of there being a fact about communal consistency across time.

In line with this, for Wright it is not the case that “rules of language ... determine how such-and-such circumstances may correctly be described” (W, p.20). If anything, it is *our interpretation* of the rules of language that would determine correctness here, and I suspect Wright might even want to deny any substance to the distinction here between ‘the rules’ and ‘our interpretation of the rules’, reducing the former to the latter.

What are the arguments behind this view? As we know, McDowell sees Wright as aiming to undermine a strong realist notion of objectivity, something like that involved in Problematic Realism, but accidentally making impossible any intuitively plausible notion of objectivity. How does Wright characterise this target himself, though? He writes:

to think ... of the shape of some unobserved object as determinate, irrespective of whether or not we ever inspect it, is to accept that there are facts about how we will, or would, assess its shape if we did so correctly, in accordance with the meaning of the expressions in our vocabulary of shapes. (W, p.216)

On this view, square things count as square whether anyone thinks they do or not, because the meaning of the predicate ‘is square’ is fixed and so is the range of ways in which it is objectively correctly applied. Though the meaning of an expression may change over time, this will be met with a corresponding change in the ranges of ways it can be objectively correctly used. In this sense, the meaning is what determines the range of objectively correct applications and not the expression.

According to Wright,

This idea leads us to look on grasp of the meaning of a shape predicate as grasp of a pattern of application, conformity to which requires certain determinate verdicts in so far unconsidered cases. In that sense, the pattern extends of itself to cases which we have yet to confront. (W, p.216)

This is the realist notion he hopes to undermine, although the idea of a pattern that ‘extends of itself’ is puzzling. This image is simply meant to convey the denial of a particular type of determinacy, one where what we judge to be correct in the future is answerable to what *is* correct independently of that judgment—a ‘ratification-independent’ fact of correctness. Hopefully it is plain from this that all Wright is saying is that M1 is false. That is noteworthy, because the realist contractual picture we get

from accepting C+M1-4 is apparently far weaker than the sort of 'transcendent realism' that Wright wants to target in the following passage:

If there can be such a thing as a ratification-independent fact about whether an expression is used on a particular occasion in the same way as it has been used on previous occasions, it ought to be something which we can recognise—or at least be justified in claiming to obtain. Otherwise the correct employment of language will become on every occasion radically transcendent of human consciousness. (W, p.218)

Understood as a general claim about what kind of theories we should accept, this presents us with an indication of the anti-realist motivation for rejecting a particular 'transcendent' kind of realist objectivity. The grounds seem to be that if a fact is 'radically transcendent of human consciousness' then it is not accessible to humans, so we might doubt that such 'transcendent facts' are the facts we normally deal with, undermining any theories that do have such notions of the facts we normally deal with. I suggest that this kind of realist's contractual picture would involve adding Logical Independence to C + M1-4, so Wright's intended target seems to be more than just M1. If he does end up rejecting every account of objectivity that accepts M1, as McDowell has suggested, then the charges of overkill might be entirely justified. It seems that the Problematic Realist's version of the contractual picture would accept LI+C+M1-4, though strictly speaking we would have to add some principle that explained LI in terms of the rejection of human nature's involvement in the making of a genuinely true judgment. Acknowledging this caveat, we can safely identify the conjunction of LI+C+M1-4 with Problematic Realism.

In any case, Wright clearly needs to demonstrate that the possibility of any of our judgments about correctness changing over time, if indeed it is a general possibility, constitutes a reason to reject M1, or else provide further reasons for doing so. However, the 'implosion' objection I raise below (§§2.2 & 2.3) offers a reason to doubt the possibility that our judgements do change over time, particularly from Wright's perspective.

It is worth spending a little time to understand how 'rule-following' fits into this discussion. Imagine someone tries to continue the pattern '2, 4, 6, 8, 10'. There is an algebraic *rule* that generates this pattern, though the term 'n+2' that expresses this rule does not figure anywhere in the pattern itself. We might see a pattern emerging in the series but be unable to say what exactly the pattern is, or what its underlying rule is, or predict what the next term will be. Moreover, someone who predicts that it will be 12 might not have done so for the correct reason—for example, if they simply guessed. To

'grasp' this pattern is, so it seems, to be able to continue the series indefinitely with some degree of reliability and for the right reasons, i.e. because the rule that generates the series is 'n+2'. For any n, someone who has grasped the pattern should be able to predict the next n in this pattern whether they have ever performed that calculation previously or not. Furthermore, for any n, someone who has grasped the pattern should be able to tell whether any number presented to them is the next n or not. He or she is in a position to know whether any new case 'fits' the pattern. It is in this sense that the notion of 'grasping a pattern of application' relates to the idea of 'understanding a rule'.

A possible gloss on the 'pattern' version of C+M1-4 realism might be that the objective correctness-value of any possible application of (e.g.) the shape predicate 'is square' is determined by the rule that generates the objectively correct pattern of application for that predicate, in conjunction with facts about which objects are square and the relation between the rule and those facts. To adapt this for Problematic Realism (LI+C+M1-4) we would have to add that the conjunction of the rule, the facts, and the relation between them determines objective-correctness-values in a way that entirely excludes the involvement of human beings, by virtue of the experience-transcendent fact that certain things are rectangular and that the predicate attaches exclusively to those things.

2.2 COMMUNALLY ESTABLISHED CORRECTNESS

Wright's reinterpretation of the pattern idea, C+W1+M2-4, is effectively characterised by the following passage:

...an understanding of φ may be harmlessly, if unilluminatingly, described as grasp of a certain general pattern of use, provided that this grasp is then reductively construed as the ability to participate in an on-going community of assent in the use of φ . To grasp the pattern of use for φ which one's training aimed to illustrate and explain is just to come to understand φ ; and the knowledge of use in which that understanding consists is manifested as, and so goes no further than, the ability to participate in the linguistic community as far as the use of φ is concerned. It is an error to conceive ... of understanding φ as explanatory of the latter ability, to think of shared grasp of a pattern as underlying and issuing in our propensity for consensus in the use of φ .
(W, p.227)

A word on 'reductive construal' is appropriate here. Wright apparently intends for the 'grasp of a pattern' in question to be conceived in a particular way. Having a 'grasp' of

this kind is not to 'grasp a ratification-independent pattern of use', but rather to have the ability to participate in a community of assent. Apparently, the attribution of a 'grasp of a pattern' is not correct unless it can be further analysed as the attribution of such an ability. In this sense, 'having a grasp' can be *reduced* to 'having a certain ability'—so although we can accept talk of grasping patterns, this should be taken as standing for some other idea.

A rough minimalistic notion of reducibility might be that X is reducible to Y just in case X can be fully analysed, or perhaps fully explained, in terms of Y. In this minimal sense, water is reducible to H₂O. However, Wright seems to be trying to say more, that reducibility here is important for factual accuracy. The claim seem to be that it would be erroneous to accept it as a fact that one grasps a pattern while denying the reducibility of this to one's ability to participate to a community in question; the original judgment that one grasps a pattern would apparently then be false on this view. Call this idea '*explanatory reducibility*'. Crucially, it is also being denied that grasp of a pattern explains that ability, and that at most the envisioned 'grasp' is a complex description of that same ability; the two phenomena are identical. In similar terms, Wright might say that language practices are reducible to certain patterns of behaviour and psychology, in order to deny that linguistic competence *explains* these patterns rather than being identified with them. Call this an '*explanation denial*'.

A related but distinct idea is that the reducibility of X to Y entails that the X is somehow 'less genuinely real' than Y. The idea is that a greater degree of accuracy would require us to talk about X in terms of Y, about X as a complex of Y, or even just about Y. This understanding of reduction is puzzling. We seem to be moving from the analysability of X in terms of Y to the failure of X to genuinely exist—but it is hard to imagine many situations where this inference would be obviously sanctioned. It is sometimes argued that moral properties are less genuinely real than non-moral properties if they can be analysed in terms of non-moral properties, but that inference is not obviously licenced without further argument.¹⁶ We might say that an animation is not accurately described as something that moves but as a successive display of many still images, but this does not immediately licence the conclusion that the animation exists in some lesser way than the successive display of still images with which it is to be identified—nor indeed, that it is 'less true' that the animation moves than that the series of images are displayed in succession. However, the inference from the fact that something is

¹⁶ For example, Blackburn (1981) argues this in defending of his 'quasi-realism' against McDowell (1981).

reducible to other kinds of things to its being *less-than-genuinely real* is a surprisingly common one in discussions of objectivity and mind-dependence, as revealingly critiqued by Gideon Rosen (1994). Call this slightly dubious kind of reduction an '*ontological reduction*'.

For Wright, the idea that my understanding the meaning of an expression commits me to a limited range of use merely tells us that I understand the meaning of an expression just in case I can participate in a linguistic community that accepts only certain applications as correct. The explanation denial says that my behaving in certain ways relative to the linguistic community *constitutes* my understanding the correct pattern of use, rather than being explained by it as a realist would suppose, because the understanding in question is reducible to an ability to participate. The normativist's intuition that in understanding the meaning of an expression we share common normative commitments to a pattern of usage, while assumed by Wright to be true, is not expressed here in a way that reflects the facts of the matter. To the extent that commitments are involved, these amount to the fact that ceasing to use an expression in the same way as our peers results in the cessation of our proper understanding.

From what we have seen of Wright's views on worldly features that extend beyond experience, it should be apparent that he reserves a certain "hostility towards any notion of truth for certain statements dissociated from criteria for their verification" (W, p.28). As part of the general anti-realist programme, his aim is to offer an *ontological* reduction of intuitively realist concepts (e.g. the existence of unverifiable phenomena) that cannot be 'manifested' in terms of things deemed to be immediately accessible to us, to concepts that can. The things deemed immediately accessible exclude objects conceptualised in a certain way, on the understanding that conceptualisation requires linguistic competence, and we aim to explain all linguistic competence eventually—so a full explanation cannot rely on that. This means that the basic components of meaningful activities are taken to be things like the particular patterns of behaviour exhibited by linguistically competent human beings, the sorts of thing that can be taken as 'natural' phenomena, having no meaning in themselves. The anti-realist 'manifestation challenge' Wright formulates (1980b, p.112) is a twofold condition on when we should accept characteristically 'realist' concepts as corresponding to genuinely real features of reality:

- Manifestation Challenge**
- 1) The realist should explain how, when our training is necessarily restricted to confrontation with experienceable situations, we are supposed to be able to form a conception of what it is for an experience-transcendent situation to obtain.
 - 2) The realist should explain what evidence there is that we actually possess any such conception—what in our use of any particular statement would distinctively manifest that we understood what it was for that statement to be true in a manner transcending our capacities for verification.

The most striking question about this, apart from what some of its more imposing jargon means, is why a realist should accept the challenge thus formulated. Perhaps surprisingly, this is not on McDowell's agenda; as I outlined (§1.4), his approach in WFR is to reject anti-realism on the basis of the unappealing consequences it entails. Still, McDowell does go on to argue (WFR §§10-11) that the unappealing consequences are traceable to the notion of 'manifestation' at work in anti-realism, and he connects this to assumptions that he thinks we should reject. I return to this in Chapter 4.

As McDowell puts it,

...the behaviour that counts as manifesting understanding to others must be characterizable...without benefit of a command of the language in question. (WFR346)

The challenge is supposed to be met without appealing to the realist concepts under examination in the explanation. Interestingly, that seems to adopt the assumption that those concepts do not play such an essential role in our language that we would be unable to characterise them in this way. In effect, it is ultimately this kind of assumption that McDowell targets—I return to the idea in Chapter 4.

The anti-realist's supposition is apparently that for a large class of realist notions, including M1, the manifestation challenge cannot be met. A possible application, understood as one which has not been considered, seems to be beyond our experience—apparently by virtue of the stipulation that it is 'unconsidered'. But the terms here are very slippery, and it might well be that the realist can make room for the idea of 'experienceable' unconsidered cases, or account for how we 'directly manifest' a grasp of a ratification-independent fact. I suspect that if the manifestation challenge is offered on the supposition that it cannot be met, it is simply because the realist account imagined is something like Problematic Realism. Indeed, Wright's

thought that realism is *characterised* by LI reveals a tendency to assume that realist objectivity must involve verification-transcendence, which I have suggested is not necessarily the case. That M1 can be characterised in a way that meets this kind of challenge seems perfectly plausible.

