OXFORD

Oughts and Thoughts

Rule-Following and the Normativity of Content

ANANDI HATTIANGADI

OUGHTS AND THOUGHTS
RULE-FOLLOWING AND THE NORMATIVITY OF CONTENT

This page intentionally left blank

Oughts and Thoughts

Rule-Following and the Normativity of Content

CLARENDON PRESS · OXFORD

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

© Anandi Hattiangadi 2007

The moral rights of the authors have been asserted
Database right Oxford University Press (maker)

First published 2007

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Typeset by Laserwords Private Limited, Chennai, India
Printed in Great Britain
on acid-free paper by
Biddles Ltd., King's Lynn, Norfolk

ISBN 978–0–19–921902–5

1 3 5 7 9 10 8 6 4 2

To Krister and Vikram

This page intentionally left blank

Preface

This book began its life as a doctoral dissertation, carried out under the supervision of Martin Kusch and Peter Lipton, at the Department of History and Philosophy of Science, Cambridge University. I would like to thank Martin for the animated disagreements which fuelled the arguments in the thesis and continue to fuel the arguments in the book, as well as his invariably thorough scrutiny of my work. I am grateful to Peter for his encouragement, his remarkable gift for making confused ideas lucid, and his ability to deliver devastating criticism gently. I would also like to thank Simon Blackburn and Bob Hale—who examined the dissertation—for their penetrating criticisms and suggestions for improvement. Research towards the dissertation was financially supported in the form of an Overseas Research Studentship, grants from the Cambridge Commonwealth Trust, The Social Sciences and Humanities Research Council of Canada and Trinity College, Cambridge. The process of turning the dissertation into a book manuscript was largely completed during the four glorious years while I was a Research Fellow at Trinity College, Cambridge.

Portions of this work draw upon material that has been published previously. Much of the material in Chapter 7 appears in 'Is Meaning Normative' (published in *Mind and Language*, 2002, pp. 220–40), and the discussion of Brandom in Chapter 6 draws on 'Making It Implicit: Brandom on Rule Following' (published in *Philosophy and Phenomenological Research*, 2003: 419–31). Earlier drafts of the material that makes up this book were presented at various seminars, including the Departmental Seminar at the Department of History and Philosophy of Science, Cambridge and the Moral Sciences Club, Cambridge. I am grateful to the audiences of both sessions for their incisive comments. I would also like to thank Sören Stenlund, who invited me to present my work at his seminar in philosophy of language at the Department of Philosophy, Uppsala University. Peter Pagin and Kathrin Glüer kindly invited me to discuss my work at the philosophy of language seminar at the Department of Philosophy, Stockholm University. I would like to thank them both for the discussion on that occasion, as well as for all of the discussions we have had over the years since then. They are wonderful allies to have—despite our broad agreement, they are very much alive to errors and potential objections. I would like to thank Martin Kusch for inviting me to the Symposium on The Normativity of Meaning, Wissenschaftskolleg zu Berlin. I profited tremendously from discussions with the participants at that workshop. Finally, I would like to thank John Broome for inviting me to present my work at his graduate seminar at the Faculty of Philosophy at Oxford University. I am grateful to the comments of participants of this session and for their helpful comments and suggestions.

Many friends and family members have contributed to this project with comments and criticism—or the odd question which has triggered major revision. These include Arif Ahmed, Gustaf Arrhenius, Anita Avramides, Stephen Butterfill, Erik Carlson, Anjan Chakravartty, Cathy Gere, Martin Gustafsson, Jane Heal, Henry Jackman, Carrie Jenkins, Neil Manson, Christina McLeish, Hugh Mellor, Alex Miller, Amartya Sen, Mark Sprevak, Åsa Wikforss, Tim Williamson, and my father—Jagdish Hattiangadi. Thanks to Marta Weiss for suggesting the cover illustration. I am grateful to Peter Momtchiloff, Jacqueline Baker, and Victoria Patton at Oxford University Press, and their anonymous readers, whose detailed comments made a great difference to the book.

Finally, I owe a tremendous debt of gratitude to Krister Bykvist, who has been the first person I turn to with any of my problems, philosophical or otherwise. It was Krister who first suggested that I look closely at the normativity thesis, and who helped to bring me up to speed on meta-ethics and deontic logic. Krister read and commented on countless drafts of this work, and has discussed both minor and major points at greater length than they no doubt deserve, at times when he probably would have preferred to be doing something else. He is a tremendous source of strength, support, encouragement, rigorous arguments, and ingenious ideas for getting out of tight spots; he is a wonderful father to our son and my closest friend.

Contents

1. **Introduction** 1

2. **The Sceptical Argument** 11

The sceptic's challenge 11
The scope of the sceptical argument 15
The sceptical argument 21
Conclusion 34

3. **Norms and Normativity** 37

From normativity to non-factualism 38
The first horn of the dilemma: the 'naturalistic fallacy' 39
The second horn of the dilemma: queerness 47
From norms to normativity 51
Norm-relativity 61
Conclusion 63

4. **Can we do Without Semantic Facts?** 65

The sceptical solution 66
The incoherence charge 69
Semantic non-factualism and deflationism 74
The social construction of meaning 87
Conclusion 103

5. **Reductionism** 105

Dispositionalism realism 105
Dispositions and conditionals 109
Extended dispositionalism 116
Success semantics 120
Teleosemantics 126
Asymmetric dependence 133
The causal theory of reference 141
Collectivism 144
Conclusion 149

6. Anti-Reductionism 151

The Capacity View 151
Intensionalism 153
Extension-determining intentions 158
Implicit rule-following 161
Bedrock norms 168
Internal relations 172
Conclusion 176

7. Is Meaning Normative? 179

The Normativity Principle 179
Essentially semantic categorical oughts 181
Mistakes and lies 185
Prima facie obligations 188
The norm of truth 190
Communication and community 193
Contractual commitments 197
Rules 201
Conclusion 206

8. Conclusion 209

Bibliography 213
Index 219

1

Introduction

Our practice of ascribing meanings to people's utterances and contents to their beliefs is commonplace. There is even a technical term to describe such discourse: it is called 'gossip'. And whether malicious or not, gossip about what people say or believe purports to describe, truly or falsely, what people say or believe. Moreover, we very often care deeply what people say or believe, and what they mean by their words. To take an example from close to home, some time ago my son, Vikram, broke out in a terrible rash, so I took him to see his doctor, who said that Vikram had chicken pox. That is, the doctor uttered the sentence 'Vikram has chicken pox', and I naturally assumed, when the doctor uttered this sentence, that he meant 'Vikram' to refer to Vikram (as opposed to Adam or Orlando) and 'chicken pox' to refer to chicken pox and only chicken pox—not to meningitis, or hepatitis, or any number of other diseases. Given my assumptions about what the doctor's words meant, I took the sentence 'Vikram has chicken pox' to mean *that Vikram has chicken pox*, and not *that Vikram has meningitis* or *that Adam has hepatitis* or anything else.[1] Given its meaning, the sentence the doctor uttered is true if and only if Vikram had chicken pox at the time. Since I took the doctor to be sincere and reliable, my assumptions about what the doctor meant by his words had an effect on how I subsequently acted. Had I thought that what the doctor had said was *that Vikram has meningitis*, I would have rushed Vikram to the hospital.

In this report of my conversation with the doctor, I used a 'that-clause' to specify what the doctor meant. I said that what he meant was *that Vikram has chicken pox*. To specify the meaning of a sentence in this way is, in effect, to specify its truth condition: to say that 'Vikram has chicken pox' means *that Vikram has chicken pox* is in effect to say that the sentence is true if and only if Vikram has chicken pox. Similarly, I specified the meanings of the doctor's words by giving their correctness conditions. For instance, I said that 'Vikram' refers to Vikram, which is to say that it correctly applies to Vikram and only Vikram. Furthermore, my understanding of the truth condition of the doctor's sentence (together with my assumption that the sentence was likely to be true) led me to act as I did. All of this lends support to one of the dominant traditions in the philosophy of language and mind, according to which correctness conditions and truth conditions play an essential role in the theory of meaning and understanding. I call this position 'semantic realism'.[2]

---

[1] I will employ the convention of using italics when I specify meanings or contents.

The semantic realist is someone who holds that many of the assumptions I made in the course of my visit to the doctor were literally true: the doctor did use 'chicken pox' to refer to chicken pox and only chicken pox, and when he said 'Vikram has chicken pox', he took this sentence to mean *that Vikram has chicken pox*, which is true if and only if Vikram had chicken pox at the time. More generally, the semantic realist holds that what it is to understand the meaning of a declarative sentence is to grasp its truth conditions, and what it is to understand the meaning of a word is to grasp its correctness conditions. Furthermore, the semantic realist typically holds that our ascriptions of meaning and truth conditions are themselves capable of truth or falsity. There is a 'fact of the matter' whether the doctor meant 'chicken pox' to refer to chicken pox, so there is a 'fact of the matter' whether my ascription of meaning to the doctor's utterance is true or false.

Intuitive though semantic realism might seem, it has been subject to a powerful sceptical argument. In his influential elaboration of Wittgenstein's rule-following considerations, Saul Kripke argues that so long as we assume semantic realism, it will turn out that there is 'no fact about me that distinguishes between my meaning [something] … and my meaning nothing at all',[3] and hence that 'sentences attributing meaning and intention are themselves meaningless.'[4] Kripke challenges the semantic realist to come up with an account of what makes it the case that someone means something by any word, such as 'chicken pox'. In particular, he challenges us to cite the facts that make it true that I mean *chicken pox* by 'chicken pox', such that 'chicken pox', given what I mean by it, applies correctly to chicken pox and only chicken pox.

---

[2] Cf. Dummett 1978; Field 1994; Wilson 1998. Note that on Dummett's (1978) characterisation of semantic realism, truth-conditions are conceived of as potentially evidence-transcendent. I will not primarily be concerned with this aspect of semantic realism.

[3] Kripke 1982, p. 21.

[4] Ibid., p. 79.

Any adequate fact must be capable of ruling out the sceptical possibility that I really mean *schicken pox* by 'chicken pox', such that 'chicken pox' applies correctly to chicken pox until the present time, or meningitis thereafter. He maintains that there is a basic condition that any account of what I mean must satisfy: it must capture the normativity of meaning. Whatever constitutes my meaning *chicken pox* as opposed to *schicken pox* must imply that I ought to apply 'chicken pox' to all and only cases of chicken pox—otherwise, some sceptical hypothesis would have equal claim to truth. Kripke argues that no theory can meet this apparently intuitive constraint. The result is a scepticism about meaning that is not only radical, but contagious: although formulated in terms of linguistic meaning, the sceptical conclusion extends to the content of mental representations as well.

Scepticism about meaning is outrageous, even *prima facie* self-refuting: if the sceptical conclusion is true, then it is itself meaningless, and if the sceptical conclusion is meaningless, then it cannot be true.[5] How can Kripke even purport to conclude something as nonsensical as this? The only way we can make sense of Kripke's argument is to see it as a *reductio* of semantic realism.[6] The sceptical argument purports to show that semantic realism implies the paradox that '[t]here can be no such thing as meaning anything by any word.'[7] To avoid this paradoxical conclusion, Kripke urges us to reject one central tenet of realism: the idea that the meaning of a word can be given by its correctness conditions; that the expression 'chicken pox' refers to chicken pox and only chicken pox. In place of realism, Kripke suggests what he calls a 'sceptical solution' to the sceptical problem. On this view, whether some use of an expression can be called 'correct' on some occasion does not depend on what it means, or on what its correctness conditions are, but on whether others in the linguistic community would agree in its use. The upshot of this alternative picture of meaning, Kripke claims, is Wittgenstein's famous argument against the possibility of a private language. Since there is no fact of the matter whether I mean 'chicken pox' to apply correctly to chicken pox and only chicken pox, I cannot be said to mean *chicken pox* by 'chicken pox' if I am 'considered in isolation,' that is in the absence of any comparison between my uses of 'chicken pox' and someone else's. However, if Jones judges that I use 'chicken pox' as he does, then he is entitled to say that I mean *chicken*

---

[5] Cf. Boghossian 1990; Wright 1984.

[6] Cf. Soames 1998*a*; Wilson 1998.

[7] Kripke 1982, p. 55.

pox by 'chicken pox'. Since we are only entitled to ascribe meanings on the basis of agreement in use, no one can be said to mean something by a word independently of any such agreement. Thus, there can be no such thing as a 'private language'.

Kripke's book initiated a discussion that continues at a furious pace over twenty years after its publication. This is no wonder. Kripke discovered in Wittgenstein a devastating sceptical argument against our intuitive picture of meaning, which leads him to question the reality, determinacy, and privacy of meaning. While Kripke's sceptical conclusion is bizarre, his argument is both lucid and powerful. And although sceptical arguments against semantic realism have dominated twentieth-century philosophy of language and mind, Kripke's argument stands out as being the most ambitious and comprehensive. A closely related view is W. V. O. Quine's famous argument that translation is indeterminate and reference inscrutable.[8] With these slogans, Quine meant that for any translation of a foreign speaker's sentences into English, there will always be an empirically equivalent but incompatible translation manual. Far from an ordinary case of under-determination, Quine held that in linguistic ascription, there was no determinate meaning to translate, no determinate reference to scrute. The conclusion is arguably as radical as Kripke's. However, unlike Quine's argument, Kripke's makes no empiricist assumptions. Kripke purports to consider *any* fact that might constitute what someone means—even those accessible only to the mind of an omniscient God—and finds that there can be no fact that constitutes what someone means. He purports to rule out both reductive theories—which take true statements about what people mean to be true in virtue of non-semantic, non-intentional facts—as well as anti-reductive theories—which take meaning facts to be *sui generis* and irreducible. Kripke's argument has a breathtaking scope, and if it succeeds, utterly devastates the intuitive view.

Another reason for the widespread interest in Kripke's discussion no doubt has to do with its pedigree. Indeed, some commentators have been primarily concerned with the accuracy and scope of Kripke's interpretation of Wittgenstein.[9] However, Kripke does not purport to give a comprehensive or systematic interpretation of Wittgenstein's writings—and the consensus seems to be that he does not succeed inadvertently. Instead, he suggests that his book 'should be thought of as expounding neither "Wittgenstein's" argument nor "Kripke's": rather Wittgenstein's argument as it struck Kripke, as it presented a problem for him.'[10] Indeed, the question whether Kripke correctly interprets Wittgenstein is quite irrelevant to the force and interest of the sceptical argument that Kripke puts forward, and irrespective of its exegetical accuracy, Kripke's argument has proved to be of enduring philosophical interest in its own right.

---

[8] Quine 1953, 1969.

[9] Wittgenstein's rule-following considerations can be found principally in Wittgenstein 1953 and 1956.

This book defends semantic realism against Kripke's sceptical attack. According to the semantic realist, to understand the meaning of a word (mental representation) is to know its correctness conditions, and that to understand the meaning of a sentence is to know its truth conditions. Semantic realism, on this definition, is compatible with a variety of metaphysical theories of what grasp of correctness conditions or truth conditions consist in. That is, a semantic realist could also be a semantic naturalist, who claims that what makes it true that I grasp the meaning of a word are ordinary 'natural' facts, which are ultimately physical, causal, or functional. Alternatively, a semantic realist could be an anti-reductionist about semantic facts, holding that semantic facts are *sui generis* and irreducible. My claim is that Kripke's argument against specific 'metaphysical' theories of what constitutes meaning, although powerful, ultimately fails to achieve full generality and *a priori* status. Hence, the sceptic is unable to show that semantic realism leads to the paradoxical conclusion that there is no such thing as meaning anything by any word. Moreover, I argue, semantic realism is indispensable. The denial of semantic realism is either self-refuting, or presupposes semantic realism. As a consequence, we have a positive reason to remain committed to semantic realism, even in the face of the sceptical argument Kripke finds in Wittgenstein.

Given the vast literature on Kripke's sceptical argument, it may seem as though no stone has been left unturned. Perhaps yet another contribution to this discussion is unwarranted. However, there is one stone that has been left relatively unexamined. This is the thesis that meaning is normative.[11] Granted the assumption that meaning is normative, I shall argue, the sceptic is able to marshal *a priori* considerations against all possible substantive theories of meaning—not merely those that come directly under attack, nor even just those that have been hitherto presented. However, without the thesis that meaning is normative, the sceptical argument amounts to no more than criticisms of a few theories of what constitutes meaning. Even if we do not now have an adequate account of what constitutes meaning, the sceptic is not entitled to conclude that there is no fact of the matter what anyone means by any word. For, if we are allowed to consider facts accessible only to an omniscient God, the fact that constitutes what someone means may well be a fact of which we are not currently aware. If the sceptic is to convince us that all ascriptions of meaning and belief are neither true nor false, he needs to do more than simply criticise the current theories of what makes it the case that someone means something by a word. Since the thesis that meaning is normative provides the sceptic with an *a priori* argument against all theories, it lends the sceptical argument sufficient force to show that our practice of ascribing meanings and beliefs is entirely without basis in fact.

---

[10] Kripke 1982, p. 5.

[11] Critics of the normativity thesis include Dretske 2000; Glüer 1999*a*, 1999*b*, 2001; Glüer and Pagin 1999; Hattiangadi 2002; Papineau 1999; Wikforss 2001; Wilson 1994.

Kripke's sceptical argument can have full generality and *a priori* status, so long as we grant the assumption that meaning is normative. We are forced to conclude that semantic realism leads to the paradox that there can be no such thing as meaning anything by any word. Kripke's suggestion is that we avoid the paradox by rejecting semantic realism and embracing, instead, a 'sceptical solution' to the sceptical paradox. However, I argue that this sceptical solution is irremediably incoherent. It is *prima facie* incoherent because if the sceptical conclusion is true, then it is meaningless, and if it is meaningless, it cannot be true. I will argue further that any attempt to rehabilitate our practice of ascribing meaning and content either fails to do so, or must presuppose semantic realism. Rejecting semantic realism is therefore not a legitimate option. The sceptical argument must go wrong somewhere. The question is, where?

The answer lies in the examination of the thesis that meaning is normative. The normativity thesis, which plays such a decisive role in the sceptical argument, is ambiguous, and this ambiguity leads to the sceptic's undoing. In order to advance the *a priori* argument against all possible theories, the sceptic must assume that meaning is normative in a strong sense, that is as inherently motivating or prescriptive. To say that meaning is normative in this strong sense is to say that what a speaker means determines which uses of an expression she ought to make, where this 'ought' is understood to be 'categorical' in that it is not contingent on the agent's desires or ends. However, there is a weaker interpretation of the normativity thesis, according to which meaning is ‘norm-relative’ in the sense that there is a norm which determines which uses of an expression are correct and which incorrect. I call the first principle, *Normativity*, and the second, *Norm-Relativity*, and argue that the distinction between these two principles is the crack in the keystone of the sceptical argument. *Norm-Relativity* and *Normativity* are not equivalent—the correctness of some use of an expression does not imply a categorical ‘ought’. And while *Norm-Relativity* is intuitive, probably true, it is anodyne—assuming only *Norm-Relativity*, the sceptic cannot rule out all theories of meaning *a priori*. In contrast, although *Normativity* would rule out all possible theories of meaning, it is untenable. Once we repudiate *Normativity*, the sceptic can no longer show that there can be no fact of the matter what anybody means.

Thus, I will conclude that we have no reason to believe that, if we assume semantic realism, there is no fact of the matter what anybody means. But do we have any positive reason to believe that semantic realism is true? If we had a positive account of what constitutes meaning—whether naturalistic or non-naturalistic—this would give us positive reason to embrace semantic realism. However, I argue that we do not have an adequate account. Although I cannot exhaustively consider every proposal on offer any more than Kripke could, I consider many of the most plausible proposals and argue that none of them succeeds. Nevertheless, I maintain that we do have a positive reason to endorse semantic realism: the attempt to do without semantic realism leads to self-refutation. Thus, even though we cannot, now, cite the facts that make it true that I mean *chicken pox* by ‘chicken pox’, we have very good reason to believe that there is a fact of the matter whether I mean *chicken pox* by ‘chicken pox’. Furthermore, given the assumption that there *is* a fact of the matter what I mean, nothing the meaning sceptic says undermines the natural assumption that I *know* what I mean. It is not necessary for me to be able to cite the fact that makes it the case that I mean *chicken pox* by ‘chicken pox’ in order to know that I mean *chicken pox* by ‘chicken pox’. Similarly, even if biology is ultimately reducible to physics, it is not necessary for me to know the truths of biology that I be able to cite the physical facts that make them true. The upshot, then, is that scepticism about meaning is indefensible and leads inevitably to incoherence, and thus that the intuitive view that we can ascribe contents to our beliefs and utterances, that such ascriptions can be true or false, and that we often know the contents of our own minds, is not touched by the sceptical argument that Kripke finds in Wittgenstein.

The structure of the book follows the structure of the argument presented above. In the next chapter, I will present the sceptical argument as Kripke formulates it. That argument is deficient in a number of respects, and I will suggest how the thesis that meaning is normative—to which Kripke subscribes—can remove the deficiencies. In Chapter 3, I lay out the meta-ethical arguments and assumptions Kripke would need to make in order to remove the gaps in the sceptical argument as he presents it. I then turn to a more detailed account of the thesis that meaning is normative and its role in the sceptical argument. Here, I argue that the assumption that meaning is normative does not follow directly from semantic realism, but from the assumption that understanding the meaning of a word is analogous to following a rule for its correct use. This, however, gives rise to two alternative interpretations of the claim that meaning is normative: *Norm-Relativity* and *Normativity*. I then argue that if meaning is *normative*, arguments commonly made in meta-ethics with regard to moral statements, can be applied, *mutatis mutandis*, to meaning statements. These arguments can be made against both reductive and non-reductive accounts of the facts that putatively make meaning ascriptions true. I conclude that if the sceptic is entitled to the thesis that meaning is normative, and if he is entitled to certain meta-ethical claims, he seems to be able to argue, *a priori*, that there is no fact of the matter what anybody means. In contrast, I argue, if meaning is merely norm-relative, no such disastrous conclusion follows.

In Chapter 4, I turn to Kripke's sceptical solution. Supposing that the sceptical argument is sound, what prospect is there of a sceptical solution, that is one that embraces the conclusion that there is no fact of the matter what we mean? I will argue that the 'no fact thesis' is irremediably incoherent, since, if we reject semantic realism, no statement can be true, or justified, even in the weakest sense. Thus, there is no hope for a 'sceptical solution' which purports to show that although semantic realism is false, our ascriptions of meaning and content are nevertheless legitimate. Since the appearance of paradox in the sceptical conclusion cannot be removed, I provisionally conclude that the argument must falter somewhere.

In Chapter 5, I turn to more sophisticated reductionist responses to the sceptical argument—that is those which seek to find the fact that constitutes someone's meaning something by a word among the causal, physical, or functional facts. I consider a wide variety of the most compelling reductive theories that have been presented in response to Kripke's sceptic and argue that each of them fails. That is, each theory fails to find facts that uniquely determine that I mean *chicken pox* rather than *schicken pox* by ‘chicken pox’. In Chapter 6, I turn to anti-reductionist theories and argue that each of these, similarly, fails. Anti-reductionists, who maintain that there are semantic facts over and above the causal, physical, and functional facts, seem equally unable to uniquely determine that I mean *chicken pox* rather than *schicken pox* by ‘chicken pox’. The problems that beset reductionists and anti-reductionists are different, but yield the same, unfortunate result. So, the question remains: where does the sceptic go wrong?

In the following chapter I argue that the sceptic goes wrong in assuming that meaning is normative. In Chapter 7, by considering, and rejecting all of the most compelling reasons one might have for believing *Normativity*, I argue that it is untenable. Since it is *Normativity*, but not *Norm-Relativity* that engages the meta-ethical arguments against meaning facts, by rejecting *Normativity*, I show that the sceptic’s only hope of a wide-ranging *a priori* argument against all possible candidate meaning facts fails. Thus, I conclude that despite the failure of both reductionists and anti-reductionists to find the facts that constitute meaning, we have no reason to suppose that there is no fact of the matter what we mean.

This page intentionally left blank

2

The Sceptical Argument

Kripke's sceptic assumes semantic realism, and argues that it leads to the paradox that there is no such thing as meaning anything by any word. To show that there is no such thing as meaning anything by any word, Kripke must either exhaustively rule out all theories of what meaning consists in, or find an *a priori* justification for doing so. Unfortunately, Kripke's explicit arguments fall short of meeting either of these requirements. Yet we should not conclude too quickly that the argument fails. The resources for the requisite arguments can be developed, in a sceptical spirit, from some of Kripke's hints and suggestions. So, perhaps the present chapter should be thought of as expounding neither Kripke's argument nor mine: rather Kripke's argument as it struck me, as it presented a problem for me.

THE SCEPTIC'S CHALLENGE

Kripke illustrates the sceptical problem with the help of a thought experiment. He asks us to imagine that he is asked to compute a sum he has never computed before: for simplicity, he suggests the sum of 68 and 57. After a moment's thought, Kripke gives the answer '125'. He is confident that this is the correct answer both in the mathematical sense and in what Kripke calls the 'metalinguistic sense'. That is, given what Kripke means by 'plus', and given that he meant to apply the *addition* function to the arguments 57 and 68, '125' is the answer that accords with what Kripke meant. Now, Kripke asks us to imagine a bizarre sceptic who comes along and questions his use of 'plus' in this metalinguistic sense. The sceptic suggests that what Kripke means by 'plus' is not *addition* but *quaddition*. The *quaddition* function (symbolised by '$\oplus$' below) is defined as follows:

For any numbers m, n,

$$m \oplus n = m + n, \text{ if } m, n < 57$$

$$m \oplus n = 5 \text{ otherwise}$$

If Kripke means *addition* by '+', '125' will be the correct answer, but if Kripke means *quaddition*, the correct answer will be '5'. Kripke, of course, is quietly confident that he really means *addition*, not *quaddition* by 'plus'. However, the sceptic says that if it is *true* that Kripke means *addition*, and not *quaddition* by 'plus', then it must be possible to cite the fact that makes it true. The sceptical challenge is to find that fact.

By hypothesis, Kripke has never before computed sums whose arguments exceed 57, so the computation he makes in this case is entirely novel. Although this assumption is implausible, particularly in Kripke's case, the infinitude of the addition function guarantees that some sums exceed his past experience. The point is that if Kripke has never computed sums whose arguments exceed 57, he cannot cite his past behaviour as direct evidence for his claim that '125' accords with what he means and has meant all along by 'plus'. The wily sceptic can always argue that the hypothesis that Kripke meant *quus* all along is consistent with his past use of the word 'plus'. And, the sceptic goes on, if Kripke did mean *quus* all along, and if he is to accord with the meaning he has always given to the word, he should now say that $68 + 57 = 5$.

The sceptical problem is designed to put pressure on semantic realism.[1] I have given a rough formulation of semantic realism in the introduction, but to be more precise, semantic realism comprises at least the following three theses:

1. What someone means or understands by a word (mental representation) can be given by the correctness conditions of the word (mental representation) as it is understood.
2. What someone means or understands by a sentence (mental representation) can be given by the truth conditions of the sentence (mental representation) as it is understood.

---

[1] This is not to say that it puts pressure exclusively on semantic realism. As Boghossian (1989) points out, the sceptical argument tells against any view according to which the meaning of a representation can be given by a correctness condition, whether this is a truth condition or a condition for warranted assertion.

3. Ascriptions of meaning to linguistic utterances and mental states are 'factual', that is, they can be either true or false, and when true, are true in virtue of objective (i.e. judgement independent) facts.

Kripke indicates clearly that semantic realism bears the brunt of the sceptical argument. For instance, he says that the Wittgenstein of the *Philosophical Investigations*, who mounts the sceptical argument, is criticising the Wittgenstein of the *Tractatus*, who accepted a variant of what I am calling semantic realism. Wittgenstein's early view is characterised by Kripke as follows:

> The simplest, most basic idea of the *Tractatus* can hardly be dismissed: a declarative sentence gets its meaning by virtue of its *truth conditions*, by virtue of its correspondence to facts that must obtain if it is true. … So stated, the *Tractatus* picture of the meaning of declarative sentences may seem not only natural, but even tautological.[2]

The Tractarian thought is that the meaning of a declarative sentence is given by the conditions under which it is true.[3] 'Grass is green' is true if and only if grass is green; *that grass is green* is what the sentence 'grass is green' means. It may be difficult to see how this idea could be the target of Kripke's sceptical argument, since his argument focuses on the meaning of the sub-sentential expression, 'plus', throughout, and the meanings of sub-sentential expressions, such as 'grass' and 'green' cannot be given by their truth conditions—because 'grass', on its own, is neither true nor false. However, it is possible to give an analogous analysis of the meanings of sub-sentential expressions by looking at their semantic relations to the world: 'grass' refers to grass, and nothing else, 'green' is true of all and only green things. Sub-sentential expressions, such as 'grass' and 'green' do not have truth conditions, but correctness conditions. And if we add the assumption that truth conditions of sentences are a function of the correctness conditions of the words in them, then it is obvious that the truth conditional picture of the meanings of sentences bears the brunt of the sceptical argument, albeit only indirectly.

Kripke's sceptical argument puts direct pressure on the realist thesis that what someone means or understands by a word can be given by its correctness conditions—thesis number 1, above. In the *plus/quus* contrast, the correct uses of 'plus' converge for sums with values less than 57 and diverge for all values greater than 57. Because the *addition* and *quaddition* functions are both infinite—they are defined for all pairs of positive integers—they each give rise to an infinite list of correct uses of ‘plus’. Moreover, any account that purports to refute the sceptic must be capable of ruling out not only the *quus* hypothesis, but also all other such ludicrous hypotheses (diverging for values greater than the speaker could grasp, for instance). In order to do so, any account that can refute the sceptic must show how what constitutes someone’s meaning *addition* by ‘plus’ determines the infinite list of correct uses of the word ‘plus’. Any account that failed to yield the full list of correct uses would simply leave open some sceptical alternative.

---

[2] Kripke 1982, p. 72.

[3] Wittgenstein 1922.

The sceptical problem is not peculiar to the case of mathematics or mathematical terms. It might be tempting to think that the infinitude of the addition function poses a peculiar problem that would not arise if Kripke had chosen a different example. However, even for such words as ‘elephant’ and ‘green’, the list of all the possible correct uses is indefinitely large, and certainly exceeds the number of uses that have already been made. There is an infinite number of sentences, for instance, in which ‘elephant’ can be used, and there is an infinite number of possible situations in which a sentence containing ‘elephant’ would be correct. The important point is that there are always uses of a given term that a speaker has yet to make. Given that this is so, it will be possible to construct sceptical alternatives for any term. For example, suppose that you have only ever seen elephants in zoos. In that case, the sceptic can suggest that what you really mean by ‘elephant’ is *schmelephant*, which refers only to elephants in zoos. If you happen to find an elephant in your back garden, the question whether it is correct to call it an elephant will depend on whether you mean *elephant* or *schmelephant* by ‘elephant’.

More unexpectedly, perhaps, the sceptical problem arises for proper names, which refer to just one individual. For instance, according to the semantic realist, ‘Socrates’ refers to Socrates; it applies correctly to Socrates and only Socrates. However, if you have only ever met Socrates in Athens, then the uses you have made of ‘Socrates’ in the past are consistent with your having meant it to apply only to Socrates in Athens. Whatever it is that constitutes your denoting Socrates by ‘Socrates’ should determine that the name applies correctly to Socrates, no matter where he is. The question invariably is this: what determines whether some use of an expression is correct or incorrect in a novel instance? And this question can be raised of any meaningful expression whatsoever.

Indeed, mental representations are equally susceptible to this sort of sceptical attack. Although Kripke formulates his argument primarily in terms of linguistic meaning, it is clear that it is the notion of representational content as such that is at stake. If I have a mental representation with the content *square*, according to the semantic realist, my mental representation applies correctly to all and only squares. The list of all the possible correct applications of the mental representation 'square' is infinite, and there are no doubt some applications of 'square' which I have yet to make. Thus, the sceptical argument Kripke raises against linguistic meaning can be applied, *mutatis mutandis* to the content of mental representations as well. The reason is that, for the semantic realist, the meaning or content of any representation (mental or linguistic) can be given by its correctness conditions, and to understand any linguistic expression, or to have a particular mental representation, is to know, at least tacitly, the conditions under which it correctly applies. If Kripke's argument shows that nothing makes it the case that a word such as 'plus' has the correctness conditions we take it to have, then it will equally show that nothing makes it the case that any mental representation has the correctness conditions we take it to have. This is why most people take Kripke's arguments to attack the more general idea that any representation, whether mental or linguistic, has conditions of correct application.[4]

THE SCOPE OF THE SCEPTICAL ARGUMENT

Kripke's sceptic purports to prove a negative—that there is *no* fact of the matter what anybody means by any word. He maintains that, given *all* the facts, there will nevertheless be no way to refute a sceptical alternative interpretation of what anyone means. The strength of the sceptical conclusion is thus directly proportional to how liberal a domain of facts Kripke admits. And Kripke maintains that he is as liberal as can be:

> [An] important rule of the game is that there are no limitations, in particular, no *behaviourist* limitations, on the facts that may be cited to answer the sceptic. The evidence is not to be confined to that available to an external observer, who can observe my overt behaviour but not my internal mental state. It would be interesting if nothing in my external behaviour could show whether I meant plus or quus, but something about my inner state could. But the problem here is more radical. … So whatever 'looking into my mind' may be, the sceptic

---

[4] Cf. Boghossian 1989.

asserts that even if God were to do it, he still could not determine that I meant addition by 'plus'.[5]

Kripke explicitly compares his sceptical argument with a similar argument made famously by Quine.[6] Like Kripke's Wittgenstein, Quine argued that there is no fact to be found that will decide between a number of competing translation manuals for a speaker's utterances—with regard to this case, Quine would say that there is no fact that could arbitrate between the translation of '+' as 'plus' or 'quus'. However, Quine started from an empiricist position, according to which the only admissible evidence consists of a speaker's behaviour and dispositions to behave under observable circumstances. Thus, more precisely, Quine argued that given only evidence of a speaker's dispositions to overt behaviour under publicly observable circumstances, it is indeterminate what the speaker should be said to mean. For instance, Quine said:

> For naturalism the question whether two expressions are alike or unlike in meaning has no determinate answer, known or unknown, except insofar as the answer is settled in principle by people's speech dispositions, known or unknown. If by these standards there are indeterminate cases, so much the worse for the terminology of meaning and likeness of meaning.[7]

Quine imagined a linguist encountering a foreigner who speaks a language of which the linguist has no prior knowledge. On Quine's view, the linguist must begin by first making observations of the speaker's utterances and the conditions under which those utterances are made. Based on these observations, the linguist devises a translation manual that correlates sentences of the foreign language to sentences of the linguist's home language (which Quine assumes to be English). For example, if the linguist observes that the foreigner repeatedly utters the sentence 'Gavagai' in the conspicuous presence of rabbits, the linguist will correlate 'Gavagai' with the English 'Rabbit'. However sensible this may seem, Quine argues that this translation is under-determined by the totality of evidence. For, there is an alternate translation manual, which correlates 'Gavagai' with 'un-detached rabbit part' (along with relevant changes elsewhere in the translation), and which is both empirically equivalent to the intuitive translation manual and incompatible with it. Quine claims that no matter how much behavioural evidence the linguist gathers, it will be insufficient to determine that one translation manual is correct, while ruling out all the others. The question of what

---

[5] Kripke 1982, p. 14.

[6] Quine 1960, 1969.

[7] Quine 1969, p. 29.

'Gavagai' translates to in English, Quine claims, 'remains undecided by the totality of human dispositions to verbal behaviour. It is indeterminate in principle; there is no fact of the matter.'[8]

Quine's empiricist assumption is evident in his claim that if a choice between two translation manuals is undecided by dispositions to observable behaviour, it is indeterminate in principle which translation manual is correct.[9] This follows only if nothing other than dispositions to observable behaviour could determine which translation manual is correct. Quine defends this assumption on the grounds that a child learning a first language has only the linguistic behaviour of adults to go by.[10] Be that as it may, it does not follow that the *correctness* of a given translation manual must be determined exclusively by dispositions to verbal behaviour. For one thing, it could be argued that children are able to learn a language not by induction from the observable behaviour of others, but because they have innate mechanisms which come into play.[11] Second, one might reject the assumption that the only admissible evidence consists in a speaker's dispositions to *verbal* behaviour under publicly observable circumstances. Donald Davidson, for instance, has argued that we go by non-verbal behaviour as well as verbal behaviour.[12] Third, one might deny Quine's assumption that the correctness of a translation manual depends only on dispositions to behaviour, verbal or otherwise. One might think, for instance, that other 'natural' facts determine meaning, even if dispositions are insufficient. Quine does not attempt to rule out non-dispositional, but equally naturalistic facts that might constitute what someone means. Finally, one might think that no natural facts determine which of several translation manuals is correct, but non-natural facts about a speaker's mind do. Quine would claim that such an appeal to occult properties is unscientific—it is a myth that meanings are like objects on display in a museum. However, the rejection of this view is based, ultimately, on a commitment to empiricism and a naturalistic worldview and so, arguably begs the question against the non-naturalist.