It is undoubtedly an associated requirement of verifiability for genuine facts, whatever 'facts' amount to ontologically, that allows Wright to make most of the important claims of his argument. The requirement allows Wright to specify that if someone is trying to follow a ratification-independent pattern of application he will only count as having done so if he can justify the claim that he really has followed it. If it can be argued that no-one could ever provide that kind of justification, then it would follow that adherence to ratification-independent patterns is not possible. In our terms, the challenge is first and foremost to justify the realist's claim that any given application of an expression E is objectively correct by virtue of the objective correctness-value according to M1. However, 'ratification-independence' is something that applies also to a realist notion of the range of correct ways of using E specified by any particular understanding, as this too reaches to merely possible applications. So there is a second challenge here, to show that any particular application of an expression E is correct relative to the determinate range of correct applications specified by a given understanding of the meaning of E (whether proper or faulty).¹⁷

Wright's argument takes the form of a *reductio* of the claim that following a rule for the use of an expression is adhering to a determinate pattern of application which 'extends of itself' to new cases. It begins with the idea that in teaching someone how to use an expression objectively correctly, I can give him instructions for its use which he initially seems to understand but to which, from my perspective, he subsequently fails to adhere in new cases. If his applications are correct relative to his understanding of the meaning of the expression, then by M3 this would be a faulty understanding. Since having a faulty understanding involves undertaking to restrict one's usage to a slightly different range from that which is objectively correct, there seems to be a sense in which he would not simply have failed to follow the instructions I gave, but rather, have followed *different* instructions from those issued. Wright will ultimately aim to show here that the only way we know whose understanding is faulty is by referring to the fact that I was teaching and the student was learning, such that if we were both

¹⁷ In fact, this last is one way of interpreting the sceptic's challenge in K. See §4.1 below.

competent users of that expression we could not know for sure whose understanding was faulty in a case of divergence like this.¹⁸

Imagine the rule of application for the predicate 'is unwell' corresponds in a simple way to the semantic rule

(G) 'S is unwell' \leftrightarrow S is unwell

However, our student's deviant rule of application corresponds to the following 'gerrymandered' semantic rule, like that of Goodman's 'grue' property:¹⁹

(G*) 'S is unwell' \leftrightarrow S is unwell between t and t₊₆

Now clearly someone who is unwell (assuming G is true) will be identified as such by the student if he follows G*, but only within a certain time frame; if S is unwell after t₊₆ then the predicate does not apply to S according to G*. Of course, from the instructor's perspective the student's divergence at t₊₆ from the instructions given will seem totally mysterious so long as it is assumed that they were originally understood by the student as amounting to something like G and not G*.

The problem is that the student could quite conceivably justify his interpretation of the instructions given with an appropriate set of 'rules for interpreting instructions' (in a sense, second-order rules).

That is, if somebody suddenly makes what seems to be a deviant application of the rule, there may be an interpretation of the instructions which we gave him which both explains his use of the terms in question which coincided up until now with the use which we intended, and also the further deviant use. (W, p.23)

Let's assume, for illustrative purposes, that the instructor specifies that 'S is unwell' materially implies that 'S is not well'. Perhaps the student has similarly mislearned how to use the predicate 'is not well', so that the semantic rule his usage reflects is that 'S is not well' iff S is not well between t and t₊₆, resulting in the error in his understanding of how to use 'is unwell'.

However, the problem could lie, as Wright imagines, with some more fundamental aspect of his ability to follow instructions. If the instructions can only convey their intended meaning under the correct interpretation, then it must be possible for the

¹⁸ Cf. W, p. 35.

¹⁹ For 'grue' see Goodman 1983.

instructor to guide his student to the correct interpretation by corrections, explanations, demonstrations and so on. But, we suppose, these too can only convey their intended or proper meaning if correctly interpreted. Unless we are willing to posit some level at which the interpretations of student and instructor *must* coincide, the ensuing regress leaves it up to the student simply to make whichever interpretation of the instructions as seems best to him, and he is likewise at liberty in how to understand any attempts on the part of the instructor to correct him, as no watertight justification for choosing one interpretation *over* another can be offered.²⁰

In the face of this possibility it is uncertain how we arrive at a common understanding of language. Beyond this, it is also unclear how we are able to use words consistently in the same way from one case to the next; any particular interpretation of the instructions for use of 'is unwell' is also subject to interpretation. Imagine that, when our student hears the instructor saying things like, "Only describe someone as 'unwell' if they are not feeling well," he says to himself, "'S is unwell' iff S is unwell between t and t₊₆". We might at least hope that this will result in him consistently following G* when he meant to follow G, but if the interpretation is just some further expression that supposedly conveys the meaning of the original then repeating "'S is unwell' iff S is unwell between t and t₊₆" like a mantra will not guarantee that the student even obeys the semantic rule G* on his most sincere efforts. If expressions in the language are always subject to interpretation in this way then something the student says to himself is just as prone to ambiguity as the words of his instructor.

In the face of this difficulty we naturally appeal, Wright claims, to the idea that the learning process is some kind of 'cottoning on', "that is, a leap, an inspired guess at the pattern of application which the instructor is trying to get across" (W, p.216). The idea is that when I try to learn the use of an expression it seems 'as if' I form hypotheses about how it should subsequently be used, one of which I ultimately come to believe reflects a correct interpretation of the application rule employed by my instructor. The instructor himself cannot guarantee that I will form the correct 'hypothesis' rather than some subtly incorrect variant but, so we think, I will somehow arrive at it eventually.

Wright does not mean that I actually *do* form a hypothesis in the common language, for in the end, if such a hypothesis took a linguistic form it would be subject to an

²⁰ The idea of a point at which the interpretations 'must coincide' is what McDowell calls a 'super-rigid machine mythology'. See WFR §4 and also my analysis thereof (§3.2).

interpretive regress as before. Rather, it is merely ‘as if’ I form a hypothesis: the envisaged leap is “something idiolectic” (WFR 327).

On this picture, then, “each of us knows of an idiolectic pattern of use, for which there is a strong presumption, when sufficient evidence has accumulated, that it is shared communally” (W, p.217). I myself can be certain of what I understand the rule to mean, so I can at least act in accordance with my own idiolectic version of it on the assumption that it reflects the rule my instructor intended. Confident in my ability to maintain a pattern of relatively correct applications relative to my own idiolectic understanding, I eventually settle on the idea that my idiolectic understanding is a *proper* understanding; that the relative correctness of my applications entails their objective correctness.

However, so the argument goes, even adhering to this kind of private pattern is problematic. If someone thinks to himself, “I want to call *that* sort of thing ϕ ,” where ‘*that* sort of thing’ means an idiolectic classification of objects, then we assume that he imposes constraints on himself to call ϕ only ‘*that* sort of thing’. The trouble is, whenever he sincerely attempts to apply this rule in new cases he cannot be sure that something which he is convinced to be of ‘*that* sort’ really does fall under his earlier specification. With respect to what we presume is a determinate pattern of application to which he commits himself, he apparently cannot be certain that what *seems* to him to be obviously a correct application, in accordance with that pattern, will really be one. He cannot appeal to anyone else to check whether he is following his idiolectic rule, as by its idiolectic nature he is unable to effectively communicate what he means by ‘*that* sort of thing’; it has a meaning only for him.²¹

Wright later observes that “there is no scope for a distinction here between the fact of an application’s seeming to me to conform with the way in which I understand it and the fact of its really doing so” (W, p.355). It is important that we try to get clear on this idea. So first, imagine that I set up one of these idiolectic rules by thinking to myself “I want to call *that* sort of thing Q”. Questions of vagueness aside, this rule sorts all particulars into two sets: ‘*that* sort of thing’ and ‘not *that* sort of thing’. In fact, we should be able to derive the following semantic rule:

(H) $\forall x (x \text{ is } Q \leftrightarrow x \text{ is } \textit{that} \text{ sort of thing})$

²¹ Cf. W, pp.216-17

If H follows from my idiolectic rule, then for every attribution I make of the property Q to a particular there will be a determinate verdict as to whether that particular really is Q according to the dictates of my idiolectic rule. The extension of the predicate 'is *that* sort of thing' is firmly established by H.

In order to reject M1 as we revised it—i.e. as also involving determinacy in *relative* correctness-value for possible applications—Wright aims to show that any determinate verdict generated by H is irrelevant here, on the basis that the notion of such a verdict can be rejected as failing the manifestation challenge. Now, the best I can do in applying Q to new cases is to try my hardest to be accurate in any judgment that some x is *that* sort of thing, but a consequence of the notion of 'idiolectic classification' at work here is that inaccuracy in such a case is not measurable. Here, if it seems to me conclusive that ' α is *that* sort of thing' then, having eliminated *ex hypothesi* the possibility of anyone other than I checking that α really is *that* sort of thing, the question of accuracy is answered entirely by whatever *my* most sincere judgment is on the matter. The actual truth value of that statement becomes an irrelevant theoretical artefact, having no detectable consequences of any kind for my judging practice.

Nevertheless, since the notion of an idiolectic rule precludes any scope for a distinction between relative correctness and mere *seeming* relative correctness in its application, it seems plain that such a rule could not really guide someone's actions, and would thus lack any normative influence. Yet it might seem that if others had access to my idiolectic rule and could compare my Q attributions to the class of things designated Q by my rule, it might bring some kind of substance to the idea of my successfully following the rule, and so to the idea of its being normative. For as Wright notes,

The only circumstance in which it makes sense to think of someone as correctly applying a non-standard understanding of, for example, 'square' is where there is a community of assent about how, given that that is what he means, he ought to characterise this object. (W, p.217)

We might paraphrase this as: to say that someone has a non-standard understanding of 'square' means that there is a recognised *proper* understanding of 'square' which differs from his. It seems to be important to Wright that whoever makes final assessment of my performance can draw the distinction between a sincere effort and success. Harking back to his assumption about the importance of a ratification-independent fact being *recognisably* true, it seems that here a language community is able to provide a less biased standard than I could provide for myself, as what may seem a correct application to me may not seem so to anyone else.

However, if we have to appeal to communal assent in this way then we have to modify our conception of understanding. Rather than acting in accordance with an idiolectic pattern of use, properly understanding an expression then becomes acting in accordance with the communally agreed rule; what is objectively correct is determined by the community. If people can agree on how an expression should be applied in a new case and my own usage conflicts with this (produces relatively correct applications relative to my understanding which are objectively incorrect), then I have failed to understand the rule and so failed to adhere to the pattern respected by the community.

Wright's objection now is that, as McDowell puts it, there seems to be a "precise parallel between the community's supposed grasp of the patterns that it has communally committed itself to and the individual's supposed grasp of his idiolectic commitments" (WFR 328). Wright asks us (W, p.218) to imagine a new case where the community agrees that my actions do accord with the communal rule—that they are objectively correct, because correct relative to the communal understanding, which we assume to be *proper*. Here it seems that the same 'privacy' problem arises again, as communal assent does not tell us that my actions really *did* accord with that rule, only that the community sincerely believes that it did—the community is just as blind to correctness-values relativized to its understanding.

It seems at this point that we have run out of options. If we imagine the community to encompass every human agent, then we have exhausted the pool of arbiters who could recognise any deviations from our communal understanding, and yet M1 entails that the whole community could still be incorrect in its judgments relative to its understanding; we may all agree that a new case fits the pattern, and yet be wrong, while having no idea that we are wrong.

To put the point in the starkest possible way: how can we penetrate behind our consensus verdict about a particular decidable question to achieve a comparison of our verdict with the putative investigation-independent fact? (W, p.220)

The answer Wright gives is simply that we cannot. I suppose the argument for this is that if judgment is separable from the truth, and every instance of knowing the truth is a judgment separable from the truth, then there is always the possibility that our best judgments about truth are false. On the reasonable assumption that we successfully ascribe truth to things, there then might seem to be something wrong with the notion of correct judgments about truth depending on verification-transcendent facts of this

kind: either a) that it requires us to make an unjustifiable claim to be in line with such facts whenever we make true judgments, or posit some mysterious faculty of intuition that provides such justification, or b) that it forces us to address the sceptical possibility that the truth of our judgments is ultimately not knowable. I think a) is more of a problem here than b); the fact that this notion permits scepticism is not a problem unless it can also be shown that the scepticism is intractable. However, for Wright b) is apparently the priority, and I presume that the explanation is based on some methodological claim to the effect that we should prefer anti-realism because it prevents these issues arising. However, holding that kind of principle may prove dangerous for Wright if McDowell can substantiate the claim that anti-realism leaves no room for objectivity of a recognisable kind, for there would then be a case for preferring a kind of realism where objectivity is adequately accounted for. Wright's argument should amount to more than this, and in effect the extra hidden premise is (again) that investigation-independent facts fail the manifestation challenge.

If ratification-independent patterns cannot be what actually determine the objective correctness-values of our applications for the same reason, then the normativity of meaning cannot be explained by a commitment we undertake to accord with such patterns. This is where the anti-realist alternative steps in:

What does it add to describe the situation in two-fold terms, of the fact of conformity to the pattern and the community's recognition of that fact, rather than simply saying that there is communal agreement about the case? (W, p.219)

If we fail to find a use for pre-ordained patterns to which our actions should accord, the anti-realist concludes that the notion is misleading and must be explained away. It is proposed that the contractual picture can only amount to the fact that non-normative and apparently contingent facts about human beings 'commit' us to certain patterns of behaviour. Understanding an expression's meaning commits us to using it in certain ways in the future only in the sense that we only count as understanding it if our use satisfies the communal consensus, and we are driven to maintain our understanding. Objective correctness is denied as explaining whatever the communally accepted understanding dictates, being reducible to it, and whatever the communally accepted understanding dictates is in turn reduced to whatever the community agrees that it dictates. Range R (of M2-4) thus contains whatever the community agrees it contains. Our distinction between objective and relative correctness becomes unhelpful at this point, as any legitimate kind of correctness is reducible to communal consensus, being

denied any explanatory role: correctness relative to personal understanding is denied legitimacy here; correctness relative to the communal understanding is reducible to communal consensus on correctness; and objective correctness is also reducible to communal consensus on correctness. However, it is important to note that the reductions described here are not clearly of the dubious *ontological* variety. Reducing objectivity to communal consensus does not immediately licence the conclusion that objectivity is not something genuine—I presume that Wright does not intend it to, and his only additional commitment seems to be that objectivity does not *explain* the consensus. This will have an impact on my assessment of McDowell’s ‘Illusion Argument’ in the next section.

It should now be clear what McDowell means when he describes this account as a ‘reinterpretation’ of the contractual picture—the “allegiance to conceptual commitments” he thinks is an essential part of the contract intuition becomes ‘allegiance to communal verdicts’.²²

2.3 SOME PROBLEMS FOR WRIGHT’S ANTI-REALISM

I will take this opportunity to outline some things that McDowell finds objectionable about this picture. I will then try to explain some of the counter-intuitive consequences that seem to follow from Wright’s account, and begin to tie those consequences to the general line of attack McDowell offers in WFR.