Kripke's form of scepticism is intended to be even more radical than Quine's. Kripke's Wittgenstein agrees with Quine that meaning ascriptions cannot be justified by the totality of evidence concerning a speaker's dispositions to behaviour. However, Kripke purports to consider and rule out *anything* that could be thought to constitute a speaker's meaning something by a word, including non-natural facts about a speaker's mind, even facts that are accessible only to the 'mind of God'. As Kripke puts the contrast:

---

[8] Quine 1969, p. 38.

[9] Quine even accepts that his conclusions follow from his behaviourism. See Quine 1987, p. 5.

[10] Quine 1993, p. 38.

[11] See, e.g., Chomsky 1980.

[12] Davidson 1984, pp. 147–8.

> Quine bases his argument from the outset on behaviouristic premises. He would never emphasize introspective thought experiments in the way Wittgenstein does, and he does not think of views that posit a private inner world as in need of elaborate refutation. For Quine, the untenability of such views should be obvious to anyone who adopts a modern scientific outlook.[13]

Kripke allows us to cite any fact whatsoever in response to the sceptic; he does not, like Quine, assume behaviourism. Nevertheless, some of the things that Kripke says suggest to some that he does restrict the domain of legitimate facts to those about the individual. For example, he says that an answer to the sceptic must 'give an account of what fact it is (*about my mental state*) that constitutes my meaning plus, not quus.'[14] Later, he says that 'the sceptic holds that no fact *about my past history—nothing that was ever in my mind, or in my external behaviour*—establishes that I meant plus rather than quus'.[15] Some people have taken these passages to suggest that Kripke really meant the sceptical argument to attack only an internalist picture of meaning and content, a picture according to which *intrinsic* facts about a person alone could constitute meaning.[16] This is clearly not the case, however. When he is summarising the results of the sceptical argument, Kripke concludes that 'there is no such fact, no such condition in either the "internal" or the "external" world' that determines what someone means.[17] Where Kripke suggests that facts 'about me' are at issue, it is unclear whether this should be taken to mean that only *intrinsic* facts about me are at issue; extrinsic facts are just as much facts 'about me', even if they are facts about me in relation to other things. Even the externalist must explain what *I* mean by adverting to facts about *my* relation to my linguistic community or the external world.

It also sometimes seems as though Kripke restricts the domain to *past* facts, to facts about what someone did or said in the past.[18] The way that Kripke formulates the sceptical problem in the first instance gives the impression that this is so. For instance, Kripke says that we can characterise the sceptical problem as follows: 'When asked for the answer to "68 + 57", I unhesitatingly and automatically produced "125", but it would seem that if previously I never performed this computation explicitly I might just as well have answered "5".'[19] However, he later jettisons this way of presenting the problem, on the grounds that if the problem arises in the past, it arises in the present also. He says: 'when we initially presented the paradox, we perforce used language, taking present meanings for granted. Now we see, as we expected, that this provisional concession was indeed fictive. There is no fact as to what I mean by 'plus', or any other word at any time.'[20] Furthermore, it is clear from Kripke's lengthy discussion of dispositionalism that we are not even implicitly restricted to considering only *past* facts. The fact that I am now disposed to add is hardly a *past* fact, since my past dispositions may be different from my present ones. According to the dispositionalist, what makes it the case that I mean *addition* by 'plus' now has little or nothing to do with anything that happened in the past, but with what I would do in a host of possible situations, given my current dispositional make up. Since Kripke takes this suggestion seriously, spending far more time discussing it than any other, and since he nowhere suggests that this solution fails simply because it does not restrict itself to past facts, it is clear that there is no such restriction, even implicit, in Kripke's discussion.

---

[13] Kripke 1982, p. 56.
[14] Ibid., p. 11, emphasis added.
[15] Ibid., p. 13, emphasis added.
[16] Bloor 1997. See also McGinn 1984.
[17] Kripke 1982, p. 69.
[18] Tennant 1997; McGinn 1984.

Kripke purports to consider all the facts, all the putative theories of what constitutes someone's meaning something by a word. Indeed, he must do so if his argument is to support the conclusion that there is *no* fact of the matter what anybody means. If Kripke were to argue that the hypothesis that I mean *addition* is under-determined by all the observable evidence, he would only be entitled to conclude that we cannot *know* that I mean *addition* by 'plus'. We could be said to have no reason to choose the *addition* hypothesis over sceptical alternatives, not that none of the hypotheses is capable of being *true*. If we say that scientific theories are under-determined by the observable evidence, this only implies that we cannot *know* that they are true—they might be true nevertheless, and we might hit upon the true theories by chance, even if our methods are hopeless. Similarly, even if all the observable evidence supports the hypothesis that suspect A alone committed the crime as well as the hypothesis that suspect B alone committed the crime, we are not licensed to conclude that there is no fact of the matter who did it. Kripke's sceptical conclusion simply would not follow from the observation that the observable evidence does not justify our ascriptions of meaning. This brings us to another observation about the nature of Kripke's scepticism: the challenge is metaphysical, not epistemological. Kripke says, 'it is clear that the sceptical challenge is not really an epistemological one. It purports to show that nothing in my mental history of past behaviour—not even what an omniscient God would know—could establish whether I meant plus or quus.'[21]

---

[19] Kripke 1982, p. 15.

[20] Ibid., p. 21.

The reference to an omniscient being here is a literary flourish. The argument is that an omniscient God could never have sufficient evidence for believing that Kripke means *addition* by 'plus' because there simply are no facts for the ideal being to know: with all the facts at His disposal, an omniscient God would still find no fact that constitutes Kripke's meaning *addition* by 'plus'.

Moreover, as Crispin Wright has remarked, if the sceptical argument were construed as epistemological, Kripke would seem to be illicitly asking for an *inductive* justification for the belief that he means *plus* by 'plus'. But knowledge of what I mean or intend is not typically thought to be inductive—only a crude behaviourist would think that I have to look at what I do in order to know what I mean, and Kripke explicitly denies that he assumes behaviourism.[22] The metaphysical interpretation makes the sceptical argument more cogent. On this interpretation, the sceptic asks whether my meaning *addition* can be a factual matter at all, independently of whether I could cite inductive reasons for my knowledge of that fact. Thus, the sceptic asks what *constitutes* my meaning *addition* by 'plus', not what justifies my belief that I mean *addition* by 'plus'. If, as the sceptic argues, there can be no adequate account of what constitutes someone's meaning something by a word, then there is simply nothing to be known, nothing in the world that makes it true that anybody means anything by any word. A metaphysical argument, rather than an epistemological one, is better suited to support the radical sceptical conclusion that there is *no* fact of the matter what anybody means.

All facts are up for grabs, yet there are no meaning facts. The conclusion of the sceptical argument is a negative, and as such, it is difficult to defend. One option would be to consider all the facts. This, however, is hardly feasible. Not only are there too many facts to consider, it is not clear that we now know what all the facts are. Given that we are allowed to consider facts only an omniscient God would know, there

[21] Kripke 1982, p. 21.

[22] Wright 1984, p. 774.

may very well be candidate facts as to what we mean that we do not know and perhaps could never know. Kripke certainly does not purport to consider all the facts in his slim volume. Apparently, all he does is consider very few candidate theories as to what constitutes someone's meaning something by a word. Against these theories, he presents a number of objections, some of which are more convincing than others. None of this quite seems to add up to an *a priori* argument against all possible theories. However, I think that it is possible to construct such an argument from the most powerful criticisms Kripke makes. After reviewing some of Kripke's arguments, I will present what I take to be the *a priori* argument that lies at the heart of Kripke's sundry objections.

THE SCEPTICAL ARGUMENT

Kripke first considers the suggestion that what I mean by 'plus' is determined by what I told myself to do in the past. We can imagine, for example, that I told myself the following: 'When you are faced with two numbers to be added, m and n, count out m marbles and put them in one bowl and count out n marbles and put them in another bowl; mix the contents of the two bowls into a single bowl and count the number of marbles in the amalgamated heap.' Kripke argues that these instructions cannot determine what I mean by 'plus' because the sceptic can always cast doubt on what I *meant* when I told myself to follow this rule. The answer assumes that 'count', as I used the word in the past, meant what we normally take ourselves to mean by it. The sceptic can suggest, instead, that what I meant by 'count' was *quount*, where to quount a heap is to count it unless it was formed as the union of two heaps, one of which had 57 or more items, in which case, the correct answer would be '5'. This might be taken to yield a general restriction. If we try to cite any state with representational content in our account of what makes it the case that someone means something by a word, the sceptic can always offer a reinterpretation of that representation. Thus, it seems, any appeal to representations is ruled out *a priori*.

Next, Kripke considers a theory that allegedly does not presuppose representations: dispositionalism. Kripke's dispositionalist holds that the fact that I mean *plus* by 'plus' is the fact that I am disposed to say that $68 + 57 = 125$; if I meant *quus*, I would be disposed to say '5'. The fact that I mean *plus*, then, is the fact that I am disposed to say that $68 + 57 = 125$, that $122 + 145 = 267$ and so forth for all of the infinite correct uses of 'plus'. Kripke argues that this suggestion faces two main difficulties: (1) it is unable to capture the potential *infinitude* of the correctness condition; and (2) it is unable to rule out *errors* from inclusion in extensions. This latter observation leads Kripke to claim that the dispositionalist cannot rule out error because he treats meaning and use as descriptive rather than normative.

Kripke's argument from the infinitude of the addition function goes as follows. According to the dispositionalist, the fact that I mean *addition* by 'plus' is the fact that I am disposed to respond with the sum of any two numbers when asked. However, the addition function is defined for all pairs of positive integers, no matter how large. And there are some numbers so large that I cannot even grasp them, let alone add them. If asked to add extremely large numbers, I simply would not respond with their sum. The sceptic can then invent another function, call it *skaddition*, which is consistent with the addition function for small numbers, but diverges for numbers too large for me to compute. Now, what makes it the case that I mean *addition* rather than *skaddition* by 'plus'?

The dispositionalist may be inclined, at this point, to idealise. True, there are some numbers too large for me to grasp. However, under epistemically ideal conditions—where I *could* grasp such large numbers—if I were to be asked to add any two numbers, I would respond with their sum. However, Kripke argues that the necessary idealisation is far too radical—how I would behave under these conditions is under-determined by my current dispositional state. Kripke says:

> How in the world can I tell what would happen if my brain were stuffed with extra brain matter, or my life were prolonged by some magic elexir? ... The outcome really is obviously indeterminate, failing further specification of these magic mind-expanding processes; and even with such specifications, it is highly speculative.[23]

In response to this argument, several people have argued that Kripke seems to presuppose that you need to know *exactly* how something would behave under an idealisation in order to say what it is disposed to do.[24] Of course, this is far too strong a requirement. No one knows exactly how an ideal gas would behave; yet, we have no compunctions about saying that the volume of such a gas would vary with its temperature. This response misses Kripke's point, however. What Kripke claims is not that we need to know exactly how I would behave under an idealisation, but that the dispositionalist needs to make unwarranted assumptions about how I would behave under an idealisation that is far too radical. I am a finite being, with finite capacities, whereas the addition function is defined for the infinite number of positive integers. Thus, I would respond with the sum of any two numbers when asked only if I could grasp arbitrarily large numbers. Since there is an upper limit to the numbers I can now grasp, to make assumptions of what I would do if there were no upper limit, is to make assumptions about what I would do if much more intelligent than I really am. And it is difficult to see what the dispositions of this intelligent creature can tell us about me. It is a bit like saying that if I had wings, I would fly—which, although true, does not imply that I am a bird.

---

[23] Kripke 1982, p. 27.

[24] Fodor 1990; Blackburn 1984; Boghossian 1989; Mellor 2000.

Blackburn's response to this objection is more compelling.[25] According to Kripke's dispositionalist, someone who means *addition* by 'plus' is someone who is disposed to respond with the sum of any two numbers, *m* and *n*, when asked to add them. This is what is sometimes called a multi-track disposition because it is a disposition to give a different answer in response to different situations. Similarly, the mercury in a thermometer has a multi-track disposition to expand or contract in response to changes in temperature. The mercury is disposed to respond in accordance with a law; likewise, as Kripke's dispositionalist would have it, I am disposed to respond in accordance with the addition function. If we characterise the disposition to add as a multi-track disposition, we emphasise the *difference* between one addition sum and the next. Infinitude poses a problem, because on this characterisation, in order for me to mean *addition* by 'plus', I must be disposed to respond with the sum of any two numbers, *no matter how large*, which is something I am clearly not disposed to do.

But there is an alternative way to characterise the dispositions involved in adding, which does not force us to idealise by imagining what a speaker would do if infinitely more intelligent than she really is. Instead of treating the disposition that constitutes meaning *addition* as one multi-track disposition, which takes arbitrarily large numbers as inputs, Blackburn suggests that we regard grasp of the addition function as realised by several dispositions each of which takes only small digits as inputs.[26] More precisely, I have the disposition to give the answer '2' to '1 + 1', '3' to '1 + 2', and so forth. I have 90 such dispositions; one for each pair of single digit numbers, from 0 to 9, inclusive. Notice that each one of these dispositions is a single-track disposition, taking a single digit as input. When I am faced with sums involving two digit numbers, I am disposed to write down one above the other, add single digits in each column, going from right to left. When the sum of one column is greater than 9, I am disposed to 'carry'.

---

[25] Blackburn 1984. See also Mellor 2000.

[26] Ibid.

On Blackburn's proposal, someone who means *addition* by 'plus' need not be disposed to respond with the sum of any two numbers when asked. All she is disposed to do is respond with the sum of any two small numbers when asked. According to Blackburn, the fact that I mean *addition* is the fact that, if I were to repeatedly manifest dispositions I currently possess, such as those described above, I would arrive at the sum. Now, the infinitude of the addition function does not in itself pose a problem, because the dispositionalist no longer needs to make a *radical* idealisation in order to specify what answers I would give to sums involving arbitrarily large numbers. Instead of imagining answers I would give if infinitely more intelligent than I currently am, we need only imagine the answers I would give if I had the time and patience to manifest, the requisite number of times, dispositions that I currently possess. Moreover, if I meant *skaddition* by 'plus', I would have to have a further disposition to alter my dispositional make up when the numbers exceed those that I can grasp. If I now lack such a disposition, then I mean *addition* by 'plus'. The problem of infinitude is thus not pressing for the dispositionalist.

This does not show that the dispositionalist solution to the sceptical challenge is beyond reproach. For, the above response to the infinitude problem presumes that, for small numbers at least, I am disposed to respond with the sum. And unfortunately, even with relatively small numbers, most people are disposed to make occasional mistakes. Mistakes are still more frequent when I am asked to compute sums that are longer or more complex. However, the dispositionalist says that what I mean is a function of what I would do; hence, what I would do is correct and determines what I mean—the dispositionalist rules out the possibility of error. Whatever I am disposed to do, on this view, is correct.

Imagine, for example, someone who forgets to 'carry' when adding—someone who says '$68 + 57 = 115$'. Normally, we would say that such a person has made a mistake. But the sceptic can always offer an alternative interpretation, such as that the speaker means a non-standard function, and that when she forgets to carry, she makes no mistake. After all, the dispositionalist says that what she means is a function of what she would do, and what she would do is say '$68 + 57 = 115$'. Any ordinary speaker will have more than one disposition with respect to the use of a word: one disposition to respond with the sum of two numbers when asked and another disposition to respond, on occasion, with something other than the sum. Now, the dispositionalist needs to find some non-arbitrary way to specify which of the speaker's dispositions are 'meaning-constituting' and which are 'error-producing'.[27] And this, Kripke claims, the dispositionalist cannot do. Any attempt to specify which disposition is meaning-constituting, Kripke claims, will be circular.

How are we to establish that some dispositions are meaning-constituting, whereas others are error-producing? One natural thought is that we might specify ideal conditions, such as those in which a speaker is not tired, bored, inattentive, heavily sedated, and so forth. We could then say that under *ideal* conditions, the responses the speaker gives constitute what she means, while under other circumstances—that is when tired, bored, etc.—she may give answers that diverge from those she would give under ideal circumstances, in which case, her answers are liable to be mistaken. That is, if the dispositionalist can specify the ideal conditions in non-semantic, non-intentional terms, she can then say that it is only the dispositions manifest under those conditions which are meaning-constituting.

Kripke claims that the dispositionalist cannot specify the ideal conditions without making the circular assumption of what a speaker means or believes. However, his argument rests on the assumption, discredited by Blackburn, that the dispositionalist must assume that a speaker who means *addition* by 'plus' is one who is disposed to respond with the sum of any two numbers, no matter how large. That is, Kripke argues that the ideal conditions must be those under which I would respond with the sum to arbitrarily large addition sums, and nobody knows what I would do if my brain were stuffed with so much additional material. To assume what I would do if so much more clever than I am now is to already assume that I mean *addition* by 'plus'.[28] Given Blackburn's solution to the infinitude problem, this argument of Kripke's seems less compelling.

Boghossian, however, has reached the same conclusion as Kripke, but by different means. Boghossian argues that 'there could not be naturalistically specifiable conditions under which a subject will be disposed to apply an expression only to what it means; and hence, that no attempt at specifying such conditions can hope to succeed.'[29] The problem for the dispositionalist, as Boghossian sees it, is that for the optimal conditions to genuinely yield the result that is required—that a speaker would use expressions correctly under those conditions—requires that the conditions be *epistemically* ideal; being in those conditions must preclude the possibility of error. Boghossian then argues that it is impossible to specify conditions that preclude error in terms that are entirely non-semantic and non-intentional.

---

[27] Cf. Boghossian 1989.

[28] See Kusch 2005.

[29] Boghossian 1989, p. 537.

Boghossian's main objection has to do with the holistic character of belief fixation. What someone believes in a given circumstance is always mediated by background theory, which consists of a further set of beliefs. In optimal conditions, there can be no interfering background beliefs, so that no beliefs would lead the speaker to make a false judgement. However, because there is an infinitude of potentially interfering groups of background beliefs, there can never be a non-semantic, non-intentional specification of a situation that would rule them all out. As Boghossian puts it:

> [a] dispositional theorist has to specify, without use of semantic or intentional materials, a situation in which a thinker will be disposed to think, *Lo, a magpie* only in respect o magpies. But the observation that beliefs are fixed holistically implies that a thinker will be disposed to think *Lo, a magpie* in respect of an indefinite number of *non-magpies*, provided only that the appropriate background beliefs are present. Specifying an optimality condition for 'magpie', therefore, will involve, at a minimum, specifying a situation characterized by the absence of *all* the beliefs which could potentially mediate the transition from non-magpies to *magpie* beliefs. Since, however, there looks to be a potential infinity of such mediating background clusters of belief, a non-semantically, non-intentionally specified optimality situation is a non-semantically, non-intentionally specified situation in which it is guaranteed that none of this potential infinity of background clusters of belief is present.[30]

Does this objection rule out dispositional theories? Boghossian suggests that the problem arises because the dispositionalist must refer to potentially interfering background beliefs in order to specify the optimality conditions of any one belief—thereby violating the restriction on specifying these conditions in non-intentional, non-semantic terms. In response, the dispositionalist would argue that she can reduce *all* the beliefs involved to dispositions. That is, the dispositional theory could be expressed by what is called a 'Ramsey Sentence'.[31] The Ramsey Sentence of the theory starts with 'There exist $B_1, B_2, B_3, \ldots B_n$ such that … ' and continues with a dispositional specification of each belief $B_1, B_2, B_3, \ldots B_n$. The dispositional specification of each belief would contain a reference to further beliefs, each of which would be given a dispositional specification. For example, call the *lo, a magpie* belief B_1. This belief will be specified roughly as follows: in circumstances where the subject has B_2, B_3, (etc.) and none of B_4, B_5 (etc.), the subject will have B_1 only if there is a magpie present. This sentence makes reference to further beliefs, but each of these can be given a non-semantic specification of the form given for B_1. By conjoining the dispositional specifications of each belief, the dispositionalist would obtain the Ramsey sentence—an extremely long sentence in which the beliefs are collectively given a dispositional specification in purely non-intentional, non-semantic terms.

---

[30] Boghossian 1989, p. 540.

[31] See Ramsey 1965, and Lewis 1970, 1999.

If the dispositionalist can appeal to Ramsey sentences, then she need not appeal to any *intentional* phenomena in order to specify the optimality conditions of beliefs; she would need only to refer to dispositions, making sure that the optimal conditions were those in which the speaker would not have or not manifest certain of her dispositions. Thus, the dispositionalist arguably can specify optimality conditions without violating the requirement that she do so in purely non-semantic, non-intentional terms. Hence, Boghossian's objection does not show that there can be no naturalistic property that distinguishes meaning-constituting dispositions from the rest.

Perhaps Boghossian's objection is more importantly that there is a potential *infinity* of mediating background beliefs.[32] That is, he could be taken to argue that it would not be possible to give a dispositional theory of even one belief because a specification of the optimal conditions would have to rule out *every* potentially interfering background belief, and these are not denumerable. Even with the help of Ramsey sentences, a dispositionalist cannot show that content *is* reducible to dispositions—the task simply cannot be carried out. By the same token, however, the infinitude of the potentially mediating clusters of background beliefs does not show that the dispositional theory cannot be *true*. The infinitude of potentially interfering background beliefs presents a *practical* obstacle to formulating a full dispositional analysis of content, but it does not rule out such an account in principle. Moreover, the dispositional theory of content is no worse off, in this respect, than analyses of non-semantic dispositional properties, such as solubility. To say that

---

[32] Miller 1998, p. 189.

x is soluble in liquid, *l*, is to say that in circumstances *C*, *x* would dissolve if immersed in a sufficient quantity of *l*. However, as C. B. Martin has argued, there is a potentially infinite number of factors that would inhibit the dissolving of *x* in *l*, making it impossible to rule out all of these in the specification of circumstances *C*.[33] Some of these inhibitors will also be dispositional properties, the analysis of which will also require specification of circumstances *C*, which will contain references to further dispositional properties. Nevertheless, no one is about to deny that solubility is a dispositional property, only that an analysis of it cannot be fully carried out, and our best analysis will only be approximately true.

Hence, neither the holistic character of beliefs, nor the infinitude of mediating background beliefs gives us a reason to rule out dispositional theories of content. However, at the conclusion of his discussion of the dispositionalist response to the sceptic, Kripke suggests a further argument against any dispositionalist account. Kripke claims that the fundamental difficulty for the dispositionalist is to account for the *normativity* of meaning. He says:

> The moral of the present discussion of the dispositional account may be relevant to other areas of concern to philosophers beyond the immediate point at issue. Suppose I do mean addition by '+'. What is the relation of this supposition to the question how I will respond to the problem '68 + 57'? The dispositionalist gives a *descriptive* account of this relation: if '+' meant addition, then I will answer '125'. But this is not the proper account of the relation, which is *normative*, not descriptive. The point is *not* that, if I meant addition by '+', I *will* answer '125', but that, if I intend to accord with my past meaning of '+', I *should* answer '125'. Computational error, finiteness of my capacity, and other disturbing factors may lead me not to be *disposed* to respond as I *should*, but if so, I have not acted in accordance with my intentions. The relation of meaning and intention to future action is *normative*, not *descriptive*. ... Precisely the fact that our answer to the question of which function I meant is *justificatory* of my present response is ignored in the dispositional account and leads to all its difficulties.[34]

Kripke says quite clearly that the source of the dispositionalist's difficulties lies in its inability to account for the normativity of meaning—that my answering '125' is *right*, or *justified* by the rule that determines the meaning of '+' for me. The dispositionalist can only give us an account of what the speaker *will* do, not what the speaker *ought* to do. The problem of reducing this 'ought', according to Kripke, is not even touched by the dispositionalist solution. Moreover, if Kripke is right to diagnose the dispositionalist's difficulties in this manner, the same problem will arise for any theory that restricts itself to a descriptive account of the relation between what I mean and what I say. If Kripke is correct in supposing that no descriptive account of this relation can account for its normativity, and if Kripke is right in assuming the relation to be normative, he can use this argument to rule out a large part of the theoretical landscape.

---

[33] Martin 1994.

[34] Kripke 1982, p. 37.

Kripke suggests that the failure of the dispositional theory can be attributed to its treating the relation between meaning and use as causal, and thus that the normativity of meaning rules out all attempts to determine the relation between meaning and use causally.[35] Functionalist accounts of the relation face the same difficulty. A speaker whose function it is to add can also malfunction. In the case of a machine, we can distinguish functioning from malfunctioning by reference to the intentions of the designer, but this is not available to us in response to the sceptic. The sceptic can always cast doubt on our interpretation of the intention of the designer, and all of a sudden, what was a malfunction on the ordinary interpretation, becomes a proper function on the sceptical alternative. 'Malfunction' is analogous to 'mistake', so the functionalist needs to find a non-arbitrary way to specify what are functions, and therefore 'meaning-constituting' and what are malfunctions, and therefore 'error-producing'. Even if the functionalist could specify optimal conditions in non-semantic, non-intentional terms, she could not tell us that answers which deviate from those made under sub-optimal conditions are *wrong*, that they are *mistakes*.[36]

Other people can offer no help either; introducing the community at this stage is also doomed to failure. For, if I can be disposed to make a mistake, so too can everybody else.[37] We might all be disposed to make mistakes when the sums get very, very complex, or very, very long. Certainly, since the addition function is infinite, there will be numbers too large for any of us to accurately compute, and for those numbers, all of us may be disposed to make mistakes. Then, the communal version of the dispositional theory will find itself in the same straits as the individual version—unable to gerrymander the dispositions so that we mean *addition* by 'plus', but nevertheless make mistakes. In general, any attempt to say that my meaning *addition* by 'plus' consists in my being in some physical (dispositional, functional, causal) state, or to say that it consists in my standing in some causal relation (to others, for instance), is doomed to failure. Indeed, Kripke goes on to argue, the prospects look no better for attempts to identify meaning something with being in a psychological state.

---

[35] Kripke, 1982, p. 53

[36] Ibid., pp. 32–4.

[37] Boghossian 1989.

Suppose we imagine that my meaning *addition* is constituted by my being in a particular mental state with a peculiar qualitative feel. Headaches, tickles, and pains are typical examples of mental states with peculiar qualitative feels. Likewise, we might suppose that meaning addition is to be in just such a state, and by introspection, I can simply know what use I ought to make. Kripke objects that this suggestion does not even begin to satisfy the sceptic. Even if, every time I added in the past, I had this particular feeling, the sceptic could easily argue that what I really had was the feeling of quadding, not adding, and so I ought to answer '5'. The fact that this feeling accompanied my past uses of 'plus' does not determine how I should go on. Moreover, if I make a mistake in computing a sum, insofar as I do not realise that I have made the mistake, my adding in that circumstance will be accompanied by the very same qualitative feel. I will feel as if I am adding, even though my uses of 'plus' will deviate from those I ought to give. If having the feeling constitutes my meaning *addition*, we have no way of ruling out unnoticed errors. The feeling is neither here nor there, since I can have the feeling that I am adding, even though I am not in fact giving answers that accord with the function that I, by hypothesis, meant.

In effect, the *quale* theory treats the relation between meaning and use as descriptive, in much the same way as the dispositionalist does. A description of my natural (dispositional) state does not determine what I ought to say. Similarly, a description of my mental state does not determine what I ought to say. I may introspect all I like, but the quality of this experience does not tell me that I ought to say that $68 + 57 = 125$ rather than that it is 5. My being in a certain state—physical or mental—treats the relationship between meaning and use as descriptive when, as Kripke argues, it is normative. Such theories are simply incapable of capturing the normativity of meaning.

The alternative to such theories, of course, is to identify my meaning something by an expression by my being in a psychological state that has representational content. Suppose, for instance, that my meaning *green* by 'green' is constituted by my having an image of something green before my mind when I use the expression—such as an image of grass, or a green coloured square. The natural thought is that on an occasion when

I am going to use the expression 'green' I consult my mental image to see whether the visual image is sufficiently like the mental image. However, the mental image does not determine how it should be compared with visual images. Even if mental representations are like images before the mind, there must be some assumed method of projection of the mental image onto visual images. Whatever representation appears before my mind, it is a representation of a particular thing, and that image can have multiple similarity relations to the things before my eyes. What I need is a rule to tell me how to compare the image before my mind with the things before my eyes.

If I do have a rule in mind, this will indeed determine the correct application of a representation for an infinite number of cases. If I have the rule for addition, for instance, in my mind, then that rule is defined for all positive integers and determines a correct answer to every sum. Though I am a finite being, I can have a rule in mind that determines an infinite number of correct uses. However promising this may seem, it leads to an infinite regress. If understanding any content requires grasping a rule for its use, the sceptic can always cast doubt on my grasp of the rule itself. Suppose that I mean *addition* by 'plus'. In that case, the rule for applying 'plus' must be represented in my mind. The trouble is that the sceptic could give a non-standard interpretation of my mental representation of the rule for applying 'plus'—he could argue that my mental representation tells me to apply 'plus' in accordance with the quaddition function. If the sceptic is right that, if we do not appeal to a rule, then there is nothing which will distinguish between my meaning *plus* or *quus*, then there is equally nothing (apart from a further rule) that will distinguish between my mental representation of the rule meaning one thing rather than another. What constitutes my understanding of the rule itself in one way rather than another? If we needed to appeal to a rule to distinguish my meaning *plus* rather than *quus*, it seems that we need a further rule to determine whether my mental representation is of the rule for addition rather than quaddition. If understanding is to be uniformly cashed out in terms of rules, it seems that I will need a further rule to interpret the rule for applying the expression 'green'. As Kripke puts it:

> It is tempting to answer the sceptic by appealing from one rule to another more 'basic' rule. But the sceptical move can be repeated at the more 'basic' level also. Eventually the process must stop—'justifications come to an end somewhere'—and I am left with a rule which is completely unreduced to any other. How can I justify my present application of such a rule, when a sceptic

could easily interpret it so as to yield any of an indefinite number of other results?[38]

Hence, Kripke could conclude, whatever state we appeal to in response to the sceptic, we will face one of two difficulties. If the state represents a rule, we will initiate an infinite regress; if the state does not represents a rule, we will be unable to account for the normativity of meaning. This completes a fairly general, *a priori* argument against any attempt to claim that someone's meaning something by an expression consists in her being in some sort of physical or psychological state. Kripke then considers two further possibilities: Primitivism and Platonism.

Perhaps we might recoup, suggests Kripke, and insist that my meaning something by an expression is a primitive state, not reducible to any other and not to be assimilated to any representations, qualitative states, or dispositions, and yet by nature introspectable. Kripke discusses this suggestion very briefly. He says that it is in a sense irrefutable, but that it leaves entirely mysterious how any such state could be contained in a finite mind and yet determine what I ought to say in an infinite number of possible situations. Kripke suggests that such a response to the sceptic would be question-begging. The sceptic asks what makes it the case that I mean *addition* rather than *quaddition* by 'plus' and the primitivist answers that it consists in the fact that I am in some primitive state—let's say, the state of 'grasping' *addition* rather than *quaddition* by 'plus'. But 'grasp' is just another word for 'understand' and to say that the fact that I understand 'plus' to mean *addition* is the fact that I grasp the concept of *addition* is merely to assume precisely what is to be shown.

Kripke's response to the primitivist can be countered, however. As Wright points out, it is not the primitivist who begs the question against the sceptic, but the sceptic who begs the question against the primitivist. The sceptical challenge is to find some *further* fact that constitutes my meaning *addition* by 'plus'. Indeed, the primitivist agrees that there is no *further* fact to be found that constitutes my meaning *addition* by 'plus', but this is simply because my meaning *addition* by 'plus' consists in a primitive state which cannot be reduced to any further facts. In response to the charge that the primitivist leaves us with a mystery, Wright responds that the mystery is resolved when we reflect that this is simply a feature of our standard intuitive notion of intention. He says that 'the ordinary notion of intention has it that it is characteristic of mind—alongside thought, mood, desire, and sensation—that a subject has, in general, authoritative and noninferential access to the content of his own intentions, and that this content may be open-ended and general, may relate to all situations of a certain kind'.[39] Perhaps here is no mystery, after all.

---

[38] Kripke 1982, p. 42.

That leaves Platonism. The Platonist maintains that there are abstract objects—*senses*—which determine an infinite extension. For the Platonist, if I mean *addition* by 'plus', then it is the concept of *addition* that determines which uses of 'plus' are correct and which are incorrect. But the sceptic asks, what is it for *me* to mean *addition* by 'plus'? Granted, the sceptic may say, the concept of *addition* determines its extension. But the concept of *quaddition* determines *its* extension as well. So, what determines whether I stand in the appropriate relation to the concept of *addition* and not to the concept of *quaddition*? The Platonist, at least on Kripke's characterisation, says that it has to do with which concept I grasp. It is in virtue of my grasping the concept of *addition* rather than *quaddition* that I mean *addition* rather than *quaddition* by 'plus'.

Once again, Kripke suggests that this is question-begging. The sceptical challenge was to explain what it is that constitutes my understanding one concept rather than another. Yet, once again, 'grasp' seems to be another word for 'understand'. If the Platonist tries to explain what constitutes grasp of a concept, then, he seems to face the same difficulties as all the others. The sceptical problem, Kripke says:

> [c]annot be evaded, and it arises precisely in the question how the existence in my mind of any mental entity or idea can *constitute* 'grasping' any particular sense rather than another. The idea in my mind is a finite object: can it not be interpreted as determining a quus function, rather than a plus function? Of course there may be another idea in my mind, which is supposed to constitute its act of *assigning* a particular interpretation to the first idea; but then the problem obviously arises again at this new level.[40]

Kripke's objection, though compelling, presupposes that grasping a concept must consist in some kind of psychological state. Thus, the sceptical arguments against treating meaning as a psychological state can be simply reapplied. The Platonist could insist that grasping a concept is not reducible to a psychological state, but that it is a primitive relation between the concept and me. My grasp of the concept cannot be further reduced. Kripke does not consider this possibility—his discussion of Platonism is undeservedly brief—but he might be inclined to raise some of the objections made above. If my grasp of a concept is a primitive relation, this leaves entirely unexplained how my meaning something by an expression can determine what *I* ought to say. The concept of *addition* determines its own extension, but it is the fact that *I* grasp that concept that must make it the case that *I* ought to say that $68 + 57 = 125$. More importantly, for the rule for *addition* to guide my use of 'plus', I must have some way of knowing what the rule tells me to do. Otherwise, I would not be motivated by the rule, and thereby motivated by the meaning that I grasp, to say that $68 + 57 = 125$.

---

[39] Wright 1984, p. 776.

[40] Kripke 1982, p. 54

Nevertheless, given Wright's response on behalf of the primitivist, the Platonist could also claim that it is the sceptic who begs the question. If understanding a concept is a primitive relation, or a primitive state, then it should come as no surprise that no further fact can be found that constitutes someone's understanding a concept. Moreover, the charge of mysteriousness is easily avoided. On the Platonist view, the infinite list of correct uses need not be 'contained' in my finite mind, since it is 'contained' in the concept itself. All we need to assume is that I am capable of 'reading off' the correctness conditions via my primitive grasp of the concept.

CONCLUSION

The sceptic allows us to consider *any* fact, even one accessible only to an omniscient God, and argues that there is no fact of the matter what anybody means. This makes Kripke's argument for meaning scepticism the most ambitious argument put forward. Unlike Quine, who places a behaviourist restriction on the facts to which we can appeal, Kripke's argument is far more radical. But the radical conclusion of the sceptical argument needs a more comprehensive argument in its defence. Kripke needs to be able to show that there really is no fact whatsoever that constitutes what someone means. Not only is it possible that there are facts of which we have no knowledge, but the sceptic has no adequate response to the claim that there just are *sui generis* semantic facts. Since Kripke's list of theories is not exhaustive, he is open to the objection that he does not rule out all possible theories of meaning, and thus is not entitled to conclude that there is no fact of the matter what anybody means.

One way to improve the sceptical argument would be to consider more specific, detailed proposals as to what facts constitute meaning. However, even if it turned out that all of the extant proposals fail, this would be insufficient to justify the conclusion that there is no fact of the matter what anybody means. First, it is possible that there are 'natural' facts of which we are currently unaware, and hence it is possible that meaning is constituted by natural facts other than facts about dispositions or functions. Second, it is open to the anti-reductionist to argue, along the lines presented above, that the sceptic illegitimately asks for a reductive answer to the question of what makes it the case that I mean *addition* by 'plus'. We are left with a stand off. The sceptic needs an *a priori* argument against all *possible* naturalistic reductions of semantic statements, and *some* grounds for arguing that the anti-reductionist response is inadequate, other than that it begs the question. This looks like something of a tall order.

Perhaps the gaps in the sceptical argument can be filled, given the assumption that meaning is normative. Kripke seems to believe that meaning is normative, as do many of the participants in the discussion of Kripke's sceptical argument. Moreover, in the course of Kripke's discussion of the dispositional theory, he says that it is the normativity of meaning that yields peculiar problems for dispositionalism, and suggests that all reductive theories would face precisely the same problems as the dispositional theory. Many commentators on this discussion concur that there is reason to believe that there can be no adequate reduction of meaning precisely because meaning is normative.[41]

The trouble is that although it is frequently assumed that the normativity of meaning is the key to Kripke's sceptical argument, neither the claim that meaning is normative, nor its potential implications are carefully spelled out. Hence, in the next chapter, I will consider the thesis that meaning is normative in greater depth and ask whether the sceptic could exploit the thesis that meaning is normative to yield an argument of sufficient generality and *a priori* status. The conclusion of the present chapter is that without some further argument, the sceptical argument fails to establish the radical conclusion that, if we assume semantic realism, there can be no fact of the matter what anybody means.