Wright’s appeal to a linguistic community is supposed to explain how we arrive at the idea that some applications of an expression are objectively correct or incorrect, as it provides some kind of authority to which each person’s usage must conform whatever their own judgments of correctness might be. Normativity is supposed to enter the picture at the level of communal authority, a consequence of the fact that if one falls behind the way that language usage evolves one ceases to mean what one intends with one’s words. Introducing the possibility of an individual’s error relative to this community introduces the possibility of his actions being guided, and so for the normativity of meaning.

The main trouble with this picture of normativity is that it entails the falsehood of the intuitive claim that the objectivity of a fact means that it is not merely agreed upon by us, due to the operative explanation denial. But can we rightly call communal consensus a kind of objectivity at all? That would seem to involve totally revising

²² For ‘allegiance’, see WFR 327, 352.

everything we think about objectivity, particularly regarding the objectivity of facts about reality, and it seems likely that the end product would be unrecognisable. What about keeping the old notion but accepting the truth of the proposed analysis? So long as Wright presents communal agreement as the real truth behind our intuitive notion of objectivity it is unclear how the two ideas could both be held consistently. The view does not obviously make room for anything less than a complete denial of the intuitive notion, and the 'objectivity' Wright does make room for is unsatisfying because it excludes the possibility of our being jointly mistaken in applying an expression like 'That box is square'—that is, in our judgment as to its truth-value when applied to a particular box.

McDowell seems to think that this is a problem for Wright in itself, and perhaps the view really is too counter-intuitive. However, I think Wright would just deny that sacrificing the intuitive kind of objectivity is a weakness. He explicitly considers the case of applying shape predicates with objective correctness and still comes to this conclusion that the relevant kind of objectivity consists in communal consensus, so is clearly unfazed by its impact on our intuitions about it. McDowell's criticisms will only make an impact if he can show that rejecting the intuitive notion of objectivity has *further* consequences that Wright cannot accept. Indeed, he attempts to show that the rejection entails the non-normativity of meaning, which would be a showstopper for Wright if demonstrable.

McDowell's claim is that the normativity of meaning seems to become a kind of sham, properly analysed as non-normative dispositions and propensities to satisfy the majority view, associate sound patterns with patterns of behaviour, feel constrained by certain scenarios etc. If Wright denies that normativity explains these tendencies we identify with meaningful language use, then it is tempting to think that he could not explain courses of action by appeal to action-guiding prescriptions. The normativist's desired contrast in bringing normative considerations into language use is between dispositions to act in a certain way and prescriptions to act in a certain way; there are supposed to be ways we *should* use an expression according to the range of objectively correct ways of using it, which explain why we *do* use it in such ways when we count as using it correctly. But Wright seems simply to deny that there is an explanation here. For him, these tendencies are not explained by the idea that I should do certain things—things like 'feelings of constraint' shared among a community are all that the normativity of meaning amounts to on a more thorough analysis. That seems a merely dispositional explanation of why some things are counted by us as correct.

The reinterpretation of the 'commitments' of the contract picture in terms of my psychological or behavioural dispositions to keep pace with the attitudes of my peers might seem to suggest that the idea that we are subject to commitments is really an illusion. McDowell claims that if we accept Wright's picture, we will 'struggle to distinguish it' from a similar picture where

...the possibility of going out of step with our fellows gives us the *illusion* of being subject to norms, and consequently the *illusion* of entertaining and expressing meanings (WFR 336).

McDowell uses this kind of 'conjectural' argument several times in WFR. The conjecture that we cannot keep Wright's picture of consensus-based objective correctness, as constructed out of non-normative facts, *separate* from this one is used as part of McDowell's important '*Illusion Argument*':

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

The claim seems to be that we cannot accept a notion of objective correctness-values alongside Wright's reduction of that notion to communal-consensus, so we cannot accept that the one is reducible to the other. This psychological claim is perhaps true, but that alone would not make Wright's account false. What McDowell seems to think is that the removal of any explanatory role for norms results in our coming to conceive of their putative action-guiding properties as less-than-genuinely existing. This would be an *ontological* reduction of the normative to the non-normative. The Illusion Argument seems to say that this manoeuvre is unavoidable if we accept Wright's account. That is rather odd, however: why should we make that inference, and if we do, why could it not be disputed?

What McDowell apparently wants to say is that Wright's reinterpretation of the contractual picture in terms of behavioural propensities means that it amounts to giving no more than thick descriptions of a fundamentally non-normative picture, and that this is the *genuine* picture of the world. So to say that the activity of understanding genuinely involves commitments would be false: the activity genuinely involves only attitudes and dispositions, facts about what we do and not what we *should* do. Wright may have explained why we *think* that we are subject to commitments, but the upshot

is that we are deluded; our actions are in no way genuinely guided, and there is nothing that we genuinely 'should' do.

However, Wright is not obviously committed to this ontological reduction—the entire Illusion Argument rests on McDowell's conjecture that we will be unable to resist that move ourselves, which is apparently not the same as saying that Wright's argument entails it. Furthermore, aside from doing extreme violence to our intuitions, which might plausibly motivate a rejection on methodological grounds, it is unclear why the truth of this conjecture would indicate that Wright is mistaken.

To substantiate the Illusion Argument, McDowell would apparently have to show that Wright's argument that his anti-realism offers a more accurate analysis of what we call norm-governed language practice does actually entail the ontological reduction. This is bad news for McDowell, as the Illusion Argument is essentially the basis for his conclusion that "a condition for the possibility of finding real application for the notion of meaning at all is that we reject 'anti-realism'" (WFR 347). The onus is very much on McDowell to demonstrate that the ontological reduction is unavoidable.

The next question, though, is whether the Illusion Argument, if made sound, would actually present a threat to Wright. After all, normativity still has a place in his picture just as much as objectivity and commitments do, having been subject to a degree of analytical reinterpretation. Wright could just say that everything we mean by 'normativity' is accounted for in this picture, and that the reason it's not an illusion is the same reason that communal consensus on correctness is not an illusion of correctness: the contrast between what seems like normativity to the community and what *is* normativity does not arise, because the ratification-independent fact of which it is fails the manifestation challenge. The Illusion Argument can be defeated from within the anti-realist model.

This seems an effective response so long as we grant that the fact of illusion would be a fact of the Logically Independent (transcendent) kind, though as I argued (§1.2), we need not grant that there are any such facts.

Wright can use this kind of argument form to respond to other realist worries about anti-realism too. So for example, one thing that might concern us is that adherence to the fixed commitments of an understanding is made irrelevant by anti-realism, because if *S* commits to some *fixed* range of ways of applying *E*, the fact that the range is fixed will render *S*'s understanding faulty if the communal consensus on the correct range

changes, whether S's commitments were originally shared communally or not. The complaint is that consistency gets you nothing here, when intuitively if you understood properly before and you maintain that understanding then you should still understand properly. Wright's answer to this would be simple: the problem supposes that there is some fact of your consistency of usage beyond the communal judgment of consistency, when there is not. Presumably if S is out of line with the community understanding when he formerly agreed with it, then the community will judge that he did *not* maintain his understanding as the argument supposes, and that is the fact of the matter. So consistency is rewarded, but there is no fact of consistency beyond the community judgment.

I will call this argument form a '*consensus override*'. Provided we do not undermine the premises of Wright's account and accept the coherence of the picture, the consensus override allows him to dodge any worries that intuitive realist notions are absent from his picture on the basis that if they intuitively obtain, then the communal consensus will be that they obtain, and so they *will* obtain. It can be summarised by the following inference-rule:

(CO) (It is the majority view that P) \leftrightarrow P

The inference from majority view to truth is permitted so long as Wright's contractual picture is assumed true. This allows Wright to answer any worries about preserving McDowell's 'familiar, intuitive notion of objectivity'. That the notion is intuitive means that its apparent insolubility with Wright's picture is subject to consensus override, such that its intuitive status entails its truth. So McDowell's objection appears unfounded.

That there is something very wrong here should be apparent. If we have correctly established that normal intuitive objectivity is incompatible with Wright's view, how on Earth could accepting Wright's view dissolve the contradiction? The answer, of course, is that it doesn't. In fact, the derivability of CO from Wright's commitments causes his account to collapse from the inside out. I call the problem '*Implosion*'.

The problem is that, to the extent that anti-realism is revisionary and counter-intuitive, the majority of people do not presently believe it is true. Accordingly, if CO is true *ex hypothesi*, then we can derive the falsehood of anti-realism by CO. We cannot dodge this by stipulating that the truth of anti-realism transcends our consensus verdict, as that would be inconsistent with CO. So if CO is derivable from Wright's commitments,

then his position is incoherent unless he can secure communal agreement that anti-realism is true.

In fact, I suggest that CO is the source of the implosion to begin with, because it is precisely the derivability of facts from consensus and vice versa that is so counter-intuitive. Unless public opinion changes, CO will always entail its own denial, so Wright's best hope is to revise his account to prevent the licencing of inferences from communal consensus to truth. However, he may be hard pressed to achieve this without totally destroying the reinterpreted contractual picture he has presented. The best option would be to develop the picture to include some account of our acceptability criteria for judgments, and licence inferences from communal judgments to facts only for those judgments that satisfy 'our most refined criteria of acceptability'. There may then be scope for arguing that such criteria would exclude the communal consensus that anti-realism is false (or whatever the precise threat may be) as indicative of truth. Perhaps we all agree on something but would prefer to know more before genuinely assenting to its truth. It would then become an open question whether our best deliberations would lead us to anti-realist conclusions, and Wright's task would be to show that they should do.

I suspect this is the kind of path Wright would want to take here, but I'm not sure it will work. If the community sincerely judges that anti-realism is false, it seems to follow that the most refined criteria of acceptability have already been met. Stipulating new criteria that *should* have been met for any such judgment will put us in a minority position, apparently revising things that the community already accepts. So once again, the only thing that can save Wright's anti-realism is a favourable change in communal consensus. If McDowell can substantiate his claim that Wright's position exceeds some limit of acceptable counter-intuitiveness, then there will be little scope for preventing the implosion of anti-realism.

2.4 THE BASIC LEVEL

In WFR §5, McDowell investigates what he calls the ‘basic level’ of Wright’s account. If Wright presents meaning as explanatorily reducible to facts about our propensities to agree that certain patterns of application are better than others, the facts about meaning must be describable in terms of those propensities. In fact, the anti-realist ‘manifestability criterion’ means that we must be able to characterise our language practices in terms of what is directly manifest to someone in behaviour, where a manifestation is an observable phenomenon whose intelligibility does not presuppose a grasp of the language in question.²³ I suppose the kind of thing that would fit this specification might be the kinds of thing a new-born baby seems able to recognise—aggressive behaviour, passive behaviour, repeated patterns of behaviour, and so on. These kinds of ‘manifest’ phenomena are the items that McDowell treats as having fundamental status in Wright’s account, such that Wright must deny that a general account of language practice in terms not ontologically reduced to non-linguistic phenomena could give us the real facts about language. I suggest McDowell might be right about this even if the Illusion Argument is unsound, given that Wright’s broader anti-realist project is to *ontologically* reduce any concepts of which a grasp cannot be made fully manifest in behaviour to concepts that can. As such, any genuine concepts will be *explanatorily* reducible to these ‘basic level’ behavioural terms.

McDowell leads into his analysis by discussing ‘calling things yellow’ in contrast with ‘things being yellow’. I will trace the reasoning, because it motivates some of his criticisms of Wright with respect to the ‘basic level’. First, McDowell posits a supposedly intuitive distinction between fulfilling the order “bring me a yellow flower” and fulfilling the order “bring me the flower that gave you a feeling of satisfaction after what I said”.²⁴ The aim is to show that although the possibility of this distinction reinforces Wright’s critique of the ‘idiolectic’ conception of understanding, his own positive account ultimately makes it unintelligible.

Note that there is supposed to be an obvious parallel between the idea of executing orders and using expressions correctly, as follows: fulfilling an order is like applying an expression correctly in that certain ranges of actions are prescribed by an order and by the meaning of an expression, and respecting either means performing an action of located within the range. Any concerns we might have about this parallel can be

²³ Cf. §2.2 above.

²⁴ WFR 335; the Wittgenstein passage quoted is *Philosophical Investigations* §460.

suspended on the assumption that a language-use example could be substituted in, as the example is merely illustrative.

According to Wright, the problem with idiolectic understanding is that one cannot distinguish between the correctness of one's actions relative to such an understanding and the mere *impression* of such relative correctness. If it seems to me that an action is relatively correct (in this case, 'if the action gave me a feeling of satisfaction') then on the 'idiolectic' picture I will form the conviction that it really is correct.

However, McDowell points out that for Wright there is a precise parallel between the inability of an individual to distinguish objective correctness from the impression thereof and the inability of the whole linguistic community to do the same, where correctness is conceived in realist terms as determined by something more than communal agreement. But if there is an irreducible distinction between the fact of something's being yellow and the fact of its being called 'yellow' by most speakers of English, inference rule CO leaves no room for it because they are true (or 'applicable') in precisely the same situations.

At this point Wright might respond that room is made for the distinction by virtue of the fact that the predicates 'is yellow' and 'would be called yellow by most speakers of English' are used in different ways or have different cognitive significance. But that will lead Wright into difficulty, because it undermines his strategy of *explanatorily reducing* 'truth' to 'communal consensus', on the basis that the distinction is not a distinction in phenomena but a distinction in ways of describing communal consensus. If we accept the distinction between 'is yellow' and 'would be called yellow by most speakers of English' on Wright's terms we will have to accept that the latter is an analysis of the former. But McDowell's very plausible claim is that we should not accept this, precisely because those two predicates are used in different ways or have different cognitive significance. So this point apparently works against Wright here.