---

[41] Boghossian 1989, 2003; Brandom 1994*a*; McDowell 1993, 1994; Millar 2004; Wright 1984.

This page intentionally left blank

3

Norms and Normativity

The slogan 'meaning is normative' is catchy. However, it is not always clear what it means, since the word 'normative' is ambiguous. First, it can mean 'prescriptive' or 'action-guiding'. Moral judgements, such as 'you ought to give to charity' are usually thought to be normative in this sense. Second, 'normative' can mean 'relative to a norm or a standard'. For example, the metre bar in Paris was 'normative' in the sense that it was a standard relative to which we could say that my kitchen table is one metre wide. The distinction between these two senses of 'normative' will turn out to be crucial. To mark the distinction, I will take 'prescriptive' or 'action-guiding' to be the primary sense of 'normative', and I will use 'normative' henceforth exclusively in this sense. I will use 'norm-relative' to mean 'relative to a norm or standard'. Thus, we are really concerned with two slogans: 'meaning is normative' and 'meaning is norm-relative'.

Kripke's discussion does not respect this distinction. On the one hand, Kripke clearly attributes to the semantic realist the assumption that understanding the meaning of a word is analogous to following a rule—that when someone means something by a word, such as 'plus', she has a specific rule in mind (either the rule for *addition* or *quaddition*) which determines whether any answer she might give to an addition sum is correct. This assumption is evident, for instance, when he says '[t]his is the whole point of the notion that in learning to add I grasp a rule: my past intentions regarding addition determine a unique answer for indefinitely many new cases in the future.'[1] On the other hand, Kripke attributes to the semantic realist the view that meaning is normative, that is, that what I mean is prescriptive, or action-guiding. This is evident, for instance, when Kripke expresses the 'eerie feeling' someone under the sway of the intuitive, semantic realist view, might have. He says: 'Even now as I write, I feel confident that there is something in my mind—the meaning I attach to the "plus" sign—that *instructs* me what I ought to do in all future cases. I do not *predict* what I *will* do ... but instruct myself what I ought to do to conform to the meaning.'[2] That is, Kripke seems to maintain that, at least according to the semantic realist, meaning must be both norm-relative and normative.

---

[1] Kripke 1982, pp. 7–8.

Despite the ambiguity in Kripke's discussion, the distinction between normativity and norm-relativity is worth respecting. As I will argue, if meaning is normative, a plausible, if controversial, case can be made for the sceptical conclusion that there is no fact of the matter what anybody means. To fend off the naturalists, Kripke needs an *a priori* argument against all possible reductive accounts of what constitutes someone's meaning something by a word. I will suggest that if we grant that meaning is normative, Kripke's sceptic could make use of arguments due originally to Hume and Moore, which purport to rule out all naturalistic reductions of normative moral concepts and properties. Kripke's sceptic also needs to justify a presumption against anti-reductionism, and here Mackie's Queerness argument might just do the trick.

The next obvious question is whether the assumption that meaning is norm-relative gives rise to an argument for semantic non-factualism. I will argue that it does not.

FROM NORMATIVITY TO NON-FACTUALISM

Normative statements are practical, they tell us what to do, what to avoid, who to admire or blame. In contrast, non-normative statements tell us not what to do, nor what attitudes to take, but how things *are*. For example, 'roses are red' is a non-normative statement, whereas 'you ought to respect your elders' is a normative statement. If meaning is normative, then meaning statements, such as 'I mean *addition* by "plus" ' must tell us not how things are with me, but what I ought to do—in this case, to answer with the sum of any two numbers when asked to add them.

Moral statements include paradigmatic examples of normative statements. Moreover, it is the normativity of many moral statements that has led some to argue that there can be no moral facts to correspond to those statements. If semantic statements are normative, this suggests an intriguing line of argument from the normativity of meaning to the view that there can be no semantic facts. There is a widespread intuition that there can be no objective moral facts, because nothing in the world determines what we ought to do, or what attitudes we ought to take. If semantic judgements are normative in just the way that moral judgements are normative, then the meaning sceptic could exploit the strong intuition that there are no objective, normative facts.[3]

---

[2] Kripke 1982, pp. 21–2.

Indeed, Kripke's argument against semantic realism is structurally analogous to A. J. Ayer's argument against moral realism. Ayer holds that moral realists, in committing themselves to the existence of objective, moral facts, face a dilemma. If the putative moral facts are said to be 'natural' facts, the realist commits the 'naturalistic fallacy'. If the putative moral facts are said to be non-natural, moral facts turn out to be unlike ordinary empirical facts and therefore unknowable.[4] Similarly, if we grant that meaning is normative, Kripke's sceptic could be taken to show that the semantic realist faces the same dilemma as the moral realist: if she says that semantic facts are natural, as the dispositionalist does, she commits a fallacy, and if she says that semantic facts are non-natural, she makes them inherently mysterious and unknowable.

In the following two sections, I shall present a series of arguments against moral realism in somewhat more detail. In so doing, I hope to draw attention to two features of these arguments which will be crucial in assessing whether they are of use to the semantic sceptic. The first feature of the arguments that I wish to highlight is that they presuppose that moral statements are prescriptive, or that moral judgements are action-guiding. Second, the arguments presuppose an 'internalist' account of action-guidingness—that is, they presuppose that if someone judges that she ought to do something, then she is *ipso facto* motivated to do it. I highlight the first feature, because it will be important when we come to assess whether meaning is normative. I highlight the second because internalism is a highly controversial doctrine to which the semantic sceptic will be committed if she wishes to make use of the arguments against moral realism.

THE FIRST HORN OF THE DILEMMA: THE 'NATURALISTIC FALLACY'

The ethical naturalist holds that moral statements are true or false in virtue of the 'natural' facts. Although it is difficult to distinguish precisely the natural facts from the rest, it is usually supposed that the 'natural' facts are those that form the subject matter of the natural sciences, most notably, physics, chemistry, and biology.[5] Crucially, these facts are not, so to speak, inherently normative. All of the observational and theoretical statements of the natural sciences (as opposed to directions for setting up experiments) are descriptive, non-normative statements. Thus, all the facts that form the subject matter of the natural sciences can be stated in descriptive, non-normative statements.

---

[3] This connection has been noted by Miller 1998.

[4] Ayer 1936.

One of the best known arguments against ethical naturalism is due to Hume.[6] Hume argued that you cannot derive an 'ought' from an 'is'; or, more precisely, that a normative 'ought' statement cannot be derived from a (consistent) set of non-normative 'is' statements. This is sometimes called 'Hume's Law'. Hume complained that in every system of morality he has met with, the author purports to derive normative conclusions—concerning what one ought to do, or how one ought to be—from non-normative premises—concerning how things *are*. He says:

> This change is imperceptible; but is, however, of the last consequence. For as this *ought*, or *ought not*, expresses some new relation or affirmation, it is necessary that it should be observed and explained; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it.[7]

Hume's view is that normative statements express some further relation, which non-normative statements do not express. Since, in deductive inference, you cannot get out more than you put in, any argument that derives normative statements from non-normative premises must be invalid; a normative conclusion can be validly derived only from a consistent set of premises, at least one of which is a normative statement.

Hume's argument is, of course, controversial. For our purposes, the most important controversy concerns the distinction between normative and non-normative statements. It is clear that the superficial distinction between 'is' statements and 'ought' statements will be inadequate—since some obviously normative statements are 'is' statements, and some obviously non-normative statements are 'ought' statements. For example, 'murder is wrong' is an 'is' statement, but is intuitively normative; it seems to imply that one ought not to commit murder. On the other hand, 'it ought to rain', a mere comment about the weather, contains an 'ought', but is intuitively non-normative; it tells us that it will rain or is likely to rain. Hence, Hume's Law requires a more substantive distinction between normative and non-normative statements. What is this distinction? What is the 'new relation' that normative statements introduce?

---

[5] In the discussion of moral realism, some people include the facts of psychology and social science as 'natural'. This poses obvious problems for the application to Kripke's discussion. For, if the facts of psychology are admissible as 'natural', then why can't the naturalist simply say that it is a psychological fact that I mean *addition* by 'plus'? However, the important point in the characterisation of naturalism is the claim that natural facts are non-normative (in the meta-ethical discussion), for then, if meaning is normative in the right sense, the meta-ethical arguments against ethical naturalism, so defined, should apply to semantic naturalism as well.

[6] Hume [1888] 1968.

[7] Ibid., Bk. 3, Pt. 1, Sec. 1, Para. 28, p. 469.

Hume suggests that the distinction is this: normative statements (and judgements) influence the will in a way that non-normative statements (and judgements) do not. He says, for instance, that normative judgements are:

> supposed to influence our passions and actions, and to go beyond the calm and indolent judgements of the understanding. And this is confirmed by common experience, which informs us that men are often governed by their duties, and are deterred from some action by the opinion of injustice, and impelled to others by that of obligation. … Morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular.[8]

This passage suggests that Hume endorsed a position which has come to be called 'motivational internalism'; the view that there is a necessary, conceptual relation between moral judgement and motivation; that it is a conceptual truth that if an agent judges that it is right for her to φ in circumstances C, then she is motivated to φ in C. The Humean thought, then, is that since normative statements have a necessary connection with motivation, and non-normative statements do not, a normative statement cannot be derived from a consistent sent of non-normative premises.

Motivational internalism supplies the requisite, substantive specification of the difference between normative and non-normative judgements, but the doctrine is controversial. The plausibility of the doctrine lies in the observation that most people are motivated by their normative judgements. That is, if the woman on the Clapham Omnibus judges that, all things considered, she ought to ϕ, she will normally be motivated to ϕ, and if she judges that she has some reason to ϕ, she will normally feel some 'pull' towards ϕ-ing. It would be odd if she were to say that she judges that she ought to ϕ, but that she feels absolutely no inclination to ϕ.

---

[8] Ibid., Bk. 3, Pt. 1, Sec. 1, Paras 6 and 7, p. 457.

However, the motivational internalist makes a stronger claim. In its strongest version, it is the thesis that it is a *conceptual truth* that if someone judges that she ought to ϕ, she is *ipso facto* motivated to ϕ.[9] And this claim is more controversial—it is not just that the woman on the Clapham Omnibus will normally be motivated to do what she judges she ought to do, but that her motivation is necessitated by her normative judgement. If she is not motivated to ϕ, then it cannot be the case that she judges that she ought to ϕ. This stronger claim seems less plausible when we think of someone who is weak-willed, for one reason or the other. Many of us know someone who sincerely judges that, all things considered, she ought to quit smoking, but continues to smoke nonetheless and seems to feel no motivation to quit whatsoever, either because she is depressed, or simply lacks will power. On the assumption that normative judgements are necessarily motivating, it would seem that if someone suffers from depression or weakness of will, she cannot genuinely judge that she ought to quit smoking. But this does not seem to be correct.

Instead, we might consider a weaker version of motivational internalism that takes account of phenomena such as weakness of will. Michael Smith, for example, argues that the fundamental insight of motivational internalism is that normative judgements are practical, and this insight can be retained by an account that nevertheless accommodates cases of weakness of the will. According to Smith weakness of the will is a failing of practical rationality. So, the motivational internalist should say instead that if someone judges that she ought to ϕ, then either she is motivated to ϕ, or she is practically irrational.[10] The immediate question this raises is whether, given the weaker formulation of motivational internalism, it can still be shown that normative judgements cannot be derived from non-normative judgements. Perhaps this form of internalism is too weak: if one can make a normative judgement and still not be motivated (because one is practically irrational), then it seems as though there is no longer a conceptual or necessary link between normative judgements and motivation—normative judgements do not introduce a 'new relation'.

Unfortunately, a proper treatment of these issues would take me too far afield. Hence, I will assume, for the sake of argument, that some form of internalism will do. The point is that Hume's Law presupposes motivational internalism in some form or other, and thus presupposes a controversial, substantive account of the nature of normative judgements. If Kripke's sceptic is to make use of Hume's Law, it must be assumed that meaning statements, such as 'I mean *addition* by "plus" ' entail 'ought' statements, such as 'I ought to respond with the sum of any two numbers when asked', and it must be assumed that if I judge that I ought to respond with the sum of any two numbers when asked, then I must be motivated to do so (perhaps on pain of practical rationality).

---

[9] This is what Smith 1994 calls rationalism, or strong internalism.

[10] Smith 1994, p. 61. This is what Smith calls the 'practicality requirement.'

The Humean claims that there is an unbridgeable gap between 'is' statements and 'ought' statements. Similarly, some say that there is a gap between fact and value. Those who do, claim that evaluative statements—those containing words such as 'good', 'bad', 'right', and 'wrong'—cannot be equivalent to descriptive statements and cannot correspond to natural facts. Following this line of argument, the semantic sceptic might argue that meaning statements entail *evaluative* statements, that is statements containing words such as 'good', 'bad', 'right', and 'wrong'. So, for example, 'I mean *addition* by "plus" ' would entail 'responding with the sum of any two numbers when asked is good/right; responding with a number other than the sum is bad/wrong'. If semantic statements entail evaluative statements, the sceptic might make use of G. E. Moore's Open Question Argument against the ethical naturalist.

Moore argued that the moral concept of *good* cannot be analytically equivalent to any other naturalistic concept or set of concepts (and, since goodness and rightness are interdefinable, for Moore, the same goes for the concept of *right*).[11] Moore reasoned as follows. If two concepts are analytically equivalent, someone who understands those concepts understands that they are equivalent. For example, the concept, *bachelor* is analytically equivalent to the concept of *unmarried man* because *bachelor* and *unmarried man* are synonymous. Since *bachelor* and *unmarried man* are synonymous, anyone who understands the concept *bachelor* and the concepts *unmarried* and *man* should recognise the analytic equivalence of these concepts. Thus, someone who says 'Jones is a bachelor, but is he an unmarried man?' would display some kind of conceptual confusion akin to the kind one would display by saying 'Jones is a bachelor, but is he a bachelor?' Both of these questions are ‘closed’ given the analytic equivalence of *bachelor* and *unmarried man*. Anyone who would seriously consider either question must fail to grasp one of the concepts involved.

---

[11] Moore 1993. I have relied heavily on Feldman 1978, ch. 13 and Miller 2003, ch. 3 both of whom make Moore's arguments clear. I am grateful to Krister Bykvist for help with this section and the next.

Moore argued that there is no natural concept or set of concepts analytically equivalent to the concept *good* in the same way that *unmarried man* is analytically equivalent to *bachelor*. For example, suppose we are considering whether *good* is analytically equivalent to *such as would produce pleasure*. Now, imagine someone who grasps the concept *good* and grasps the concept *such as would produce pleasure*. Moore says that such a person could, without displaying any conceptual confusion, say: ‘eating chocolate cake is pleasurable, but is it good?’ In contrast, someone who says ‘eating chocolate cake is good, but is it good?’ or ‘eating chocolate cake is pleasurable, but is it pleasant?’ does display conceptual confusion. If the concept of being *such as to produce pleasure* were analytically equivalent with the concept of *good* then the question whether pleasure is good would be closed, just as the question whether pleasure is pleasant is closed. Moore suggests that the Open Question Argument is a device that can be used against any putative natural concept, *N*—that, for any concept of a natural property *N*, full grasp of both *N* and the concept of *good* does not preclude our seriously considering whether the property referred to by *N* is really good.

Why is Moore convinced that, for any concept of a natural property, *N*, it will always be an open question whether the property referred to by *N* is really good? Moore suggests that this is because the property of goodness is simple; it has no parts and is not identical to any other properties. Yet this is just another way of stating the claim that *good* is indefinable. In the absence of an independent explanation of why the Open Question Argument should always succeed, it seems as though Moore simply begs the question—he must assume that there is no *N* that is analytically equivalent to *good* in order to justify his claim that there will always be an open question whether something that is *N* is also good.[12] What the Open Question Argument needs is an explanation of why it will always succeed, independent of the assumption that *good* is indefinable.

Furthermore, since Moore is concerned to show that *good* is not analytically equivalent with any other concepts, his argument only addresses an analytic form of naturalism—that is one according to which evaluative concepts are said to be defined in terms of natural concepts. Such a naturalist in effect claims that the reduction of evaluative concepts is *a priori*. To know a conceptual truth we need not investigate the world, but can merely reflect on our concepts; to know that *bachelor* is analytically equivalent to *unmarried man* we need not undertake an empirical investigation—we need not check bachelors to see if they are unmarried men. However, naturalists do not all take the view that moral concepts are analytically equivalent to natural concepts; instead, they might hold that the fact that moral properties are identical to some natural properties is *a posteriori* rather than *a priori*. By analogy, consider the discovery that water is H_2O. We had the concept of *water* before we discovered that water is H_2O. Having discovered that water is H_2O, we can now say that the concepts *water* and H_2O refer to the same property in all possible worlds—that necessarily, water is H_2O.[13] As Moore formulates it, the Open Question Argument would not seem to apply to this kind of naturalistic reduction, for it is not necessary to have the concept H_2O in order to have the concept *water*, even though the two concepts refer to the same property. Someone who seriously considers whether water is really H_2O displays no conceptual confusion, merely an ignorance of the empirical facts. Similarly, a naturalist might argue, we have the concept of *good*, and we might discover that the concept *good* refers to the same property as some natural concept, *N*. Given this identity, we would then be able to say that necessarily, *N* is good, although one would display no conceptual confusion by asking whether *N* is good.

---

[12] Frankena 1939.

Thus, on Moore's original formulation of the Open Question Argument, it does not seem to show that evaluative concepts cannot be equivalent to natural concepts. However, despite the difficulties with Moore's presentation of the Open Question Argument, it still holds sway. Many people find it intuitive that evaluative concepts are fundamentally different from natural concepts, and this difference makes it seem that no reduction—either analytic or synthetic—will work. Perhaps this is due to an intuitive commitment to motivational internalism. Indeed, Stephen Darwall, Alan Gibbard, and Peter Railton have suggested that, if we grant motivational internalism, the Open Question Argument can be improved. On Darwall *et al.*'s formulation, the Open Question Argument goes as follows:

> Attributions of goodness appear to have a conceptual link with the guidance of action, a link exploited whenever we gloss the open question 'Is *P* really

---

[13] Kripke 1972.

> good?' as 'Is it clear that, other things equal, we really ought to or must devote ourselves to bringing about *P*?' Our confidence that the openness of the open question does not depend on any error or oversight may stem from our seeming ability to imagine, for any naturalistic property *R*, clear-headed beings who would fail to find appropriate reason or motive to action in the mere fact that *R* obtains (or is seen to be in the offing). Given this imaginative possibility, it has not been *logically* secured that *P* is action-guiding. ... And this absence of a logical or conceptual link to action shows us exactly where there is room to ask, intelligibly, whether *R* is really good.[14]

Assuming motivational internalism, what makes *good* an evaluative concept is that judgements of what is good are *normative*; they have a conceptual link with motivation—if one judges something to be good one is *ipso facto* motivated to pursue it. In contrast, if one judges something to be pleasurable, one is not *ipso facto* motivated to pursue it. So, one may reasonably ask whether pleasure is good (i.e. worthy of pursuit), whereas one cannot reasonably ask such a question of goodness. And what goes for pleasure will go for any other natural concept whatsoever, since natural concepts conspicuously *lack* this internal conceptual link with motivation. It follows that no evaluative concept can ever be analytically equivalent with any natural concepts.

This version of the Open Question Argument seems to offer a response to the charge of begging the question. Motivational internalism supplies an independent explanation of why it seems that it will always be an open question whether, for any natural concept *N*, something that is *N* is good.[15] Whether this improves Moore's case against the *a posteriori* reduction of evaluative concepts is less clear. For, even if the concept *good* has a conceptual link with motivation, it can still pick out a natural property.

There are, of course, independent arguments against an *a posteriori* reduction of evaluative concepts. For instance, following Hare, Smith argues that the metaphysical naturalist will have difficulty in accommodating moral disagreements.[16] Suppose that in one community, the term 'good' is consistently applied to people who are mild mannered and considerate of others, whereas in another community, 'good' is consistently applied to people who are aggressive and strong. According to the metaphysical naturalist, when members of the first community use the word 'good', it refers to the properties of being mild mannered and considerate of others, whereas when members of the second community use ‘good’, it refers to the properties of being aggressive and strong. Since ‘good’ has different meanings, relative to the two communities, they can have no genuine moral disagreement whether someone, such as Jones, is good. For Hare, the problem is that the *a posteriori* reduction of ‘good’ to natural properties fails to accommodate the intuition that to say that something is good is to say that it is commendable, or worthy of pursuit. If this is the explanation of why this form of naturalism may seem wrong, it too relies on a form of motivational internalism.

---

[14] Darwall, Gibbard, and Railton, 1992, p. 117.

[15] Darwall *et al.*,1992, p. 117–18.

[16] Smith 1994, Hare 1952.

Once again, this issue need not be decided here. The main point is that the plausibility of the Open Question Argument depends on our regarding evaluative concepts, such as *good*, *bad*, *right*, and *wrong* as normative, as having a necessary connection with motivation. If the sceptic is to make use of the argument, then, she must assume that semantic judgements involving evaluative terms, such as ‘ “125” is the right answer to “57 + 68 = ?” ’ are normative judgements and that, if I make a judgement such as the above, I in effect judge that ‘ 125’ is the answer I ought, *ceteris paribus*, to make. This judgement, in turn, must have a conceptual connection with motivation: if I judge that ‘ 125’ is the right answer, then I must be motivated to give it.

THE SECOND HORN OF THE DILEMMA: QUEERNESS

Kripke’s sceptic needs to rule out not just all naturalistic theories of meaning, but also all non-naturalistic theories as well. Moore himself was an anti-reductionist about goodness, and presented the Open Question Argument in defence of his anti-reductionist theory that goodness was an irreducible, *sui generis*, non-natural property. Similarly, an anti-reductionist about meaning and semantic correctness could very well argue that semantic correctness is an irreducible, *sui generis*, non-natural property.

Ayer argued against non-naturalism as well as naturalism. However, Ayer’s argument against the non-naturalist would be unsuitable for Kripke’s use, since it presupposes a verificationist criterion of significance, according to which synthetic statements are meaningful only if they are empirically verifiable. Ayer argued that if moral statements described non-natural states of affairs, they would be unverifiable, and therefore meaningless.[17] Unfortunately, Ayer's verificationism has been discredited, and cannot therefore be relied upon to ground the sceptical argument. However, Kripke's own discussion of non-naturalism, although inconclusive, even question-begging, points to the sort of argument he might make. Kripke argues that to treat my meaning *addition* by 'plus' as a *sui generis*, irreducible state commits us to supposing that such a state can *guide* me, not causally, but in some *queer* way.[18] Kripke suggests, that is, that non-natural semantic facts would be inherently queer.

In a similar spirit, Mackie argued that non-natural moral facts would have to be inherently queer. Our conception of moral principles—such as 'we ought to help the least well off'—is that they are both objective and categorically prescriptive.[19] To say that they are objective is to say that moral principles are true in virtue of moral facts or properties that must exist independently of our judgements about them; that moral principles are true independently of our judgements whether they are true. To say that moral principles are categorically prescriptive is to say that they are true in virtue of facts which give us reason, sometimes sufficient reason to act, quite independently of our desires. In contrast, just about any fact can give me a non-categorical reason to act, given the right desires. If I wish to get to Edinburgh in the shortest possible time, the fact that an airplane will get me there in the shortest possible time gives me a reason to fly to Edinburgh. If I wish to feel warmer, then the fact that putting on a sweater will make me feel warmer gives me a reason to put on a sweater. If I no longer want to get to Edinburgh in the shortest possible time, or if I don't want to feel warmer, then I no longer have a reason to fly to Edinburgh or put on a sweater. In contrast, moral principles seem to be categorically prescriptive in the sense that they hold quite independently of our desires. The principle that I ought to help the least well off gives me a reason to give to Oxfam even if I have no desire to give to Oxfam. It is fortuitous if I want to give to Oxfam as well as having a reason to do so, but it is not necessary that I want to give to Oxfam for me to have a reason to do so.

The thought that moral facts are categorically prescriptive is closely related to the thesis that played a decisive role in the reformulation of the Open Question Argument—that there is an internal connection between moral judgements and motivation. According to motivational internalism, my judgement that I ought to give to Oxfam motivates me to give to Oxfam, whereas my judgement that the fastest way to get to

---

[17] Ayer 1936.

[18] Kripke 1982, p. 53.

[19] Mackie 1977. Cf. Miller 2003.

Edinburgh is to fly does not motivate me to fly to Edinburgh in the absence of a desire to get to Edinburgh as quickly as possible. Thus, Mackie might say, our normative moral judgements could only have this internal relation to motivation if the putative facts that make them true were categorically prescriptive—that they give us reasons to act quite independently of our desires. Thus, if there were normative moral facts, if there were genuinely normative moral properties, they would have to be both objective and categorically prescriptive.

Mackie argues that we know of nothing in the world that is objective and prescriptive, so if there were normative moral properties, they would have to be properties of a very peculiar sort. For instance, Mackie says:

> An objective good would be sought by anyone who was acquainted with it, not because of any contingent fact that this person, or every person, is so constituted that he desires this end, but just because the end has to-be-pursuedness somehow built into it. Similarly, if there were objective principles of right and wrong, any wrong (possible) course of action would have not-to-be-doneness somehow built into it. Or we should have something like Clarke's necessary relations of fitness between situations and actions, so that a situation would have a demand for such-and-such an action somehow built into it.[20]

Mackie argues that when we consider states of affairs in the world, we do not find that they have demands for action built into them. Moreover, he argues, none of the ordinary properties and facts with which we are acquainted seem to be such that mere apprehension of them motivates rational agents to act. This is clearly the case with such properties as redness and being square—mere apprehension of those properties does not motivate rational agents to act, quite independently of any desire. So much has already been defended by the Moorean argument given above. Now, instead of natural properties and facts, consider other non-natural properties and facts with which we are familiar—such as the facts of mathematics. Suppose, for the sake of argument, that the facts of mathematics are non-natural. Let us also suppose that there are objective facts about mathematics, that is, facts that are independent of our mathematical judgements. Now, if we compare mathematical facts with putative moral facts, we still find a substantive difference. My judgements about the facts of mathematics do not give me a reason to act quite independently of any desire: my judgement that $2 + 2 = 4$ does not motivate me to act independently of my desires. Thus, even if we postulate objective, *sui generis* mathematical facts, we are not postulating facts that are categorically prescriptive. In contrast, if we suppose that there are normative facts, we are postulating *sui generis*, objective facts that are also categorically prescriptive. Putative normative facts would be utterly unlike any of the other facts with which we are acquainted, including putative non-natural facts of mathematics. Thus, to postulate *sui generis* normative facts or properties is to postulate entities that are inherently queer.

---

[20] Mackie 1977, p. 70.

The anti-reductionist could bite the bullet and say that there just are such queer facts and properties in the universe. However, this bullet is unpalatable for two reasons. First, because non-natural properties of any kind (whether mathematical or moral) are by definition non-causal and not accessible to the senses, the anti-reductionist would have to explain how I can be acquainted with them at all. The anti-reductionist is thus committed to a bizarre epistemology, according to which I must have some non-sensory faculty of intuition so that I can become acquainted with normative facts. Second, because non-natural properties are by definition non-causal, the anti-reductionist would have to explain how it could be that a rational agent's apprehension of such properties could have the effect of motivating her to act. Other motivational states, such as desires and intentions, typically come from the inside. You can of course cause me to have a desire from the outside—you could, for instance, cause me to desire to take a painkiller by punching me in the nose—but what you have done is *cause* me to have a desire. If normative properties are non-causal, it is difficult to see how they could impinge on rational agents in such a way as to motivate them to act. Thus, the anti-reductionist seems to commit himself to a bizarre epistemology and a bizarre ontology for which we have no models from any other domain.

Although Mackie's argument does not supply a conclusive refutation of anti-reductionism, it does supply us with a presumption against anti-reductionism. The anti-reductionist needs to be able to argue that she can give a suitable epistemology for our normative judgements and show that normative facts need not be inherently queer or that their queerness justifies no presumption against them. At any rate, Mackie does not simply argue that anti-reductionism begs the question against someone who asks for a suitable reduction. Similarly, Kripke could argue that semantic facts cannot be non-natural because they would be unknowable, without the assumption of a peculiar epistemology, and because they would be metaphysically queer; quite unlike anything else in the universe.

FROM NORMS TO NORMATIVITY

The foregoing foray into meta-ethics might appear to some to be a digression. To be sure, there are other arguments which might be brought to bear against both reductive and non-reductive accounts of meaning and content. For instance, non-reductive accounts of meaning and content could be objected to on the grounds that they make it mysterious how the content of people's beliefs could have any role in the causation of action. However, this objection has no basis in Kripke's discussion, whereas the argument from the normativity of meaning does. Not only does Kripke suggest that meaning is normative, others who have been closely involved in the discussion of Kripke's views have regarded the claim that meaning is normative to be one of Kripke's central insights and the key to the sceptical argument. For instance, Boghossian takes the normativity of meaning to rule out all reductive theories.[21] Boghossian even suggests that the irreducibility of meaning facts has to do with the conceptual links between judging that some use of an expression is correct and being motivated to use it correctly. This is evident when he says that 'to be told that "horse" means *horse* implies that a speaker ought to be motivated to apply the expression only to horses; whereas to be told, for instance, that there are certain select circumstances under which a speaker is disposed to apply the expression only to horses, seems to carry no such implication'.[22] And Boghossian further recognises that if Kripke's sceptic could make use of an argument like Mackie's, the sceptic would have a chance of ruling out all anti-reductionist theories as well.[23] Indeed, although this assumption is rarely explicitly articulated, most people involved in this discussion take meaning to be normative,[24] and suppose furthermore that the normativity of meaning poses peculiar difficulties for reductive theories or even rules them out entirely.

The previous discussion provides some insight into why one might think that normative judgements must be non-factual. But how are the foregoing meta-ethical arguments to be applied to the semantic domain?

[21] Boghossian 1989, especially pp. 511–15 and 527–34.

[22] Ibid., p. 533.

[23] Ibid., p. 540.

[24] Blackburn 1984; Bloor 1997; Boghossian 1989, 2003; Brandom 1994*a*; Gibbard, 2003; Glock 1996; Lance and O'Leary Hawthorne 1998; McDowell 1993; McGinn 1984; Millar, 2004; Miller 1998; Pettit 1990*a*; Soames 1998*b*; Wright 1984.

A semantic judgement, such as 'Jones means *addition* by "plus" ' contains none of the words that signal a normative judgement, such as 'ought', 'good', or right. In order for the sceptic to employ the argument from the normativity of meaning to semantic non-factualism, she would have to show that the semantic realist is committed to the view that meaning is normative. In what follows, I try to reconstruct the reasoning that leads some from semantic realism to the normativity of meaning.

First, recall that the semantic realist maintains that what someone means or understands by a word (mental representation) can be given by the correctness conditions of the word (mental representation) as it is understood. So, according to the semantic realist, if 'green' means *green*, then 'green' applies correctly to all and only green things. To put it more formally, the semantic realist thesis under sceptical attack can be stated as follows (where x is an expression, F a meaning or content, a is an object or other suitable referent for x, and f is a feature or set of features):

> Semantic Realism: x means $F \rightarrow (a)(x$ applies correctly to $a \leftrightarrow a$ is $f)$

Someone who accepts that meaning is normative might wish to claim that it is simply an intrinsic part of semantic realism itself. 'Correctness', one might say, is an evaluative concept. If the fact that 'green' means *green* implies that 'green' applies *correctly* to something if and only if it is green, then 'means' implies 'correct'. If *correct* is an evaluative notion, just like *good*, then the fact that 'means' implies 'correct' shows that meaning is normative in the requisite sense.

This is far too quick, however. The 'correct' in 'correctness condition' is not obviously an evaluative notion, akin to 'good' in Moore's discussion. For the semantic realist, the conditions for the correct use of an expression must be understood as the conditions that must obtain in order for the expression to refer, denote, or be truly predicated of something. The expression 'applies correctly' is a placeholder for the various semantic relations an expression can have to the world: it stands for either 'x refers to a', 'x denotes a', or 'x is true of a'. Thus, if we substitute into the above semantic realist thesis one of the semantic relations for which 'applies' stands, there should be no normativity bells ringing:

> Semantic Realism (reference): x means $F \rightarrow (a)(x$ refers to $a \leftrightarrow a$ is $f)$

We might still call f the 'correctness conditions' of x, but 'correctness' here does not, on the face of it, seem to be an evaluative notion. Reference is not an evaluative notion—at least on our usual notion of reference. Thus, with the above substitution in place, we should see that semantic realism does not imply that meaning is normative. More importantly, as I have presented it, semantic realism does not even mention a speaker. However, action-guidingness and the link with motivation which played such a crucial role in the foregoing meta-ethical arguments, have to do with the relation between normative judgements, or putative normative properties and *agents*. Semantic realism, as it stands, says nothing whatsoever about agents. Semantic realism does not by itself imply that meaning is normative in the relevant sense.

Since the normativity of meaning is clearly not a direct implication of semantic realism, as sketched above, where does the assumption that meaning is normative come from? The answer is fairly obvious: it comes from the assumption that meaning is analogous to rule following, the assumption which gives its name to Wittgenstein's rule following considerations.[25] To say that a speaker's meaning something by an expression is a species of rule following is to say considerably more than to say that meaningful expressions have correctness conditions—it is to say what it is for a speaker to understand the meaning of an expression, and what it is for her to use the expression with the meaning that she understands. The rule following assumption tells us specifically about how a speaker's knowledge of meaning relates to her behaviour in the use of words. Semantic realism says nothing about this. These two theses are thus independent; we could do away with the epistemological view without touching the metaphysical one.[26] And it is the assumption that understanding is analogous to rule following that gives rise to the idea that meaning is normative.

If meaning is analogous to rule following, then meaning is norm-relative in the sense that a semantic rule (the correctness conditions) determines which uses of an expression are correct, that is, which uses accord with the rule. However, the semantic rule in question could be merely a statement of the correctness conditions of the expression in question. If so, semantic rules would be of the following form:

R1: $(a)(x \text{ applies correctly to } a \leftrightarrow a \text{ is } f)$

---

[25] Boghossian 1989 says that this is a misnomer for Kripke's problem—which has nothing to do with rule following, but with the nature of meaning and content. I think that he is correct that it has little to do with rule following *as such*, since it has little to do with rule following in other contexts—such as in law or playing a game—however, it does seem to be necessary to bring this assumption to the fore to make sense of the claim that meaning is normative.

[26] Cf. Davidson 1984.

R1 simply states the correctness conditions of an expression. It is only a small step from semantic realism to the idea that a speaker, in order to use an expression with a given meaning, must 'follow' a rule such as R1. If grasp of meaning is grasp of the rule that specifies correctness conditions, then my use of an expression will be 'correct' insofar as it accords with the rule that I follow. If meaning something by a word requires that a speaker 'follows' a rule such as R1, then meaning is norm-relative in that it involves rules or norms, but will it be the case that meaning is normative?

The first thing to note is that R1 is not a prescriptive rule, it does not tell an agent what she ought to do. Rather, it is a non-prescriptive, 'constitutive rule'. On John Searle's account, constitutive rules are those that define a practice or action; they are necessary for the possibility of performing that practice or action. For example, the activity of playing Mah-jong is not possible independently of the constitutive rules of Mah-jong. A group of people might go through the motions with marked tiles, but if they are in utter ignorance of the rules of Mah-jong, they are not playing Mah-jong. Before the rules of Mah-jong were invented, nobody played Mah-jong; nobody could have played Mah-jong without first inventing the rules. In contrast, there are rules that are not essential, in this way, to practices. Searle calls these 'regulative' rules because they regulate antecedently existing activities or practices. Consider, for instance, the Indian rule of etiquette that prohibits eating with one's left hand. This rule regulates the activity of eating, which certainly predates the invention of the rule, and could just as well outlive it. Searle suggests that what distinguishes constitutive rules from regulative rules is that you cannot describe behaviour that accords with a constitutive rule without at least implicit reference to the rule:

> Where the rule is purely regulative, behaviour which is in accordance with the rule could be given the same description or specification (the same answer to the question "what did he do?") whether or not the rule existed, provided the description or specification makes no explicit reference to the rule. But where the rule (or system of rules) is constitutive, behaviour which is in accordance with the rule can receive specifications or descriptions which it could not receive if the rule or rules did not exist.[27]

A rule such as R1 seems to be a constitutive rule in Searle's sense, because we cannot describe someone as meaning, say, *green* by 'green' without ascribing to her the knowledge, however tacit, that 'green' applies correctly to all and only green things. That is, if a speaker has a rule such as R1 in mind, we can describe the speaker's use of the word 'green' as *meaningful* and, in particular, as meaning *green*. A rule's being constitutive does not preclude it from being prescriptive—from telling a speaker what she *ought* to say. However, R1 is also clearly not a prescriptive rule. R1 does not tell a speaker to apply *x* to something if and only if it is *f*. As a result, it is difficult even to make sense of the idea that someone could 'follow' a rule such as R1, since it does not give a directive that can be followed. It seems, therefore, that even if a speaker must have a rule such as R1 in mind, in order to mean something by a word, it does not follow that meaning is normative.