Of course, the success of this argument depends on our acceptance that the fact that we make such distinctions in our non-theoretical usage of our language means that such distinctions are ultimately substantial. But it seems Wright could simply deny this in revisionist spirit—certainly the distinction exists in ordinary language, but the existence of two different predicates that are applied in different ways need not correspond to the actual existence of two different properties. He could argue that our tendency to think that 'being yellow' means a different thing from 'being called yellow by most speakers of English' rests on a mistake.

So we need a reason to accept that such a contrast exists in reality and is not something projected by us. One thing that might appeal is that, intuitively, 'would be called yellow' implies 'might not be yellow'. Indeed, this intuition presents a pervasive difficulty for Wright. His aim is to draw a contrast between communal consensus as truth and a realist notion of truth. But if Wright's account is correct, then what we mean by 'truth' is what he is calling 'communal consensus', and we do not call truth 'communal consensus'. Rather, communal consensus is what we call 'communal consensus'! The concept seems to be abused in Wright's hands; 'consensus' is something we normally oppose to more objective things like true judgments. And that is precisely what he aims to do in opposing it to a realist conception of truth as more than consensus. But if we reject that conception as he desires, then in what respect is it the case that we merely agree? It seems that if there is no difference for us between what our 'most refined criteria of acceptability' tell us is true and what really is true, then we should conclude not that truth is ontologically reducible to acceptability, but that those 'criteria of acceptability' are truth-conditions. It seems the only appropriate kind of reduction would be an explanatory one, but then we stumble over the intuition that consensus that P and the fact that P are non-identical, as the fact that P amounts to *more* than consensus.

This issue is a development of the implosion problem (§2.3). If Wright says that the fact of the matter is that truth is correctly analysed as communal consensus, then that fact of the matter can only be a consensus-transcending fact of a realist variety because our standing practice of considering consensus inferior to true judgement would render it false, because we would judge it false. Of course, this is a problem for any revisionary aspect of Wright's account, and equally so for the case of 'is yellow'.

The facts of this kind that Wright claims as fundamental, that we deal in communal agreement rather than truth, constitute the 'basic level' picture McDowell describes:

... a picture of human beings vocalizing in certain ways in response to objects, with this behaviour (no doubt) accompanied by such 'inner' phenomena as feelings of constraint, or convictions of the rightness of what they are saying... (WFR 336)

The picture is couched in terms of 'behaviour' and 'convictions' rather than applications of expressions and use of concepts because these are the kind of things in which the latter can be 'directly manifest'. Aside from the implosion issue, the problem we might find with this analysis is that the commitments of the reinterpreted contractual picture would not be properly analysed as action-guiding commitments at

the 'basic level', but as dispositions (acquired individually) to secure the assent of our peers. And the problem with dispositions, so the Kripke argument goes, is that they do not guide action; they are not normative.²⁵ If this is right, then the contractual picture does not involve normativity at the basic level, so at the basic level our language practices are not meaningful.

That might seem reasonable, even to a normativist. From one perspective, the basic level is just an analysis of language as a natural animal phenomenon, as a naturalist analyses the practice we call 'building a dam' in terms of the behavioural patterns and propensities of beavers, rather than in terms of engineering and aquatic infrastructure. Those latter concepts are projected onto the situation by us, and a beaver does not need them to undertake the activity we call 'building a dam'. All Wright seems to be saying is that any language practice we have should be analysed in terms of the behavioural patterns and propensities of humans, because learning that practice must not ultimately rely on things like meaning and language, which are things we project onto the situation having learned the language in question

But note that *any* conception of what the beaver is doing is projected onto his activities, even the notion 'pattern of behaviour'. That is why we describe the practice in terms of 'brute movement and causal explanation', as McDowell puts it, to avoid crediting the beaver with a perspective on his activities that he does not have.²⁶ But when any analysis we give of human language practices presupposes the linguistic competence of the analyst, who is also the subject of analysis, why is it preferable to describe what he does in terms of behaviour, when any description we can give will be one he understands? The problem is to say why concepts like meaning are projected while others like behaviour are not. What Wright apparently wants to say is that if we could imagine the world from a perspective where nothing was meaningful for us, then the facts about language practices would be facts about the way we behave and not the meaning of expressions. The reason for thinking this is that the facts in question would not involve language, because language necessarily involves meaning—but then why say that, considered from *our* perspective, those facts would be facts about behaviour and not facts about meaning? Surely it would be more appropriate to admit that the facts would be inexpressible, because they are not couched in language. The claim that behaviour is basic is accordingly under-motivated.

²⁵ Cf. K p.37.

²⁶ Cf. WFR 339

McDowell does not take this line, and instead focuses on the supposed absence of normativity at the 'basic level'. He invokes the 'Illusion Argument' introduced in my §2.3 to argue that the absence of normativity at the basic level of Wright's picture amounts to the claim that, strictly speaking, normativity does not exist. However, we have already seen problems with this strategy, as it involves identifying the reducibility of correctness to consensus as *ontological* reducibility, when Wright could apparently deny this.

Fortunately for McDowell, Wright seems to make the kinds of ontological reductions he is accused of, such as to make an illusion of notions like objective truth. The revisionism underpinning much of *Wittgenstein on the Foundations of Mathematics* is apparently intended to extend to our ordinary conceptions of the way things are. Consider, for example, the following passage:

The dilemma we confront as a community is thus essentially that of the private linguist: faced with the impossibility of establishing any technique of comparison between our judgment and the putative objective fact, we must construe the fact either as something we cannot know at all or—the classic choice for the private linguist—as something we cannot but know. Wittgenstein's response is to urge that both courses are equally disreputable; and that we should therefore abandon the conception which leads to the dilemma. If we do so, we shall reject the idea that, in the senses requisite for investigation-independence, the community goes right or wrong in accepting a particular verdict on a decidable question; rather, it just goes. (W, p.220)

On the highly plausible assumption that Wright endorses Wittgenstein's response here, the claim he makes is that we, *the community*, should reject the conception of the whole community 'going right or wrong'. His idea is apparently not just to make some merely academic claim that leaves our everyday linguistic practices intact, but rather to totally revise our ideas about correctness. The idea of investigation-independent facts is thus construed as an *illusion*—which confirms McDowell's claim that we lose out on an intuitive conception of objectivity.

So long as the picture of the 'basic level' follows from this commitment, it seems fair to say that revising our conception of facts and correctness in this way, as Wright demands, would mean accepting the basic level picture as 'the real truth'. So McDowell is right: it would be impossible to distinguish the kind of commitments Wright describes from the *illusion* of commitments, since properties constructed on top of the basic level do not reflect the way the world really is. As such, Wright's reinterpretation of the contractual picture amount either to a concession that it does not really obtain,

thus failing to meet the minimal adequacy conditions for an account of meaning and understanding here, or that meaning is not normative, denying the basic premise of the debate.

The original problem of distinguishing between the fact that a flower is yellow and the fact that it would be called 'yellow' by most speakers of English is a scaled down version of the problems Wright encounters as I have presented them here. He must maintain that there is a difference between commitments and the illusion of commitments, but McDowell argues that they are unavoidably construed as illusion because putative basic level facts cannot "be prevented from purporting to contain *the real truth* about linguistic behaviour" (WFR 336). However, this argument would be undermined if Wright simply toned down his revisionism, to identify the concepts of objective correctness and communal consensus on an equal plane. I have suggested an alternative line of criticism to the effect that 'basic level' facts are not obviously facts about behavioural propensities rather than facts about correctness. If we do not presuppose a command of the language in question then the events underlying language are not describable at all, so if we are willing to go as far as offering a description using the language we know, then why limit the range of concepts we use to characterise the situation to concepts that describe non-linguistic phenomena?

I have argued that Wright constantly has to grapple with the problem that the terms he uses to characterise intuitively 'realist' concepts tend to characterise them poorly, the most pressing case being the use of 'consensus' to characterise true judgment. This issue resurfaces here when we consider the contrast between activities that lack a meaning, like 'behaving in such and such a way', and meaningful activities like 'claiming that P'. Wright needs to say that the former kind of thing, properly arranged, amounts to the latter kind of thing. But from the perspective of someone located within Wright's reinterpretation of the contractual picture, it would just seem false that the former could amount to the latter because the latter is meaningful to him, and the former is stipulated as having no meaning in itself. His experience of language is apparently not one of people making noises and exhibiting positive behavioural responses, but of people communicating and speaking correctly—asserting, denying, ordering, expressing, and so on. How could a configuration of non-meaningful things amount to meaning? This and other questions motivate McDowell's positive account of the contractual picture, which I will investigate in Chapter 4.

3 UNDERSTANDING AS INTERPRETATION

3.1 INTERPRETATIONS & PARADOX

As McDowell proceeds in his analysis, he becomes noticeably preoccupied with what emerges as a central question, namely, how it is that one can understand something without ‘interpreting’ it. He claims (WFR §§3-4) that failing to answer this question leads to a regress, one that cannot be defused without adopting unappealing additional theses about meaning and understanding.

The postulated regress says that if every kind of understanding is an interpretation of some kind of input, then an understanding of whatever it is that makes such interpretations correct is also open to interpretation itself, and so on, generating a ‘regress of interpretations’. But what is an ‘interpretation’ here? McDowell does not explain this. Indeed, the term is Wittgenstein’s, and he did not explain it either.²⁷ We will have to divine the meaning of ‘interpretation’ from the use McDowell, Kripke and Wright make of it. Fortunately, the task is not so difficult.

On his example of understanding the ‘plus’ sign, Kripke famously writes:

...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 (K, p.107).

The process of interpretation here seems to be something like attributing a particular meaning to something one holds consciously in mind. McDowell offers the following gloss on this quotation, and on the formulation of Kripke’s ‘sceptical paradox’:

That is, whatever piece of mental furniture I cite, acquired by me as a result of my training in arithmetic, it is open to the sceptic to point out that my present performance keeps faith with it only on one interpretation of it, and other interpretations are possible. So it cannot constitute my understanding ‘plus’ in such a way as to dictate the answer I give. Such a state of understanding would require not just the original item but also my having put the right interpretation on it. But what could constitute my having put the right interpretation on some mental item? And now the argument can evidently be repeated. (WFR 329)

Before we look at the argument we can use this analysis to get an idea of what ‘understanding as interpretation’ would amount to. Clearly, that picture involves some

²⁷ McDowell’s source is PI §201.

notion of 'mental objects' that are usable in some conscious way to justify actions as correct. Zalabardo (2002, p.286) helpfully suggests that the notion of justification here involves a requirement of "*conscious engagement* with the facts". If this is correct, as it looks to be, then if 'mental objects' are the kinds of things to which we look for justification then they must be the sorts of things with which we consciously engage, or at least try to. This seems to mean that the objects in question are imagined as consciously accessible, such that I can refer back to them. Moreover, they provide me with a standard by which to judge the correctness of my actions.

In our terminology, this is a picture of correctness relative to an understanding. In the case of expression usage, we stipulated (§1.3) that on the minimal contractual picture, relative correctness arises when S undertakes to limit his usage of the expression to a particular range of ways of using it. We left it open how S makes this undertaking. The conception of understanding as interpretation seems to offer one possibility, namely that the range is something I can 'consciously engage' with. Understanding thus involves my committing to restrain my usage to a range of ways of using the expression communicated by some mental item. However, which range the item itself communicates is subject to 'interpretation': its meaning is something I must infer from the mental item itself. Similarly, regarding a particular application of an expression I can only justify my judgment that it falls within a particular range of usage by inferring that it does, appealing to mental items of this kind to justify the inference.

So McDowell's gloss on Kripke's paradox, based on this conception, says that problems arise when we conceive the justifying mental item as something of a similar kind which itself needs interpretation. If understanding an application to mean a particular thing involves interpreting it in a particular way, hence making some inference that it does, then that particular inference must be justifiable by appeal to an understanding of some mental item that prescribes the inference. If the understanding of that mental item also amounts to an interpretation of it, hence an inference to what inference was sanctioned originally, then this new inference must also be justifiable by appeal to an understanding of something else that sanctions this inference, and so on apparently into regress.

Kripke's sceptic aims to show that facts about present usage, meaning and understanding cannot justify future instances of usage, meaning and understanding. As Kripke puts it, "a fact as to what I mean now is supposed to *justify* my future actions, to make them *inevitable* if I wish to use words with the same meaning with which I used

them before” (K, p.40); but the argument is supposed to show that such facts can provide neither justification nor inevitability for my future usage because appealing to them for justification means understanding (interpreting) them in a particular way, an action that itself requires justification, and so on. Incidentally, I think we ought to understand ‘inevitability’ here in terms of the M1 determinacy of relative correctness-value for a particular application, relative to a particular understanding. If understanding is always conceived in terms of ‘interpretation’, as it is here, then it is impossible for me to justify my interpretation of the range of correct usage to which I commit as counting the application as either correct or incorrect.

The idea is that if our actions are beyond justification, then they are not really objectively prescribed so much as simply chosen by us. The premise that justification requires ‘conscious engagement with the facts’ is notable in its epistemological internalism about justifications, and an appropriate externalist account could possibly defuse the worry. However, I cannot pursue this line here.