---

[27] Searle 1970, p. 35.

One might say, instead, that for me to mean something by an expression, a rule of the form of R1 must be 'in force'.[28] And one might suppose that whether or not a rule is 'in force' is itself a normative matter: if R1 is 'in force' for a speaker, then it is 'binding'; the speaker ought to aim to comply with R1.[29] As I see it, this means that R1 will tell me which uses of *x* are correct, and an additional rule will tell me that I ought to aim to use *x* correctly, in accordance with R1. This additional rule will, indeed, be prescriptive—it tells me what I ought to do—and whether or not a rule such as R1 is 'in force' would indeed be a potentially irreducibly normative matter. However, there is an alternative way to characterise what it is for a rule to be 'in force', which does not make it a normative matter whether a rule is in force. That is, we might say that a rule—such as a legal rule—is in force in the sense that it is accepted by a relevant community, and enforced by sanctions. On this characterisation, to say that a rule such as R1 is 'in force' for a speaker is not to say that the speaker ought categorically to comply with it, but to say that it is accepted by the speaker, and perhaps her linguistic community, and that failure to comply with it will result in sanctions (including perhaps, simply being misunderstood). These are all descriptive statements, as are the statements that are made true by the fact that R1 is 'in force' (i.e. regarding which actions can be described as 'correct' and which cannot).

The point is that to assume that meaning is analogous to rule following is not sufficient to show that meaning is normative. If for a rule to be 'in force' is for it to be accepted and enforced by sanctions, then even if R1 is in force it does not follow that there must be a further prescriptive rule governing behaviour.[30] The fact that R1 is accepted and enforced by sanctions does not make it the case that I ought to comply with R1, quite independently of any of my desires, such as the desire to avoid punishment. Rather, if a rule such as R1 is 'in force' (just in the sense of being accepted by me or my linguistic community), and I apply x to some a that is not f, then I have made an 'incorrect' application of x. This is because R1 merely states the conditions under which e correctly applies. It *distinguishes* my uses of x between those that accord with its meaning and those that do not; it supplies a *description* of my uses of x as those that accord with the rule and those that do not.[31] More formally, the thesis that meaning is norm-relative, where the semantic norms are like R1, amounts to this (where S is a speaker):

---

[28] Glüer and Pagin 1999.

[29] Thanks to an anonymous referee for this suggestion.

> *Norm-Relativity*: S means F by $t \rightarrow (a)(S$ applies x 'correctly' to $a \leftrightarrow a$ is $f)$

If semantic rules are not prescriptive rules, then even if we assume that meaning is norm-relative, it does not follow that meaning is normative. For, even if a rule like R1 is in force, in the sense that it is accepted by the speaker or her linguistic community, the judgement that some course of action accords with R1 is not categorically prescriptive (since it does not enjoin any actions whatsoever); it does not *ipso facto* motivate a rational agent to pursue any course of action. For meaning to be normative in the sense required for the meta-ethical arguments against both naturalism and non-naturalism to go through, the semantic rules a speaker is taken to follow must be prescriptive rules—they must tell the speaker what she ought to say. If semantic rules are taken to be prescriptive, then it would arguably be constitutive of the concept of semantic correctness that a speaker who judges that some expression x applies correctly to some a is *ipso facto* motivated to apply x to a; the judgement that some use of an expression is correct, like the judgement that some action is right, would be categorically prescriptive. This is the sense of 'normative' we assume when we say that moral rules are normative—we mean that moral rules tell us what we ought to do—and it is this feature of moral requirements that gives rise to the arguments for the non-factuality of moral judgements. Kripke clearly suggests this interpretation of 'meaning is normative' when he says, for instance, that 'the relation of meaning and intention to future action is *normative* not *descriptive*'.[32] The contrast between 'normative' and 'descriptive' is that between 'prescriptive' and 'descriptive'. This would make no sense if 'normative' simply meant 'rule-governed', since my behaviour could be governed by a rule such as R1 which only supplies a description of my uses of an expression as those which accord and those which do not accord with the rule. If the relation of meaning and intention to future action is normative, the rule that I must be following in order to mean *addition* by 'plus' must be a prescriptive rule, such as R2:

---

[30] If we assume the 'prescriptive' characterisation of what it is for a rule to be 'in force', then what someone means by a word will determine how she ought to use it. However, if what it is for a rule to be 'in force' is for it to be 'backed up' by a prescriptive rule, this view will face all of the same objections which I raise against the thesis that meaning is normative in Chapter 7. My point here is simply that we can distinguish between the thesis that meaning is norm-relative and the thesis that meaning is normative—one according to which what a speaker means determines which of her uses can be described as 'correct' and another according to which what a speaker means determines which uses of an expression she *ought* to make. Given that there is a way to characterise what it is for a rule to be 'in force' that does not presuppose an additional, prescriptive rule, the distinction between the two versions of the normativity thesis is not threatened.

[31] Cf. Glüer and Pagin 1999.

R2: $(a)(\text{apply } x \text{ to } a \leftrightarrow a \text{ is } f\ !)$

Unlike R1, R2 prescribes a course of action; it tells me what I ought to do. If meaning something by x requires that I follow a rule such as R2, and if following a rule implies that I judge that I ought to do what it tells me to do, then in order to mean something by x, I must judge that I ought to apply x to a if and only if a is f. Furthermore, if I judge that I ought to apply x to a if and only if a is f, and this is a genuinely normative judgement, then I will be motivated to apply x to a if and only if a is f. Thus, if meaning something by x requires that I follow a rule such as R2, then there would seem to be an internal link between meaning something by x and being motivated to apply x to a if and only if a is f. If there are semantic facts, then it must be an objective fact that I ought to apply x to a if and only if a is f. On this view, the 'correct' in 'correctness condition' is being glossed as 'ought to be pursued'. So, if there is a fact of the matter in which uses of an expression are correct, these facts must be both objective and have 'to-be-pursuedness' built right into them. On this assumption, the claim that meaning is normative amounts to this:

Normativity: S means F by $x \rightarrow (a)(S \text{ ought to } (\text{apply } x \text{ to } a) \leftrightarrow a \text{ is } f)$.[33]

---

[32] Kripke 1982, p. 37.

[33] The obligation in the consequent of *Normativity* as presented here assumes that the ought takes narrow scope. The alternative, which is for the ought to take wide scope

Norm-relativity and *Normativity* seem superficially quite similar, although I maintain that they are fundamentally distinct. Their distinctness has gone largely unnoticed, however. Although Kripke clearly uses both principles, he never distinguishes them, and he is not alone in this. For instance, Paul Boghossian characterises the normativity thesis as follows:

> The point is that, if I mean something by an expression, then the potential infinity of truths that are generated as a result are *normative* truths: they are truths about how I *ought* to apply the expression, if I am to apply it in accord with its meaning, not truths about how I *will* apply it. My meaning something by an expression does not guarantee that I *will apply* it correctly; it guarantees only that there is a fact of the matter whether my use of it is correct.[34]

Boghossian vacillates between saying that there must be a truth about how I *ought* to apply an expression, and that there must be a fact of the matter whether my use of it is, or can be described as 'correct'. Similarly, Robert Brandom writes:

> For a [dispositionalist] account to weather [the sceptical] challenge, it must be able to fund a distinction between what is in fact done and what ought to be done. It must make room for the permanent possibility of mistakes, for what is done or taken to be correct nonetheless to turn out to be incorrect or inappropriate, according to some rule or practice. … What is correct or appropriate, what is obligatory or permitted, what one is committed or entitled to do—these are normative matters. Without the distinction between what *is* done and what *ought* to be done, this insight is lost.[35]

Both Brandom and Boghossian seem to conflate *Norm-Relativity* and *Normativity*. I suspect that this is because they both assume that 'correct' is a normative term. Indeed, as Brandom explicitly says, 'correct' and 'appropriate', 'obligatory' and 'permitted' are all 'normative' notions—they all have to do with values, and they can all be interdefined. Thus, to say that some use of an expression is 'correct', on this view, is to say that it is a use that 'ought to be pursued' or something

---

is: S means F by $x \rightarrow (a)(S$ ought to (apply e to $a \leftrightarrow a$ is f)). I have adopted the narrow scope reading because it is only on this reading that the ought can be detached—i.e., it is only on this interpretation that it follows from the fact that you mean *green* by 'green' that you ought to apply 'green' to something green. On the wide scope interpretation, all that follows is that you ought to avoid combining applying 'green' to a with a's being non-green. This principle, together with the fact that a is non-green, does not imply that you ought to refrain from applying 'green' to x. This does not, I think, capture the original intuition.

[34] Boghossian 1989, p. 509.

[35] Brandom 1994*a*, p. 27.

to that effect. Rules such as R1 can be taken to state which uses of an expression are 'correct' and if 'correct' means 'ought to be pursued', even a rule such as R1 could be taken to be inherently prescriptive.

This assumption is intuitively appealing. It is true that 'correct' is sometimes used as a normative term, just like 'good', 'right', 'wrong', and 'bad'. If 'good' is interpreted as a normative term, then if someone judges that something is good, she must be motivated to pursue it quite independently of any of her desires. If 'correct' is a normative term in this sense, then to judge that some use of an expression is correct, I must be motivated to make that use of the expression in question.

However 'correct' and 'good' are not always normative terms. Sometimes, to say that something is good does *not* imply a prescription; rather, it is to say that it meets a certain assumed or accepted *standard*.[36] For instance, if I say 'that's a good knife', I mean, perhaps, that it is sharp and wieldy. 'That's a good knife' does not imply an obligation on its own, such as that I ought to bring it about that more such knives exist. Of course, if I want to chop vegetables with a good (i.e. sharp and wieldy) knife, then the fact that the knife is a good knife implies that I ought to use the knife but only given my desire to chop vegetables. Here, it is the desire to chop vegetables that motivates me to use the knife; the judgement that the knife is sharp and wieldy will not, on its own, suffice to motivate me—as it does not when what I want to do is whisk an egg. In this use of the word 'good', when I say that the knife is good, I say that it meets an implicit or assumed standard, and not that the knife is worthy of pursuit or approval quite independently of any desires. And whether or not the knife meets this standard is a purely descriptive matter.

Similarly, consider theme parks where some of the rides are dangerous for children under a certain height, so that there is a standard, a minimum height that a child must meet in order to go on the ride. This standard may be 'in force' at the park, in the sense that it is accepted and enforced by sanctions, but this does not make the standard itself prescriptive, nor does it imply a further prescriptive rule. However happy the child might be to meet the standard, whether or not she does meet the standard is a straightforwardly non-normative fact—it is simply the fact that she is so many centimetres high. We might say that she is the 'correct' height, but that does not imply that it is a height she ought to achieve quite independently of her desire to ride the roller coaster.

---

[36] Foot 2002.

Boghossian and Brandom both seem tempted to assume that 'correct' is a normative term. That is, they assume that to say that some use of an expression is correct is to say that one ought to make that use of the expression. However, we should resist this temptation. It is important to remember that given the semantic realist picture of meaning, a 'correct' use of an expression is just one that accords with a rule such as R1, a rule which specifies the conditions under which an expression applies (i.e. refers to, is true of, denotes). To say that some use of an expression is correct is, on this view at any rate, to say that it *refers to* or is *true of* the thing to which it has been applied. If we keep this firmly in view, it no longer makes sense to gloss 'correct' as 'ought to be pursued' but rather as 'meets a standard'. We do not normally think that '*x* refers to *a*' should be understood to prescribe the application of *x* to *a*. Assuming the non-normative interpretation of 'correct' mandated by semantic realism, it should be clear that meaning could be norm-relative without being normative.

To make the distinction more vivid, it might help to consider the difference between saying that, for a speaker to mean something by an expression, she must follow a prescriptive rule as opposed to a non-prescriptive rule. A non-prescriptive rule merely states the conditions under which a term applies. For instance, a non-prescriptive rule such as R1 for the expression 'green' would be:

R3: (a)('green' refers to $a \leftrightarrow a$ is green)

Suppose that for me to grasp the meaning of 'green', a rule such as R3 must be in force. However, since R3 does not tell me what I ought to do—it does not tell me that I ought to apply 'green' to something if and only if it is green—the fact that something is green does not imply that I ought to apply 'green' to it. Rather, if R3 is in force, what follows is simply that if I apply 'green' to something non-green, the resulting sentence is false. That is, even if my meaning *green* by 'green' implies that R3 is in force, this at most allows us to describe my uses of 'green' as either in accordance with R3 or not in accordance with R3, where the judgement that some use accords with R3 does not in any way imply an obligation to make that use or to be motivated to make that use. In contrast, suppose that for me to grasp the meaning of 'green', I must follow a prescriptive rule. A prescriptive rule for 'green' is:

R4: (a)(apply 'green' to $a \leftrightarrow a$ is green!)

Unlike R3, R4 tells me what to do; it enjoins applying 'green' to something if and only if it is green. Supposing that, to mean *green* by 'green' is to follow R4, and that if it is a fact that I mean *green* by 'green', then

R4 expresses an objective semantic principle. In this case, to mean *green* by 'green' is to be obligated to apply 'green' to something if and only if it is green. If I fail to do so—if I apply 'green' to something non-green, then not only is my statement false, but I have violated a semantic obligation. In uttering a falsehood, I do something that I (semantically) ought not to do.

NORM-RELATIVITY

I have, up to now, sketched an argument from semantic realism to semantic non-factualism. The argument depends crucially on the assumption that meaning is normative, not just norm-relative. In Chapter 7, I will argue that it is not the case that meaning is normative, with the implication that a sceptical argument from the normativity of meaning to non-factualism cannot even get off the ground. A natural objection to my line of argument is that the norm-relativity of meaning is sufficient for the sceptical argument—that a case for non-factualism can be based on the assumption that meaning is merely norm-relative. I will argue, however, that the norm-relativity of meaning is not sufficient to justify scepticism.

First, consider the argument against naturalism. Suppose that meaning is norm-relative, that is, that if a speaker grasps the meaning of a word, then she must follow a rule that tells her the correctness conditions of the word. Why should we suppose that this rules out any naturalistic account of meaning? Kripke's text furnishes us with objections primarily to a dispositionalist theory, which shows that such a theory cannot discriminate between the 'correct' and 'incorrect' uses of a word. However, this simply is the problem of naturalising content, and even if the problem has not yet been solved, we cannot conclude, as the sceptic does, that there *can be no reductive solution*. To defend such a strong claim, the sceptic needs to do more than criticise all the extant reductionist theories—for there are no doubt further reductionist theories that have not yet been articulated. We need an argument why the objections to dispositionalism generalize to all reductive theories. So, the question is whether *Norm-Relativity* can give rise to an *a priori* argument of sufficient generality.

One thought might be that the sceptic can find an argument based only on *Norm-Relativity* by making use of the notion of a rule's being 'in force'. That is, if the sceptic were to assume that for a rule to be 'in force' is for it to be ‘backed up’ by a further prescriptive rule, then the fact that the semantic rules of English are currently ‘in force’ for me would imply that I ought to comply with the semantic rules of English.[37] The resulting argument would not, however, be a genuine alternative to one that assumed *Normativity*. For, on the current suggestion, what it is for a rule (such as R1, above) to be ‘in force’ is for there to be an additional rule according to which a speaker ought to comply with R1. The upshot of such a picture of rule following would simply be *Normativity*. If, in order to understand ‘green’, a rule such as ‘green’ applies correctly to all and only green things must be ‘in force’, and if for such a rule to be in force is for there to be another rule that tells me that I ought to use ‘green’ correctly, then if I understand ‘green’, I ought to apply it to something if and only if it is green. So, even if such an account of rule following is assumed, it will not give rise to an alternative *a priori* argument against semantic realism based only on *Norm-Relativity* and not on *Normativity*. If *Normativity* is false, *Norm-Relativity* on its own would not supply an alternative route to the conclusion that there is no fact of the matter what anybody means.

Normativity uniquely fills another gap in the sceptical argument—in the rejection of anti-reductionism. The anti-reductionist could legitimately maintain that the failure of reductionism should be taken to imply that there is indeed nothing that makes it the case that I mean *addition* by ‘plus’—other than the *sui generis* fact that I mean *addition* by ‘plus’. Why, says the anti-reductionist, should we give up altogether simply because reductionism fails? In order for the sceptic to defend the claim that there is no fact of the matter whether I mean *addition* by ‘plus’, he must be capable of answering this charge. However, if the sceptic assumes that putative meaning facts must be categorically prescriptive, he can argue that to assume *sui generis* facts about meaning is to assume an ontology that is inherently queer. It should be clear that these arguments work only if we assume that meaning is prescriptive, because it is only if meaning is prescriptive that facts about what I mean would need to be action-guiding. If meaning is prescriptive, then meaning properties would also have to have to-be-pursuedness built into them, and hence they too would be inherently queer. In this way, using *Normativity*, the sceptic would be able to rule out both reductionist and anti-reductionist accounts of the fact as to what I mean. Although these arguments make some highly controversial assumptions, they at least supply the sceptic with some *a priori* reason to maintain that there is no fact of the matter what we mean.

---

[37] Thanks to an anonymous referee for making this suggestion.

Are there other options available to the sceptic? There are, no doubt, other ways to defend a preference for reductionism. For instance, the sceptic could argue, along with many other committed reductionists, that the reductionist project is justified by the impressive results of the natural sciences. Since these impressive results give us very good reason to believe in the existence of physical objects, causes, and functions, but no good reason to believe in meanings, the sceptic might argue, we ought to seek the truth makers of our meaning statements in the physical, causal, or functional facts. The failure of such a project would then reveal that there really is no fact of the matter what anybody means, since we have so far found no truth makers for our meaning statements in the physical, causal, and functional facts.

There are several reasons why this line of defence of reductionism would not serve the sceptic's needs. First, it would be open to the anti-reductionist to argue that the impressiveness of the results of natural science notwithstanding, we also have an impressive wealth of experience of thinking and talking as if our language and thought have content. If that wealth of experience cannot be reduced to physical facts, then we ought to conclude that facts about what we mean and what we think are irreducible, not that there are no such facts. Moreover, the above mentioned line of defence of reductionism would explicitly conflict with the conclusion of the sceptical argument—that there is no fact of the matter what anybody means by any word or thinks by any thought. For, if there is no fact of the matter what any statement means, and since truth depends on meaning, there will be no fact of the matter whether any statement is true—including the statements that comprise our best scientific theories appealed to in the defence of reductionism. If there is no fact of the matter what any statement means, that is, then the statement that all there exists in the world are physical objects and properties, causes and functions, will have no meaning, and will therefore be neither true nor false.