The relative correctness-value of a given application is assigned by the understanding in question by virtue of the range of usage to which it commits us. But the challenge here is to justify any claim to having honoured the commitment when the range specified also requires interpretation. This would be true beyond the case of a merely *personal* understanding, and McDowell notes that it relates to Wright’s rejection of the idea that the community adheres to a ratification-independent pattern of application (WFR §3). If a community aims to commit to a particular range of ways of using an expression, then their joint claim that a given application is correct relative to that understanding is based on one of numerous possible interpretations of the range. Everyone may agree that an application falls within the range, but there is apparently nothing we can use to justify our agreement that does not also require interpretation of this kind. This seems to support Wright’s conclusion that we should reject the requirement to justify our agreement by something extra-judgmental, that the community does not go right or wrong with respect to agreement; “rather, it just goes” (W, p.220).²⁸

²⁸ Cf. my §2.4.

We might model the ‘interpretation’ conception of understanding with an analogy. Imagine the community makes judgments of correctness with reference to a chart:

Application	Correctness-Value
A	Correct
B	Correct
C	Incorrect
D	Incorrect
...	...

Judging whether application B is correct, they consult the chart and decide that it is. Kripke’s sceptic would then object that this judgment assumes one particular interpretation of the chart, and demand justification for choosing that interpretation. Why not read the chart in a different way? If all they can cite are further charts, i.e. charts showing how to read charts and so on, then a regress obviously ensues.

The pertinent question is why we would think understanding operates in a comparable way, and whether the realist contractual picture need assume it does. Specifically, the concern arises: why must the mental item which is supposed to justify my interpretation of a particular expression application also be something of the same ‘interpretable’ character as the application itself? The fact that I understand my commitment on the contractual model does not mean that I am free to interpret that commitment however I wish—the problem here does not seem to be so much that I have to appeal to something in my mind to justify my actions, as might be argued, but rather that we imagine that ‘something’ to need interpreting too. Perhaps this view would be inspired by the idea that the mental object is something like an instruction that I obey when I interpret the language use (perhaps also an instruction) in a certain way. That seems to be what underlies the ‘chart’ analogy, if it is a good one.

There is apparently nothing in the contractual picture that says we need conceive our mental objects in this way—we commit to restricting our use to a particular range, and in that there is not obviously any interpretation involved. However, I think the motivation for conceiving them this way comes from the idea of citing mental objects as justification. Kripke’s idea seems to be that I must in principle be able to consciously engage with some mental item that would justify my interpretation of (e.g.) the ‘plus’ sign. But if conscious engagement means the kind of thing I do when I listen to

someone speaking, and I am able to interpret his meaning in any number of ways (provided I can adequately justify myself), then it seems I should be able to interpret my mental items in any number of ways too. That is not the only possible gloss on ‘conscious engagement’ however, so it is admittedly not conclusive that this is what motivates the conception that an understanding of anything, even of mental items, involves a fallible interpretive process. McDowell does not venture to explore what the motivation is—that would apparently be a task to follow the missing account of what ‘interpretation’ amounts to here—but there is evidence that he lays the blame on a failure to reject certain core ideas of Problematic Realism.²⁹ I discuss this in more detail in the final chapter (§4.4), as it might seem a surprising charge to bring against the anti-realistically inclined.

At present we should focus on one particular claim that generate the justificatory regress: that we must be at least able *in principle* to provide *conclusive* justification for the claim (e.g.) that a particular application is prescribed rather than another. This is a characteristically sceptical challenge, as we must be able to explain what it is that makes some action correct relative to an understanding in a way that leaves no room for deviant ‘grue’ interpretations of what it commits us to. If my particular application is to be inevitable, as it apparently should be in the favoured example case where a) I am engaged in the activity of answering addition problems correctly, b) I seem to understand addition and the meanings of the terms involved according to ordinary judgments of such competence, and c) I do not make some procedural error in my assessment of whether the expression “125” is appropriately applied in response to the question “What is 68+57?”, then the sceptical challenge is to demonstrate that a full characterisation of everything that justifies my saying that “125” is correct relative to my understanding can be provided in principle. And of course, the premise that anything I could use to justify this claim is subject to various interpretations will leave the sceptic’s challenge unanswered, as we conceive every putative justification for thinking, acting, speaking (etc.) in some particular way, including the making claims of the form ‘this justifies that’, as requiring justification.

Kripke presents the sceptical paradox as an insurmountable challenge to provide a fact that shows I mean one thing rather than another by my (apparent) use of addition, the actual upshot being, on McDowell’s summary, that “there is no fact that could constitute my having attached one rather than another meaning to the ‘plus’ sign” (WFR 329). This slightly odd turn of phrase is meant to indicate that the supposedly

²⁹ Cf. WFR §11 in particular.

unavoidable regress fundamentally undermines the idea that any mere fact could make it the case, just by obtaining, that I mean one thing in particular by my use of the 'plus' sign. If relying on interpretations of the dictates of my understanding generates the regress, no fact about my understanding can commit me to meaning one thing rather than another—to meaning green rather than grue or plus rather than quus—because my understanding alone is apparently unequipped to guarantee that I do the sort of thing I committed myself to doing. McDowell seems to agree with this: there is no hope for *this* conception of understanding for precisely these reasons, so we should reject the idea that understanding is always a kind of interpretation. That is rather quick, however, and we would seem to be accepting the sceptic's demands without questioning whether they are reasonable, as well as being impressed by the putative regress. In fact, it will emerge (§4.2) that McDowell's dialectical approach here is not to engage with Kripke's sceptic at all, but to reject the challenge by presenting a superior account of the relation between meaning and understanding.

Kripke's aim in analysis of the paradox is not to show that expressions can never be meaningful, but to show that some kind of anti-realist justification-based semantics needs to replace our traditional truth-conditional accounts. Sentences like "I have attached a determinate meaning to the word 'green'" are then not false but merely misleading; they should be understood "in terms of their use to record acceptance of individuals into the linguistic community" (WFR 330). Making contact with Wright's anti-realism, Kripke's 'sceptical solution' to the sceptical paradox aims to preserve meaningful language use with the crucial caveat that 'meaningful language use' here is really a term of art that is explanatorily reducible to behavioural practices which are in line with the preferences and propensities of the linguistic community.

McDowell's arguments for rejecting the sceptical solution are highly condensed, but his main criticism seems to take the following form:

The right response to the paradox ... is not to accept it but to correct the misunderstanding on which it depends: that is, to realize 'that there is a way of grasping a rule which is *not* an *interpretation*'. (WFR 331)

He offers two arguments in support of this claim. The first is as follows:

It is natural to suppose that if one says 'There is no fact that could constitute its being the case that P', one precludes oneself from affirming that P; and this supposition, so far from being a distinctively 'realist' one, plays a central role in the standard arguments *against* 'realism'. Given this supposition, the concession that Kripke... makes to the sceptic becomes a *denial* that I understand the 'plus' sign to mean

one thing rather than another. And now—generalizing the denial—we do seem to have fallen into an abyss: ‘the incredible and self-defeating conclusion, that all language is meaningless’... It is quite obscure how we could hope to claw ourselves back by manipulating the notion of accredited membership in a linguistic community. (WFR 330)

On the face of it, this is not so different from the ‘illusion’ argument McDowell uses to criticise Wright. However, at the end of Chapter 2 I suggested that perhaps the real substance to that argument is that the kind of anti-realist claims being made are false from the perspective of someone who operates within the anti-realist picture being described.³⁰ Though the passage is a little obscure, the argument seems roughly to be that if the sceptic’s challenge leads to the conclusion that ‘no fact can be found that could constitute its being the case that P’, then the appropriate response is to revise our premises and working definition of ‘fact’ to account for this, unless we accept that the sceptic has won and that we never have justification for thinking that P. However, the first option seems to leave concepts like truth and falsehood basically intact, merely accepting that they cannot be explained in this particular paradox-generating way. To infer that they are unsalvageable is too quick from this perspective, but McDowell accuses Kripke of doing just that in “conceding that the sceptic’s negative assertions are unanswerable” (K, p.66). McDowell seems to be playing on the idea we saw in §2.4 that trying to reconstruct what we call ‘truth’ out of less imposing concepts like justification leaves us with something short of that original concept at the non-theoretical level. If we want it to be a genuine truth that justification is all truth can amount to, then we seem to have said not just that truth must be conceived in a particular way for coherence, but also that it is something less rigid and objective that we are inclined to think.

The apparent analogy in meaning and language would be that if it were a genuine fact of the matter that words have no meaning, that would seem to licence the conclusion “that all language is meaningless”. But this is just the ‘Illusion Argument’ again, and we only found grounds for that as applied to Wright because a) on Wright’s account, the derivability of CO means that generally counter-intuitive claims are false, and b) we found good reason to suppose that Wright intends his claims to be taken as facts that should be acknowledged as the ‘real truth’ about correctness etc. Given that a) is not obviously an option for Kripke’s case, McDowell would have to show that Kripke intends his revisionary ‘justification-conditions’ project for a theory of meaning to be

³⁰ Note that Wright’s apparent acceptance of CO was what leads from this to the ‘implosion’ problem. It is not so obvious that this principle is derivable from Kripke’s ‘sceptical solution’, though of course if it is he will face that problem too.

acknowledged as more than an explanation of where our perfectly well-functioning everyday concept of sentence truth comes from; that he intends it to detract from that everyday concept in some sense. This would give McDowell some grounds for applying the Illusion Argument here—but pending that, Kripke might just be able to deny that a revisionary theory of meaning need revise our intuition that language is meaningful. The absence of truth-conditional semantics from a theory of meaning does not obviously entail any weakening of the notion of truth that is held by speakers of the object language.

3.2 SUPER-RIGIDITY

If meaning is only possible by virtue of our common ability to adhere, in what we say, to the rules that govern permissible meaningful usage of expressions in language, then it cannot be the case that every rule is always open to interpretation or I could never be sure of meaning the same thing by a given expression as you do. McDowell's concern about Kripke's 'sceptical solution' seems to be that the intuitive certainty and confidence with which we use expressions is totally undermined by the sceptical paradox, and so the switch to a focus on acceptance in a linguistic community would not be able to restore it.

In order to avoid this consequence, McDowell encourages a defence of the idea that we have *fundamentally shared commitments in language usage*, and not merely in patterns of meaningless behaviour. This would mean preserving the idea, contra the paradox, that there is a fact that 'could constitute its being the case that P'. But that means challenging the sceptic's arguments rather than accepting them as Kripke does, and of course the strategy here is to challenge the premise that understanding is always a kind of interpretation. Presented in a way that also addresses McDowell's 'illusion' concerns about Wright's anti-realism, the contention is that if we are committed to more than simply keeping our peers satisfied then there must be a sense in which we understand the meaning of an expression without a mediating interpretation.

McDowell argues against the idea that the problem can be solved without rejecting the conception of all understanding as interpretation. That would require an explanation of how, at some level, we always interpret the mental object (still totally ambiguous in itself) in the correct way. This amounts to the following unlikely 'super-rigid machine' conception of understanding:

Understanding an expression ... must be possessing an interpretation that cannot be interpreted—an interpretation that precisely bridges the gap, exploited in the sceptical argument, between the instruction one received in learning the expression and the use one goes on to make of it. (WFR 332)

The mythology translates roughly as the idea that I am gifted with an inexplicably reliable ability to interpret certain action-guiding inputs of the consciously engage-able kind in the same way every time. If in understanding the meaning of an expression there is a range of ways of using it to which I commit, and that range is something I can commit to only by virtue of some mental item I possess interpreted in a particular way, then such an understanding is 'super-rigid' just in case I always interpret that mental item in that particular way. So I can break off the regress by appealing to my super-rigid understanding of one of the interpretations in the chain—and of course, the same idea might be applied not at the level described but at some further point down the chain of justifying interpretations. Put this way, the 'super-rigid machine' idea just seems an ad hoc response to the regress of interpretations; we simply postulate a level at which our troublesome freedom to interpret action-guiding inputs in more than one way is curtailed.

It is important for McDowell's attempt to build an alternative contractual picture that we understand this as problematic mainly due to its attempt to end the regress by postulating a step in the chain of reasoning which is identical to those that precede it, albeit with the crucial caveat that it ends the chain, and not due to the mere fact that it involves understanding of a kind that is not subject to (mis)interpretation. For if it should be rejected simply on the latter grounds, then McDowell's overarching aim to reject the idea that understanding is always interpretation looks like exactly that manoeuvre, an attempt to evade the problem by insisting that some kinds of understanding are not subject to interpretation in the regress-generating way.

Rather, as he puts it in a later work, here the 'meaning' of the 'plus' sign that we might claim to understand super-rigidly is conceived of "as an item of the same character as the interpretations that succeed one another when one embarks on the regress" (McDowell 2009a, p.83). In the 'chart' example (§3 above) this would mean referring back to a foundational chart which, though simply a chart like any other, is somehow totally unambiguous, such that one cannot but interpret it correctly.

Such charts clearly do not exist. If the analogy is a good one, then by 'an item of the same character as the interpretations' we would seem to mean something very much

like ‘unambiguous verbal instructions for interpreting verbal instructions for interpreting verbal instructions...,’ and so on, except perhaps that the instructions would not be verbal but instead couched in a language of thought (one of Wright’s troublesome ‘idiolects’), or in thoughts themselves, or in otherwise super-rigid mental states. Super-rigidity of any kind seems to entail total infallibility in some kinds of practices, which is thoroughly implausible. We might argue that one of the tenets of ‘conscious engagement’ would have to be that anything done consciously is done without this kind of super-rigidity, ruling out (by Kripke’s standards) any appeal to interpretations arrived at by super-rigid interpretive faculties as facts that could count as justifying claims to mean one thing rather than another. That seems plausible, though it would need development that I cannot offer here. I will merely suggest, following McDowell³¹, that this connects with the kind of criticism that Wright makes of an idiolectic conception of understanding as offering justification for absolutely any sincere verdicts I draw regarding the relative correctness of my own actions.

However, as noted earlier McDowell’s strategy is to reject the account of understanding that generates the problem to which the super-rigid conception is a response. On the assumption that super-rigidity is sufficiently problematic, McDowell presents accepting Kripke’s sceptical paradox and accepting a ‘super-rigid machine’ solution to it as two horns of a dilemma that must be confronted if we fail to reject the conception of all understanding as interpretation (WFR 331-2). But whether either presents an intractable picture in itself, hence meriting the title ‘horn of a dilemma,’ is not really up for discussion in WFR. McDowell is clearly convinced that neither option is sustainable on its own merits—and recall that he seems to see Kripke’s ‘sceptical solution’ as no better than accepting the sceptic’s own conclusion that language is meaningless. For reasons that will become apparent in the next chapter, I will refer to this as the ‘Small Dilemma’.

Presenting the dilemma as based on false premises significantly improves McDowell’s position in the dialectic. If the diagnosis is correct then it does not matter whether his previous arguments for rejecting Kripke’s sceptical solution are convincing. If the paradox is generated by a mistake, then the correct response is to correct the mistake and stop it from arising in the first place. Correspondingly, it also does not matter whether we ought to reject the super-rigid machine mythology for being an ad hoc response, unintuitive, implausible etc. because it is a response to a problem that should not arise. As such, all attempts to respond to McDowell’s prior treatment of these

³¹ Cf. WFR 332-3

Rule-Following, Normativity & Objectivity

topics will be thwarted if he can demonstrate that there is a more fundamental problem, and that we ought to reject the dilemma. This, in effect, becomes the primary goal of the paper.

4 MEANING-PERSPECTIVISM

4.1 McDOWELL'S TASK

In preceding chapters I have examined various lines of enquiry that McDowell pursues in WFR, and attempted to provide some sense of the dialectic outside of his attempt to demonstrate that Wright and Kripke have produced readings of Wittgenstein that are not historically representative. I have suggested that Wright's strong reductionism leaves him open to McDowell's charge of unacceptable counter-intuitiveness, alongside a risk that his position will undermine itself. However, McDowell's attempts to make the charge of counter-intuitiveness stick for Kripke's case are less convincing because Kripke does not seem to hold the same revisionist agenda.

He thus appeals to the broad strategy of the 'Illusion Argument', which aims to show that if claims about meaning and understanding are ontologically reducible to claims about the non-meaningful activities and attitudes of human beings, then this somehow makes the former claims less than genuinely representative of the truth, such that 'x means y' will not be strictly representative of the facts, compared to something like 'when people say "x" they are expecting others to react in a particular way w' (for example). However, the Illusion Argument is not effective against anti-realists who make merely *explanatory* reductions. If water is explanatorily reducible to H₂O, nothing controversial about the reality of water follows.

McDowell apparently needs to say why characterising meaning in terms of non-normative behaviour and other basic level elements will always be an *ontological* reduction, or why it otherwise detracts from the normativist's intuition that those who understand meaning are subject to norms, if the Illusion Argument is to be successful. Being 'subject to norms' may be a legitimate way of conceiving of being inclined, at the basic level, to do some things rather than others, and sharing that inclination with others.

In fact, I think that is the sort of account that McDowell might want to propose. His criticism of Wright and Kripke is aimed at their *rejection* of truth-conditional semantics, and the project of *replacing* it with an anti-realist or otherwise justificationist alternative, not merely the project of explaining the former in terms of the latter. Even Kripke says that claims about someone's meaning something cannot be analysed beyond the fact that there are "roughly specifiable circumstances under

which they are legitimately assertable, and that the game of asserting them under such conditions has a role in our lives,” such that “no supposition that ‘facts correspond’ to those assertions is needed” (K, p.78). It seems right to characterise this as a denial that a truth-conditional analysis is appropriate, given the charge of superfluity.

McDowell thus sees both Wright and Kripke as attempting to say that certain things are *true* about language, while simultaneously denying our apparent intuition that things like true judgments of fact are meaningful and correct by virtue of more than the significance of their component words to us. However, the *opposition* of the justification model to our standard notion of truth seems to leave it mysterious what the contrast amounts to. What can be meant by the apparent claim that what we call ‘truth’ is something less than truth—given that we also call that latter thing ‘truth’?

It seems anti-realists must be saying that things that we ordinarily take to be true are less than true, but depend merely on justifiability: but then this presupposes some other kind of truth that our justifiable judgments ordinarily fail to respect, it being at best a happy coincidence if our judgments coincide with this superior kind of truth. But this is coupled with the claim that our ordinary notion of truth is really such a notion of justifiability conditions: so then surely we should be saying that judgments we ordinarily take to be true (because we are appropriately justified in doing so) really are true, not merely appropriately justified. But then we seem to have lost our point of reference for the claim that our *ordinary* notion of truth is not really truth. We can apparently reject some transcendent notion of truth as something inaccessible as championed by the Problematic Realist—which I have argued that this is not to be identified with realism in general—because that conception seems to entail the claim that we are never justified in judging judgments true. But the denial of truth-conditional semantics seems falsified by the idea that we are respecting all and any notions of truth to which we do have access, and if we reject surprisingly strong realist principles like LI then there is no reason to think that the notion of truth is in danger here. The only way we can force that conclusion is by arguing (somehow!) that truth is not really a concept that applies in our language practices, and that justification is really what we mean by truth. But that kind of argument is error-theoretic, and seems to tell us that nothing is really true; that truth is an illusion. This is the motivation for McDowell’s ‘Illusion Argument’ strategy, which is supposed to show that if we get this far then it will follow that meaning is an illusion.

However shaky that may seem, the Illusion Argument also draws on an additional point McDowell hopes to demonstrate: that both Wright and Kripke allegedly assume, wrongly, that understanding is always something like an interpretive process. Pressing this point means attempting to deny a premise that underpins their arguments, and showing that the debate here revolves around a faulty account of understanding would be a great achievement. Strictly speaking, McDowell's aims are slightly more modest than this. He conceives his positive claims as providing an exit route from a dilemma generated by the assumption. This fact alone does not validate his alternative, though it might have the secondary effect of (further) persuading us that understanding cannot always be a kind of interpretation.

Though McDowell wishes to reject the assumption that understanding is always interpretation, his alternative account preserves most of the core premises we have seen in the debate—notably, that meaning is normative, and that some form of the contractual picture must be able to account for the way that meaning and understanding relate normatively to language practice. He aims to respect the first, and demonstrate the second. He also aims to preserve the kind of epistemological internalism that seems characteristic of the justification debate surrounding Kripke's paradox: he is apparently sympathetic, at least for the sake of argument, to the idea that if a particular application of an expression is meaningful and correct then we must be able to know that it is both meaningful and correct. I think the supposed plausibility of this premise comes from that idea that correct and incorrect language use must be at least partly administered (evaluated, encouraged, corrected and so on) by language users themselves. The disagreement between realists and anti-realists is about the extent of their role as arbiters: anti-realists think that respect for the arbitration of language-users is all that the normativity of meaning could amount to, while realists would contend that it arises, at least in some cases, out of a need to respect the genuine fact that some things just are (in)correct uses of language, whatever we judge. But the realists here are not committed to the idea that the facts transcend knowability, or that the arbiters' judgment is somehow an obstacle to the truth. All they have to admit is that the arbiters themselves owe allegiance in judgment to facts about the world, and accordingly, that their judgments are open not only to peer-review, but to comparison with the truth—and importantly, that it is possible for us to know if the judgments are false. This is the internalist premise.

Of course, McDowell does reject the anti-realist's premise that all kinds of meaning must be 'fully manifest in behaviour'. Recall that this means, roughly, that if something

has a meaning then it must be intelligible to someone without presupposing that they understand other meaningful utterances, concepts (etc.) in a way that defies analysis in non-semantic 'basic level' terms. In fact, McDowell's aim in the last few sections of WFR is apparently to challenge the coherence of this demand, and the possibility of explaining meaning in anti-realist terms.

If McDowell's alternative account of the contractual picture relies on facts about human nature to characterise the formulation of true judgments then McDowell must distance his account from Problematic Realism. Indeed, he wants to reject it outright, as should be apparent from the mocking tone of his formulation of that idea (cf. my §1.2). What McDowell wants is objectivity in correctness without something as strong as LI. The idea to be preserved is simply that a judgment about the shape of an object is correct or incorrect with reference to the object, and not solely in respect of human attitudes (and so on). This seems like a reasonable goal, but the various assumptions that frame the debate make this a serious obstacle, and a case must be made for its possibility.

4.2 THE BIG DILEMMA

McDowell offers a summary of the project he faces in the form of a 'Scylla and Charybdis' metaphor (WFR 341-2). He claims we are faced with the task of avoiding a dilemma, distinct from the Small Dilemma mentioned in §3.2. This new dilemma offers us a choice between the assumption that understanding is always interpretation, or the conclusion that the 'basic level' picture derived from Wright's account (§2.4) represents the way things really are, and (so it is claimed) that language is therefore properly understood as meaningless and normativity does not genuinely exist. I will call this the 'Big Dilemma'.

McDowell claims that the first horn of the Big Dilemma ('Scylla'), the assumption that all understanding is interpretation, leads directly to the Small Dilemma: confront Kripke's sceptical paradox and interpretive regress, or adopt some kind of super-rigid account of understanding as interpretation to avoid it. We have seen some evidence to support this claim, but it is strictly speaking an open question whether we can take understanding to be a kind of interpretation without having to face the Small Dilemma.

The second horn of the Big Dilemma, McDowell's 'Charybdis', is the account of the basic level that Wright is committed to. McDowell claims that on this account "norms are obliterated" (WFR 347) because language practices seem to reduce ontologically to a

kind of “brute meaningless sounding off” (WFR 336), a complex but merely causal result of the training human beings receive, not sanctioned by any commitments they make and thus not subject to correction. The suggestion is that if we embrace the second horn “then we cannot prevent meaning from coming to seem an illusion” (WFR 342). However, there is apparently still scope for making an explanatory reduction of language to basic level terms if we reject Wright’s claim that the reduction is ontological.

McDowell’s aim is to reject the first horn without committing us to the second; to reject the idea that understanding is always interpretation while *also* rejecting the idea that commitments, meaning and normativity are ontologically reducible to meaningless ‘basic level’ facts about behaviour. This requires not only a replacement notion of understanding, to treat the first horn, but an alternative contractual picture that can explain how meaning comes about in a way that does not permit the reduction. The account accordingly needs to allow that causal mechanisms like training aim to produce certain complex behavioural responses, a neuro-psychological thesis about humans which cannot be easily denied, while maintaining that full-fledged language practices and the commitments supposedly involved are not ontologically reducible to facts about those sorts of things. The real trouble with this is that accepting that we must explain how meaning and normativity arise out of behavioural and psychological phenomena seems to commit us, by the (questionable) reasoning of the Illusion Argument, to saying that these things *are* ontologically reducible to behavioural and psychological phenomena, because we can give a full characterisation of the former in terms of the latter. McDowell will have to maintain that such a characterisation is not possible.

The task is thus to get from the common sense psychological notion that language begins with behavioural training guided by correction-behaviour on the part of the teacher—indicating to the trainee that his or her actions were incorrect by tone of voice, reinforcement of the correct alternative by repetition and so on, comparable to the way that one trains a dog—to the idea that the contractual picture involves commitments to future use in a way different to the ‘commitments’ an infant makes to say ‘mama’ and not ‘dada’ to its mother as a result of a kind of conditioning. What needs to be established is that language, commitments and the normativity of meaning are only *explanatorily* reducible to such terms, preserving their genuine existence.

If a commitment is different from a disposition, if what I *should* do is not just an unhelpfully complicated way of saying what I and others in the community *will* do, then we should expect to find a clear distinction between the kind of behavioural and psychological facts underlying such commitments and the kind of stimulus-response behaviour of someone who is psychologically conditioned to respond to certain stimuli in certain predictable ways. And yet without any grounding in scientific theories of language acquisition, it seems McDowell is not going to be able to lend any support to that claim. Furthermore, it is not clear that a scientific theory of the difference would really help anyway: it would still talk about psychology or neurology, and not about obligations. We would still have an explanation of the behaviour I *will* exhibit, but not prescriptive reason for why I *should* exhibit it. We might try to provide some complicated theory of how communally shared dispositions to undertake certain patterns of behaviour are the basis of institutions like ‘commitments’—but then we still seem to have accepted the second horn of the Big Dilemma, because commitments to calling a spade a spade rely on facts about humans and not facts about the soil-shifting implement in question and the meaning of the word ‘spade’.³²

Introducing new basic level elements into the picture will not help, and introducing supernatural elements is clearly not an option. So how are we going to evade the second horn of the Big Dilemma? In fact, this reflection on the idea of facts underlying human activities provides a hint to McDowell’s solution to the problem. As I read him, McDowell’s answer takes the form of a thesis about the interrelation of certain kinds of ideas, to the effect that while meaning must be analysable in basic level terms, it cannot be ontologically reduced to basic level phenomena—precisely the conclusion that is required. In a sense, the claim is that the analysis of meaning and understanding in behavioural terms provides as full an explanation as is possible of it as a *non-meaningful* phenomenon, but in fact the content of that explanation is not to be identified with the concept of meaning; something is missing. Accordingly, the absence of genuine commitments and normativity from the basic level picture does not entail their non-existence, but simply that if they do exist then the basic level picture is an incomplete picture of meaning and understanding. This thesis is part of McDowell’s alternative gloss on understanding and the contractual picture.

This thesis involves the strong claim that the explanation of meaning and understanding in basic level terms involves an insurmountable obstacle. To support this kind of claim, we need to see why the obstacle is insurmountable. The thesis also

³² This is the gist of McDowell’s rejection of Kripke’s ‘sceptical solution’; cf. my §3.1.

seems to involve the claim that the relevant notions inherently defy analysis, which would need to be argued for if it is meant to be something more than the kind of *ad hoc* irreducibility characteristic of the super-rigid machine idea, the mere stipulation that things like commitments and normativity exist in an inherently analysis-defying way. I propose these as requirements on an acceptable alternative account.

4.3 MEANING-PERSPECTIVISM

McDowell aims to ground his solution to the Big Dilemma in the Wittgenstein-inspired thought that “we have to realize 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” (WFR 339). As opposed to verifying this claim, the immediate challenge is to make sense of it in a way that effectively explains why the notion of a practice can help us here. We need to see how this idea is relevant.

Consciously seeking to avoid *ad hoc* irreducibility, McDowell’s account of understanding will incorporate a notion of what we might call ‘non-interpretive comprehension’ of an instruction, commitment or prescription. When I commit, on the minimal contractual picture, to restricting my usage of an expression to a limited range of ways of using it, I act in accordance with this commitment without seeking guidance at all. The same applies when I follow an instruction: if I understand it, and am committed to acting in a particular way it prescribes, then I follow it as a kind of ‘second-nature’. So then,

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. (WFR 339)

So the training I have is somehow meant to explain how I can ‘simply act’ in a way that counts as following a sign-post. But without further elaboration, it seems like the ‘training’ in question is just an *ad hoc* device to bridge the gap between the meaning of the sign-post and my understanding it correctly, in precisely the super-rigid manner that corresponds to the second horn of the Small Dilemma. This cannot be what McDowell means, and we should suspend that concern for the moment.

A more relevant objection, which McDowell does consider here, is that the training merely provides “a causal explanation of certain bits of (what might as well be for all that you have said) mere behaviour” (WFR 339). The point seems to be that training

someone to respond in certain ways when confronted with sign-posts may cause them to do the right thing, but it does not explain why they *should* do what they do—why the ‘understanding’ produced by the training is something that commits us to acting in the right way, rather than offering a psychological compulsion to act in a way that just happens to be what we call ‘right’ most of the time. So for example, I may be conditioned to respond to squares with expressions like ‘That is a square’ and not with expressions like ‘That is a squirrel’, but the *fact* that this is what I *should* do is not what motivates me to undertake this course of action—rather, I am compelled by engendered desires and attitudes, my neurology and psychology, and so on. It seems I could exhibit these reactions without any idea of their meaning. Here it would seem that objectivity in facts about the external world would amount to this: that I am trained or conditioned to respond to certain stimuli consistently in a particular way, as is everyone else, and properties like ‘square-ness’ are ultimately explicable as complex descriptions of facts about my behaviour, psychology and the conditioning process, and not as properties of the world as it exists independently of me. Accepting that training is simply the cause of a complex behavioural effect entails accepting the second horn of the Big Dilemma unless we can avoid an ontological reduction here.

There are no commitments involved in being trained to behave in certain ways—I can train myself to drink less coffee, but this is not the same as upholding a commitment to drinking less coffee. In a case of mere stimulus-response training, we can only say what we *expect* the subject to do, not what he is committed to doing in the sense of the contractual picture—for he is not properly described as committed to anything. We could say that he ‘should’ do x, but not in the normative sense of ‘should’—only in the sense that my car *should* start if I turn the ignition key. There are no obligations here.

The challenge is thus to say how the training McDowell is appealing to can result in more than mere behaviour of the stimulus-response kind. It must certainly be something more complicated than conditioning or hypnosis, but in the end it seems it would have to be a causal process that results in a behavioural output. Where do commitments to correct action enter this picture, if not as constructs upon that behaviour?

McDowell’s answer is not particularly illuminating:

The reply...is that the training in question is initiation into a custom. If it were not that, then the account of the connection between sign-post and action would indeed look like an account of nothing more than brute movement and its causal explanation; our picture would not

contain the materials to entitle us to speak of following (going by) a sign-post. (WFR 339)

The claim is that the idea of initiation into a custom, which is conceived of as something comparable to a language practice, is what gives us the materials to be able to talk about 'following sign-posts' rather than behaving in certain complex ways in response to stimuli. It is unclear what the significance of this is supposed to be, nor how conceiving the training as 'initiation into a custom' provides us with any materials that we previously lacked. McDowell claims, disappointingly, that this is "only a programme for a solution to Wittgenstein's problem" (WFR 342). However, I have suggested (§2.3) McDowell's best strategy for substantiating his Illusion Argument is by showing how the programme might be executed.

I will now sketch an interpretation and possible development of this reply. As I understand it, the point is supposed to be something like this. On McDowell's account 'following a sign-post' is a distinct concept from 'behaviour of kind K', and these descriptions attach to different kinds of situations. This, I hope, is relatively uncontroversial. The behavioural processes exhibited in someone's following a sign-post may be identical to those exhibited in behaviour of kind K. And yet, to an observer who is party to the relevant practice—who has learned the concept 'following a sign post'—a distinction can be made between the two. We may observe the same behavioural processes, but knowing what it means to follow a sign-post allows us to say more than that S is exhibiting behaviour of kind K: we can say that S is following a sign post. When we are conducting experiments of a certain kind we may prefer to talk about 'the patterns of behaviour S exhibits' rather than 'S's following a sign-post'. But to assume that the behaviour is all that is going on is to neglect the fact that this kind of behaviour is called 'following a sign-post' by people like us, who are party to the relevant language practice of applying the predicate 'is following a sign-post'. The analogy would be that the written word has a meaning for us, and is more than merely ink-marks on paper: though of course, on a particular analysis it is simply nothing more than such ink-marks. Indeed, on a particular analysis it is not even ink-marks that compose the words you are now reading, but sub-atomic particles of all varieties—although it would be a strange conflation of levels of analysis to insist that these are what you are reading. And this is the sort of point McDowell is making: we are observing the behavioural and psychological phenomena we call 'meaningful language use' and wondering how meaningful language use can be found among these meaning-devoid events, wilfully suspending our ordinary inclination to see those

events as instances of meaningful language use. The presupposition is that in suspending this inclination we will get a better view of what meaning consists in—that is the motivation for the anti-realist’s criterion of ‘manifestability’—but it is this very disposition to find things meaningful that meaning depends upon.

On this expansion of McDowell’s programme, the training in question is induction into a way of thinking of our surroundings and responding to them, and this is the kind of thing meant by ‘custom’ or ‘practice’. Someone who is trained in the use of sign-posts is able to conceive behaviour of kind K as ‘following a sign-post’, because has been trained to recognise certain complex patterns of behaviour as ‘following a sign-post’. In a sense he is conditioned to think this way by the training—but note that the word ‘conditioning’ implies things like reduced intentionality in actions, and that this is unlikely to be how he conceives his own condition, or how we would conceive it having undergone the same process. This is how the distinction between basic level terms and terms like meaning and normativity can be accounted for without detracting from the status of the latter kind as genuine: they are genuine for us.

If two things are explanatorily reducible to one another, this can sometimes be explained by the fact that we *conceive* the same phenomena presented to us in two different ways. An example of two such explanatorily reducible conceptions would be a) the spoken claim that snow is white and b) a particular meaning-devoid vocalisation made as part of a complex of psychological and behavioural events (etc.). I will call a particular conception of things a ‘*perspective*’, although for the sake of getting an outline of the account I will not worry too much about how ‘conceiving of things’ takes place—let’s imagine it is something like a way that the thing is presented to me. S ‘*has*’ or ‘*occupies*’ a particular perspective just in case S has the disposition to conceive things in that way. By contrast, one merely ‘*consciously adopts*’ a particular perspective just in case one conceives things in that way deliberately. In addition, if S occupies a perspective then it seems to follow that S can consciously adopt that perspective, as in the case where S consciously adopts a perspective other than the one S occupies and then deliberately returns to the original conception, perhaps by thinking things like “I must now conceive of this thing in the original way again”.³³ Initiating S into the practice of following sign-posts will mean training S to occupy a perspective whereby sign-posts are things that can be followed.

³³ This raises a number of questions about intentionality that I cannot address here, such as whether one can deliberately do something one is already naturally inclined to do, so strictly speaking the matter is up for discussion. The example case seems plausible enough, though.

What McDowell needs to say is that the perspective of the linguistically competent means they are disposed to conceive things like meaningful language use in ‘more-than-basic-level’ terms, such that when they encounter what might be described as ‘a certain pattern of sounds under certain conditions’, they will conceive it as ‘the claim that snow is white’. We might say that this is explained by their conditioning to think in certain ways, but we could explain it equally well by their understanding of the expression ‘snow is white’: the difference is one of perspective. We cannot insist that the former provides a ‘more genuine’ picture of the facts without adding additional premises to the effect that consciously adopting the perspective from which their actions are meaning-devoid provides a more accurate picture. The onus would be on us to do so, because we *share* the perspective of the individual in question, so we too are disposed to conceive what might be described as ‘a certain pattern of sounds under certain conditions’ as the claim that snow is white. I have used the awkward preface ‘what might be described as...’ deliberately: an important claim on this account is that someone who understands the spoken claim ‘snow is white’ must consciously adopt a *different* perspective on the event in order to conceive it as a pattern of meaning-devoid behaviour.

This is ‘*meaning-perspectivism*’. The account incorporates the following theses regarding meaning and understanding:

- (MP1) An expression E is meaningful \rightarrow there is some perspective from which E is meaningful.
- (MP2) S understands the meaning of an expression E \rightarrow S occupies the perspective on E³⁴ required to find E meaningful.

As this is really just a sketch, I will leave it open whether MP1 allows for the existence of perspectives that no-one has.³⁵ Answering that question would mean establishing (among other things) whether new perspectives are produced whenever language evolves and new meanings are produced. Note that we talk about perspectives on expressions and not on basic level phenomena, the point being that expressions are genuinely real phenomena.

By MP2, if a particular process of training results in S occupying the perspective according to which an expression E is meaningful, that training results in S

³⁴ i.e. ‘conception of E’

³⁵ It would be an interesting project, sadly also beyond the scope of this thesis, to adapt a kind of possible-world semantics in order to express these ideas using a system of modal logic. Opening up that kind of precise analysis might help highlight improvements that could be made to meaning-perspectivism.

understanding the meaning of E. How does this sit with the minimal contractual picture (C+M2-4)?

If meaning-perspectivism is consistent with principle C then MP2 can express at most a *necessary* condition for understanding the meaning of E. Occupying the relevant perspective on E could only be a *sufficient* condition for understanding the meaning of E if doing so entails undertaking to use E only in a range of ways R, or if C is false. We might be inclined to take the latter option and reject the minimal contractual picture—but that cannot be McDowell's choice because he is committed to C.

I suggest McDowell should try to make commitments a consequence of adopting the relevant perspective on E by considering things as perceived by an occupant of that perspective. What needs to be said is that when S is trained in the meaning of E, from S's perspective it is not the case that S is being conditioned in order to cultivate the disposition to conceive certain meaning-devoid phenomena as meaningful, but rather, that S is being taught the correct use of E. Saying this would allow room for the normativist's intuition that understanding the correct use of E entails something prescriptive, and for the subsequent analysis of that prescription in 'contractual terms'. The idea is to reconcile the dispositional aspect of understanding the meaning of an expression, that training of a brute causal kind is certainly involved in the acquisition of that understanding, with the normative aspect, that one recognises certain uses of the expression as correct and that correctness is prescriptive. As before, it is supposed to be an easy move from the prescription of correct use to the idea of a commitment to restricting one's use to a certain range, and any potential problems with that will be identical to those we saw in §1.1.

This option seems open to McDowell, but the enabling claim that S conceives his training in terms of correct use will ultimately require an extended exposition and defence that I cannot offer here. The same will apply, of course, to the notion of a 'perspective'. In any case, our concern should be to see how this rough form of meaning-perspectivism might deal with the Big Dilemma on the assumption that MP2 and C are consistent, in order to assess whether proof of this would be worth pursuing by the requirements of McDowell's task.

The meaning-perspectivist gloss on the contractual picture would seem to be C+M1-4+MP1-2, but M1-4 will have to take on very different meanings in this context. What we cannot allow here is that the range R of correct ways of using an expression E that S commits to in understanding the meaning of E is some kind of mental list of ways of

using E—that would bring back all the problems associated with the interpretation model of understanding, and would seem unnecessary given our notion of causal behavioural training as underlying recognition of correctness. Perhaps we could allow that the training furnishes S with some kind of tacit knowledge of R, such that S recognises an application of E as correct just in case E is located in R. Such details would need to be elaborated in a properly formulated exposition of meaning-perspectivism

4.4 DODGING THE BIG DILEMMA

If S understands the meaning of E then S occupies the perspective from which E has that meaning, and from which certain actions are conceived entirely in terms of their being correct or incorrect applications of E. S is thus able to apply E correctly without interpreting the range of actions prescribed by his understanding of the meaning of E, because actions in that range are not conceived in terms of meaningless behaviour but simply as correct uses of E. Nothing requires interpretation because the uninterpreted materials in question, the potential behavioural patterns we would have otherwise had to assess against the range of behavioural patterns that count as correct use of E, are not 'bare' to us here at all. Someone who occupies the relevant perspective does not need to interpret the meaning of an expression any more than I must interpret the sensory information conveyed by my eyes to my brain. Taking the analogy further, the perspective in question is also not the sort of thing we could have 'interpreted' anyway. The process of understanding the meaning here is, if you like, 'passive'.

This indicates the possibility of avoiding the first horn of the Big Dilemma. We do not see uses of the expression as meaning-devoid behaviour, but as correct or incorrect actions, and we do not need to 'interpret' our understanding to check this. The problem with the 'understanding as interpretation' model is that it assumes, for example, that in understanding spoken language I am presented with patterns of meaningless vocal behaviour which I must interpret by consciously engaging with some mental object in the right way; or at least, that my understanding implies there is such an object with which I could have engaged. The paradox arises because we also imagine that the mental object in question is an item of no immediate significance to us—as indeed we must, if we assume that understanding it would mean interpreting it. I know that what I do is correct because I have been trained to think things like 'this action is correct' in response to such actions, or more likely, to prefer it when the expression is used in this kind of way—that I now think this way is my justification for claiming to mean one

thing rather than another. Even this is not quite right though: the justification I would offer, from my perspective, is simply that I understand and know how to use the word. The crucial point is that from my perspective the training was not the cultivation of dispositions to respond to certain stimuli in certain ways, but training in the recognition of correct actions. Expecting to find facts about correctness outside of a language-user's perspective will be futile, if the difference between whether something is meaning-devoid behaviour or correct action depends entirely on adoption of that perspective. And this is precisely what I think McDowell wishes to claim.

In light of this, the Big Dilemma seems to be generated by the assumption that we must be able to explain the coherence of our language practices without taking any language practices for granted—crucially denying ourselves the appeal to our engendered inclination to see 'basic-level' phenomena as meaningful. The 'engendered inclination' here, what I have been calling a perspective we occupy, is something like what McDowell calls "bedrock...where justifications have come to an end" (WFR 341). The idea is that, as competent language users, we see meaning in the world without having to look for it or check what it is. The Big Dilemma is apparently a result of an attempt to step outside our language practice and observe goings-on without getting involved in things like meaningful behaviour ourselves—but of course, any kind of analysis will require full use of our linguistic capacities, so the task is impossible.

The misguided project is motivated by the thought that there must be some fact about what happens from an extra-linguistic perspective, and that this would be the genuine fact of the matter, so we must try and characterise what that fact would be like. That conception of objective facts is characteristic of Problematic Realism—but it is also this kind of fact that anti-realists try to get hold of when insisting on the full behavioural manifestability of concepts. What both have in common is an attempt to explain things as they seem from our perspective, from a different perspective.

The problem might be brought out as follows. According to meaning-perspectivism, when someone plays a game of chess he partakes in a practice that has very clear rules, and failing to respect them will result in his being disqualified or otherwise penalised. When he plays the game properly, he respects the rules: but there is no separate action he undertakes of 'playing by the rules' as opposed to 'playing chess'. On the anti-realist model, we look at the actions he performs and conclude that we have come to agree that these actions are called legal moves in chess, but that the actions themselves are just as fairly construed as meaningless.

However, within the context of chess—the actual context of what is going on here—it is simply untrue that the actions are meaningless. ‘Moving a little piece of varnished, lathed wood away from oneself from a white-painted square to a black-painted square...’ and so on, is not the same as ‘conducting a King’s Pawn opening’, which would be described in terms of White’s first move, that the pawn rather than a knight is moved, and with respect to other chess terminology. However complex the behavioural description, if the actions are placed outside the context of a game of chess then we cannot expect to introduce chess terminology into the picture, even though someone who knows enough about chess will recognise that they correspond to the action called a King’s Pawn opening.

So how do we place these complex behavioural actions in the context of a game of chess, if the actions themselves do not require alteration? McDowell’s answer is that one must be party to the relevant practice, that of playing chess (or perhaps watching or judging games of chess). According to meaning-perspectivism, the actions take on this character not because the movements in question are different, but because they correspond to a move in a chess game from the perspective of someone who has been initiated into the practice of playing chess by means of the relevant training. The actions become moves in the game for us because we are trained to perceive them as such. This ability should, of course, be understood as the product of a kind of training in complex behavioural responses. But to identify the practice with behavioural responses of this kind is to fail to respect the fact that within the context of ‘playing chess’, we are no longer dealing with *mere* behaviour—and the context is established by means of a particular perspective.

The anti-realist wants to say that the game of chess is reducible to certain kinds of behaviour, and of course it can be analysed as such—but in undertaking that analysis we have stopped talking about chess. To call this an explanatory reduction is misleading because there is no identity between the two phenomena ‘behaving in this complex way’ and ‘playing chess’: perhaps there is some transcendent ‘thing’ they both correspond to when ‘viewed without any kind of perspective’, but that is certainly not what is meant by the claim that they are identical. The one is a complex analysis of the behaviour characteristic of the other, but it will be an *incomplete* characterisation of ‘playing chess’ so long as we neglect to include the point that one must be trained to adopt a particular perspective in order to get those concepts. And even including mention of this point precludes us from characterising ‘playing chess’ in a way that would be ‘fully manifest in behaviour’, because ‘playing chess’ will be manifest to an

observer who is not equipped to adopt that perspective. This means denying the legitimacy of the anti-realist's requirement that grasp of concepts be manifestable in this way.

According to meaning-perspectivism, searching for the action 'following the rules' alongside the action 'behaving in this complex way' is misguided in the same way that looking for an extra activity 'running quickly' alongside the activity 'running' is misguided. The two things happen concurrently, separable only with reference to two distinct perspectives on the action. The anti-realist strategy effectively commits us to imagining that we do not understand things we actually do understand, and trying to explain them without being privy to the understanding required in order to deploy the relevant concepts. Having depriving ourselves of the tools required, the endeavour is hopeless—but why should this surprise us? No one is expecting a cell biologist to find the phenomenon 'King's pawn opening' in his enquiries, because that is not the way that a cell biologist conceives of cell activity. This idea goes some way towards explaining why the characterisation of meaningful language use in basic level terms presents an insurmountable obstacle to finding meaning in actions: the picture is incomplete unless we already benefit from mastery of the practice in question, and thus, from the requisite perspective on those actions.

4.5 CONCLUSION

The sketch of meaning-perspectivism offered here does not go too far beyond what McDowell offers in *WFR*, and the most significant additions pertain to exploring what could be meant by the idea that training in the use of an expression is induction into a 'custom, practice or institution'. This is where the notion of 'perspective' came from, but there are plenty of other glosses on that central idea in the rule-following literature that practices are the solution to problems like Kripke's sceptical paradox, so I have no illusions that McDowell *must* be committed to meaning-perspectivism.

What is noticeable is that the Big Dilemma is quite well accounted for on this sketch, and all we really needed to acknowledge to avoid it was that language users are simply trained to apply their language to the world, and that a product of this training is that things are simply meaningful to them. This might seem slightly anti-climactic, but the valuable lesson here is that it is only possible to conduct a philosophical analysis of language if one has been trained in the language in question. Contrary to McDowell's ineffective Illusion Argument, though, the fact that the training is about forming

psychological and behavioural dispositions to which language can be explanatorily reduced does not make language meaningless.

McDowell aims, like his anti-realist counterparts, to reject Problematic Realism. In the account sketched above this translated as the idea that genuine facts are facts that rely on no particular perspective on things. The appropriate response, I suggest, is to accept that perspectives are necessary, but Wright takes this a step too far when he identifies correctness with consensus. As I have repeatedly emphasised, the fact that all human judgments are couched in acceptable use of language does not mean that our ordinary conception of objectivity requires revision. Wright's disastrous acceptance of CO illustrates this point very well.

I have consciously conducted this analysis in the controversial frame of the assumption that meaning is normative, a thesis that has come under fire in recent years. However, it seems to me that if there is any scope for defending and developing meaning-perspectivism then there might also be scope for preserving the thesis that meaning is normative. Glüer and Wikforss suggest that the normativity of meaning was supposed to be "a pre-theoretical requirement that any plausible theory of meaning has to respect" (2009, p.34), and I am interested by the prospect that meaning-perspectivism might lend substance to that claim without explicitly adopting the thesis that meaning is normative. If the training required to adopt a perspective on some expression had the effect of making us prefer correct application of it to incorrect application, then from within that perspective there might be scope for seeing correct applications as the ones we *should* make. That is a tantalising possibility indeed.

BIBLIOGRAPHY

- Blackburn, Simon (1981) 'Reply: Rule-Following and Moral Realism' in S. Holtzman & C. Leich (eds.) *Wittgenstein: to Follow a Rule*, pp. 163-187 (London: Routledge & Kegan Paul)
- Blackburn, Simon (2002) 'The Individual Strikes Back' in A. Miller & C. Wright (eds.) *Rule-Following and Meaning*, pp.28-44 (Durham: Acumen)
- Dummett, Michael (1981) Frege: Philosophy of Language, 2nd Edition (London: Duckworth)
- Dummett, Michael (2004) Truth and the Past (New York: Columbia UP)
- Goodman, Nelson (1983) Fact, fiction and forecast, 4th Edition (Cambridge, Mass. & London: Harvard UP)
- Glüer, Kathrin & Wikforss, Åsa (2009) 'Against Content Normativity' in *Mind* 118, pp.31-70
- Kripke, Saul (1982, 'K') Wittgenstein on Rules and Private Language (London: Blackwell)
- McDowell, John (1981) 'Non-Cognitivism and Rule-Following' in S. Holtzman & C. Leich (eds.) *Wittgenstein: to Follow a Rule*, pp. 141-162 (London: Routledge & Kegan Paul)
- McDowell, John (1984, 'WFR') 'Wittgenstein on Following a Rule', *Synthese* 58, no. 3, pp.325-363
- McDowell, John (1998, 'AEU') 'Anti-Realism and the Epistemology of Understanding' in *Meaning, Knowledge, and Reality*, pp. 314-343 (London: Harvard UP)
- McDowell, John (2002) Mind, Value, and Reality (London: Harvard UP)
- McDowell, John (2009a) 'Are Meaning, Understanding etc. Definite States?' in *The Engaged Intellect*, pp. 79-95 (London: Harvard UP)
- McDowell, John (2009b) 'How Not to Read *Philosophical Investigations*: Brandom's Wittgenstein' in *The Engaged Intellect*, pp. 96-111 (London: Harvard UP)

Rule-Following, Normativity & Objectivity

- Rosen, Gideon (1994) 'Objectivity and Modern Idealism: What is the Question?' in M. Michael & J. O'Leary-Hawthorne (eds.) *Philosophy in Mind*, pp. 277-319. (Kluwer)
- Wittgenstein, Ludwig (1989, 'PI') Philosophical Investigations tr. G. E. M. Anscombe (London: Blackwell)
- Wittgenstein, Ludwig (2007) The Blue and Brown Books (London: Blackwell)
- Wright, Crispin (1980a, 'W') Wittgenstein on the Foundations of Mathematics (London: Duckworth)
- Wright, Crispin (1980b) 'Realism, Truth-Value Links, Other Minds and the Past' in *Ratio* 22, no.2, 1980 (Oxford: Blackwell)
- Wright, Crispin (1989) 'Wittgenstein's Rule-following Considerations and the Central Project of Theoretical Linguistics' in A. George (ed.) *Reflections on Chomsky*, pp. 233-264 (Oxford: Blackwell)
- Zalabardo, José (2002) 'Kripke's Normativity Argument' in A. Miller & C. Wright (eds.) *Rule-Following and Meaning*, pp.274-293 (Durham: Acumen)

QUICK-REFERENCE GLOSSARY

I have labelled a large number of 'theses' in the text and refer back to them by label. The important ones are reproduced here for ease of reference.

- (C)** S understands the meaning of E \rightarrow S undertakes to use E only in a range of ways R whenever S uses E.
- (M1)** For any possible application *a* of an expression E, *a* is either objectively correct or objectively incorrect, and for any understanding of the meaning of E, *a* is either correct or incorrect relative to that understanding.³⁶
- (M2)** S's understanding of the meaning of E is proper \leftrightarrow S undertakes to apply E only in a range of ways R, where R is the range of correct ways of applying E.
- (M3)** S's understanding of the meaning of E is faulty \leftrightarrow S undertakes to apply E only in a range of ways R*, where R* is not the range of correct ways of applying E.
- (M4)** S undertakes to apply E only in a range of ways R \rightarrow S can accidentally apply E in a way falling outside R.
- (W1)** For any application *a* of an expression E, *a* is either correct or incorrect \leftrightarrow *a* is an actual (not merely possible) application of E.
- (CO)** (It is the majority view that P) \leftrightarrow P
- (MP1)** An expression E is meaningful \rightarrow there is some perspective from which E is meaningful.
- (MP2)** S understands the meaning of an expression E \rightarrow S occupies the perspective on E³⁷ required to find E meaningful.
- Logical Independence (LI)** For particular true statements it is either unnecessary or insufficient, or both, to meet our most refined criteria of acceptability in order to be true. (W, p.199)
- Minimal Contractual Picture** Understanding the meaning of an expression imposes some kind of constraints on one's future actions, amounting to a commitment to make only a certain kind of use of the expression.

³⁶ This is the refined version of M1 from the end of §1.3.

³⁷ i.e. 'conception of E'

Problematic Realism

A genuine fact must be a matter of the way things are in themselves, utterly independently of us. So a genuinely true judgment must be, at least potentially, an exercise of pure thought; if human nature is necessarily implicated in the very formulation of the judgment, that precludes our thinking of the corresponding fact as properly independent of us, and hence as a proper fact at all. (WFR 351)