CONCLUSION

Thus, controversies surrounding motivational internalism aside, assuming that meaning is prescriptive supplies the sceptic with the strongest available argument against all possible accounts of semantic facts and properties. Assuming merely that what a speaker means determines which of her uses is ‘correct’ on a non-normative understanding of ‘correct’ supplies at most specific arguments to specific theories, but no *a priori* argument against all possible theories.

The sceptical conclusion is radical, and the argument for it must be proportionately ambitious. Having presented the only form of the sceptical argument that seems to have some hope of success, I will now turn to the sceptical solution—the solution that embraces the thesis that, given the intuitive picture of meaning and content, there is no fact of the matter what anybody means.

4

Can we do Without Semantic Facts?

After rejecting all candidate theories of meaning, Kripke concludes that:

> There can be no such thing as meaning anything by any word. Each new application we make is a leap in the dark; any present intention could be interpreted so as to accord with anything we may choose to do. So there can be neither accord, nor conflict.[1]

This conclusion is outrageous, even self-refuting: if the sceptical conclusion is true, then there must be a fact of the matter what it means, and if there is no fact of the matter what the sceptical conclusion means, then it cannot be true. How can Kripke even purport to conclude something so nonsensical as this? I think that the only way we can make sense of Kripke's argument is to see it as a *reductio* of semantic realism. More precisely, Kripke urges us to reject one central tenet of realism: the idea that it is an objective, factual matter whether a word has such and such conditions of correct application; that it is a factual matter whether the word 'chicken pox' refers to chicken pox and only chicken pox. In place of realism, Kripke suggests what he calls a 'sceptical solution' to the sceptical problem. On the most plausible interpretation of this view, whether a given use of the expression 'chicken pox' can be called 'correct' does not depend on what it means, or on what its correctness conditions are, but on whether others in the linguistic community would agree in its use. The sceptical solution can be seen as a form of 'social constructivism' about meaning because, according to it, agreement with us is sufficient to justify our saying that a given use of 'chicken pox' is correct, or means *chicken pox*.[2]

---

[1] Kripke 1982, p. 55.

[2] This characterization of Kripke may be jarring. He is, I take it, a social constructivist in the sense that he takes ascriptions of meaning and content to be judgement-dependent. In contrast, what one might call a communal dispositionalist is someone who holds that the totality of dispositions in the community (or some subset of them) would collectively determine which uses of an expression are 'correct' in that the correct use would be the

The primary question of this chapter is whether the sceptical solution removes or avoids the apparent incoherence of the sceptical solution. To this end, I will discuss two subsidiary questions. The first question is whether the sceptical conclusion is indeed incoherent. I will argue that it is. The second question is whether the sceptical solution is tenable. I will argue that it is not.

THE SCEPTICAL SOLUTION

Kripke claims, more boldly than Wittgenstein ever does, that 'Wittgenstein holds, with the sceptic, that there is no fact as to whether I mean plus or quus.' This is what has come to be called a 'non-factualism' about meaning. The non-factualist claims that for all speakers, *S*, meanings, *F*, terms, *x*, propositions, *p*, and sentences, *s*, there is no fact of the matter whether *S* means *F* by *x*; there is no fact of the matter whether *S* means *p* by *s* and there is no fact of the matter whether *S* believes that *p*.

Non-factualism is alleged to be an unforeseen consequence of what Kripke calls the 'realistic or 'representational' picture of language—the picture that gives rise to the sceptical problem—which Kripke subsequently urges us to reject. In Kripke's words:

> In order for Wittgenstein's sceptical solution of his paradox to be intelligible, the 'realistic' or 'representational' picture of language must be undermined by another picture ... the paradox ... drives an important final nail (perhaps the crucial one) into the coffin of the representationalist picture.[3]

By the 'representational picture', Kripke means what I have been calling 'semantic realism'. That is, the paradox is said to arise if we assume that 'a declarative sentence gets its meaning by virtue of its *truth conditions*, by virtue of its correspondence to facts that must obtain if it is true.'[4] This is slightly inaccurate, however. The first part of the sceptical argument, which led to the paradox, considered not declarative sentences, but the meanings of subsentential expressions—words such as 'plus'. In fact, the sceptical argument takes issue with the very idea that any word has conditions of correct application that derive from its

---

use that all or most members of the community would be disposed to make. I think that the judgement-dependence of semantic ascriptions is crucial to Kripke's sceptical solution. I will defend this as the most plausible interpretation presently.

[3] Kripke 1982, p. 85. [4] Ibid., p. 72.

meaning. The paradox arises because we can give no adequate account of what makes it the case that a word has a determinate correctness condition, or a determinate extension. The sceptical argument seems to attack, not just an optional feature of some accounts of meaning, but the platitudinous core of most accounts; the argument drives a final nail into the coffin of the intuitive view that, for a word to have meaning, it must apply correctly under some possible circumstances rather than others. To make matters worse still, the arguments, as Kripke frequently points out, apply just as well to mental representation as to linguistic representation. For instance, Kripke says:

> [T]he paradox of the second part of the *Investigations* constitutes a powerful critique of any idea that 'mental representations' uniquely correspond to 'facts', since it alleges that the components of such 'mental representations' do not have interpretations that can be 'read off' from them in a unique way. So *a fortiori* there is no such unique interpretation of mental 'sentences' containing them as 'depicting' one 'fact' or another.[5]

Roughly, then, the sceptical argument can be summarised as follows. Kripke's sceptic begins by undermining the idea that subsentential expressions and related mental representations have any determinate conditions of correct application. He argues that nothing 'fixes' a determinate extension for a thought or a word grasped by a speaker, so no speaker ever means anything by any word. Given that thoughts and words can have no representational content, Kripke's Wittgenstein then goes on to say that mental representations and sentences fail to have truth conditional content as well.

Abandoning our intuitive picture is an undeniably radical move.[6] It seems to be to deny something that everyone accepts; it seems to be denying commonsense. What could be more ordinary than ascribing truth conditional contents to sentences and beliefs? Recall my exchange with the doctor concerning my son's health. In that exchange, I assumed that when the doctor said 'Vikram has chicken pox', he meant *that Vikram has chicken pox*. If statements of the form 'Dr Barnett means *chicken pox* by "chicken pox" ' do not describe—truly or falsely—the contents of Dr Barnett's mind, then what sense can be made of our everyday discourse about meaning and our everyday practice of belief ascription?

Kripke concurs that Wittgenstein does not wish to deny that when we speak of people as meaning something by their words, or as having

[5] Ibid., p. 85.

[6] Ibid., pp. 70–1.

beliefs, we do so 'with perfect right'.[7] Furthermore, he says, 'we do not even wish to deny the propriety of an ordinary use of the phrase "the fact that Jones meant addition by such-and-such a symbol", and indeed such expressions do have perfectly ordinary uses.'[8] The sceptical solution thus aims to save our ordinary ascriptions of meaning and associated uses of 'facts' and 'true'. These must come out as legitimate, even if there are no determinate meanings, and no facts to which our meaning ascriptions might correspond.

The sceptical solution starts with a supposed description of our linguistic practices—a description of the conditions under which we take ourselves to be entitled to use meaning ascriptions such as 'Dr Barnett means *chicken pox* by "chicken pox" '. Kripke's Wittgenstein considers primarily the situation in which a teacher would ascribe a meaning to a child, and suggests that she would do so if the child has demonstrated, in a sufficient number of cases, that she uses expressions as the teacher deems correct. This reflection leads Kripke to suggest that similar conditions apply across the language. Thus, the description of our practice of ascribing meanings culminates in the following picture. Smith is entitled to say that Dr Barnett means *chicken pox* by 'chicken pox' if she judges that Dr Barnett's uses of the words 'chicken pox' agree with her own. Dr Barnett, of course, is entitled to use the words 'chicken pox' if he has used it enough times in a way that is deemed acceptable by others in his linguistic community, and if he has acquired the confidence to say 'now I can go on!'. Dr Barnett is even entitled to say 'I mean *chicken pox* by "chicken pox" ' under these conditions. However, Smith is not bound to accept Dr Barnett's authority on the matter. If Smith judges that Dr Barnett's uses do not agree with her own, she is entitled to say that Dr Barnett does not mean *chicken pox* by 'chicken pox'.

Herein lies the argument against private language: if Dr Barnett is conceived of in isolation, independently of any interaction with any linguistic community, there is no possibility of comparison between Dr Barnett's uses and anyone else's. As a consequence, so long as Dr Barnett is considered in isolation, we cannot legitimately say that he means *chicken pox* by 'chicken pox'. On this view, the point of ascribing meanings is not to state something that is true or false in virtue of what is or isn't in someone's head. Rather, the point of ascribing meanings and beliefs is *merely* to signify membership in a linguistic community. To say that Dr Barnett means *chicken pox* by 'chicken pox' is simply to

[7] Kripke 1982, p. 69.

[8] Ibid.

say that he uses 'chicken pox' as we do, and the point of saying this is to signify that Dr Barnett is one of us.

Membership in the linguistic community is always provisional. Kripke insists that we focus not on the conditional that if Dr Barnett means *chicken pox* by 'chicken', he *will* apply the word 'chicken pox' and the concept *chicken pox* to something if and only if it is chicken pox. For, as the sceptic has argued, Dr Barnett's meaning *chicken pox* by 'chicken pox' does not imply that he will use it correctly—since he could always make a mistake. Nor do we have any reason to assume that he will continue to use 'chicken pox' as he did in the past, since we have no reason to assume that he is following a rule. If by saying that Dr Barnett means *chicken pox* by 'chicken pox' we were ascribing to him a rule that he follows, our ascription would be wholly illegitimate. Instead, Kripke says that we focus on the assertion conditions of the *contrapositive*. We are committed to revoking the attribution of meaning if Dr Barnett does not continue to use the word 'chicken pox' as we do. If Dr Barnett does not use 'chicken pox' in the way that we deem correct, we will judge that he does not mean *chicken pox* by 'chicken pox'. Of course, if we ascribe to Dr Barnett the concept of *chicken pox*, we do not expect him to come up with bizarre uses of 'chicken pox'. But, on the Wittgensteinian view, all this amounts to is our commitment to revoke our ascription of the concept if Dr Barnett uses the words 'chicken pox' bizarrely enough. The point of ascribing meanings is thus to certify membership in the linguistic community. Given that this is the point, it is clearly pointless to ascribe meanings to individuals conceived of entirely in isolation.[9]

THE INCOHERENCE CHARGE

Is the sceptical conclusion fundamentally incoherent? One reason to think so is given above: if there is a fact of the matter what the sceptical conclusion means, then it cannot be true, and if it is true, there must be a fact of the matter what it means. If Kripke's sceptic claims that no sentence has a truth condition and if truth value depends on truth conditions, then no sentence can be true, including the statement of the sceptical conclusion itself.[10] The sceptical conclusion appears to be self-refuting.

---

[9] Cf. Kripke 1982, pp. 90–5.

[10] This is, put very briefly, one of the arguments made in Wright 1984.

In anticipation of just such a difficulty, Kripke suggests that Wittgenstein urged not only the abandonment of the truth conditional picture of meaning, but the closely allied correspondence theory of truth as well. For instance, Kripke says, 'if Wittgenstein is right, we cannot begin to solve [the sceptical problem] if we remain in the grip of the natural presupposition that meaningful declarative sentences purport to correspond to facts; if this is our framework, we can only conclude that sentences attributing meaning and intention are themselves meaningless.'[11] Instead, Kripke suggests that Wittgenstein 'accepts the "redundancy" theory of truth: to affirm that a statement is true (or presumably, to precede it with "It is a fact that ...") is simply to affirm the statement itself, and to say it is not true is to deny it ('*p*' is true = *p*)'.[12] Given a redundancy theory of truth, if the sceptic can be entitled to say anything at all, he is equally entitled to say that his assertions are true, for to say that his assertions are true adds nothing further to the assertions themselves.

Boghossian argues that the instability of semantic non-factualism reemerges if we only dig deeper. He claims that there is a tension between endorsing a non-factualist thesis about any subject matter, and holding a redundancy theory of truth, or any other similarly 'deflationary' conception of truth.[13] According to Boghossian, a deflationary conception of truth is characterised 'by the claim that there is no such thing as the property of truth, a property that sentences or thoughts may enjoy, and that would be named by the words "true" or "truth".'[14] Rather, the deflationist maintains that the truth predicate is a logical device that enables indirect assertion. What all deflationary conceptions of the truth predicate have in common is the view that appending 'is true' to a sentence or proposition adds little of substance to the sentence or proposition. Boghossian takes this view to imply that the predicate 'is true' can be appended to any sentence that is declarative in form and meaningful. That is, he assumes a minimalism about truth aptness—what makes a sentence apt for applying the predicate 'true' is simply its syntax.[15] If this is correct, however, minimalism about truth and truth aptness seems to be in tension with the non-factualist thesis that our semantic judgements are neither true nor false. The sceptic does not want to deny that meaning ascriptions are declarative and significant, only that they lack truth value. Yet, on a deflationary conception of the truth predicate, it makes no sense to say, of some declarative sentence, that it is assertable but nevertheless lacks truth value. On a deflationary conception of the truth predicate, ' … is true' can be predicated of any assertable declarative sentence whatsoever.[16] Thus, Boghossian concludes, if the sceptic claims that the judgements of our semantic discourse lack truth value, he must assume a 'robust', that is, non-deflationary conception of truth.

---

[11] Kripke 1982, pp. 78–9.

[12] Ibid., p. 86.

[13] There are several forms of deflationism, which I will discuss in more detail in the next section. For the time being, I rely on Boghossian's characterization of deflationism.

[14] Boghossian 1990, p. 163.

[15] For more on the distinction between minimalism about truth aptness and minimalism about truth, see Jackson, Oppy, and Smith 1994.

Is the best response to Boghossian to give up on the minimalist notion of truth aptness, by denying that any declarative, assertible sentence is apt for truth? It does not seem as though this option is open to the semantic non-factualist. For, the semantic non-factualist cannot use the meanings of sentences to determine whether or not they are truth apt, since he claims that there is no fact of the matter what any sentence means. Given the argument for semantic non-factualism, he cannot appeal to standards of acceptance, or semantic platitudes to determine whether a given sentence is truth apt. For the sceptical arguments seem to show that there can be no fact of the matter what standard is followed or endorsed by an individual or in a community. And for any expression of putative platitudes governing truth aptness, there will be no fact of the matter what platitude is expressed. It seems most plausible, therefore, to suppose that the semantic non-factualist needs to be committed to minimalism about truth aptness. The destructive power of the sceptical argument is such that nothing else is left to do the work.

At the same time, it seems as if the sceptic cannot assume a robust conception of the truth predicate, for on such a conception, non-factualism about meaning seems to globalise to a non-factualism about anything whatsoever, and global non-factualism seems to be incoherent. Wright argues that meaning non-factualism globalises on the basis of what he calls the meaning–truth platitude, that is that the truth value of a declarative sentence depends only upon its meaning and the way the world is in relevant respects.[17] Wright goes on to argue that since the truth value of a declarative sentence depends only on its meaning and the way the world is, non-factuality in one of the determinants of truth value will induce a non-factuality in the outcome. If there is no fact of the matter what some given sentence means, then there will be no fact of the matter whether it is true. Since the semantic non-factualist maintains that there is no fact of the matter what any sentence means, it turns out that all sentences lack truth value. Global non-factualism seems to be self-refuting because, if every sentence lacks truth value, then the statement of global non-factualism itself lacks truth value, as do the sentences that comprise the sceptical argument, and the semantic non-factualist conclusion. Thus, the sceptic must assume a robust conception of the truth predicate, but, in so doing, he is forced into a position in which he cannot claim that semantic non-factualism itself is true. As Wright picturesquely concludes, 'to sustain the sceptical argument is to uncage a tiger whose depredations there is then no hope of containing.'[18]

---

[16] Boghossian 1990; Wright 1984. For more on the incoherence of the sceptical conclusion, see Read 1995.

[17] Wright 1984, p. 769.

Boghossian, like Wright, maintains that meaning non-factualism globalises, but presents a different argument for the incoherence of semantic non-factualism. First, Boghossian makes the point mentioned above, that the semantic non-factualist must assume a robust conception of the truth predicate in formulating non-factualism. Assuming such a conception of the truth predicate, Boghossian argues, commits the semantic non-factualist to truth being a robust, language-independent property, for which the predicate stands. On the robust conception of the truth predicate, whether or not a sentence is true cannot be anything but a factual matter (since it cannot but be a factual matter whether the sentence has or fails to have the robust property of being true). However, the non-factualist maintains that it is not a factual matter what any given sentence means. Since truth conditions depend on meaning, if it is a non-factual matter what any sentence means, it is a non-factual matter what truth conditions a sentence has. And since truth value depends on truth conditions, if it is a non-factual matter what truth condition a sentence has, it will be a non-factual matter what truth value any sentence has. Since this applies to all sentences, it turns out that the semantic non-factualist must concede that any sentence of the form 'sentence *s* is true' is neither true nor false, and this conflicts with the robust conception of the truth predicate, according to which, it must be a factual matter whether the sentence 'sentence *s* is true' is either true or false.[19]

Does the self-refutation charge stick? Bob Hale argues that it does not, because both Wright's and Boghossian's arguments assume that the non-factualist is committed to a robust conception of the truth predicate throughout, whereas, as Hale argues, the non-factualist should be entitled to wield *both* a robust and a deflationary conception of the truth

[18] Wright 1984, p. 771.

[19] Boghossian 1990.

predicate. Hale raises two telling objections against Boghossian's argument. First, Hale points out that although the semantic non-factualist may be committed to a robust conception of the truth *predicate* in the formulation of semantic non-factualism, he is not thereby committed to there being a robust *property* that answers to that predicate. The question whether the extension of the robust truth predicate is empty must remain open. Thus, to assume a robust conception of the truth predicate does not automatically commit one to there being a robust property of truth, nor to the view that it must be a factual matter whether any sentence (or the statement it expresses) possesses that property. Moreover, Hale argues, Boghossian places weight on the assumption that the truth of the sentence '*s* is true' must itself be genuinely factual—that the truth predicate applied to '*s* is true' must be conceived as robust. However, the semantic non-factualist can maintain that all metalinguistic attributions of 'true' or 'false' are to be understood in deflationary terms. Why, Hale asks, must the semantic non-factualist be forced to choose between robust and deflationary conceptions of the truth predicate?[20]

In an attempt to deflect such an objection, Boghossian argues that the concept of truth must be univocal. And, to be sure, there are some good reasons to favour a univocal conception of the truth predicate. For instance, given a plurality of truth predicates, it becomes difficult to assess the validity of mixed arguments in which the various premises are apt for different kinds of truth. However, it is not entirely clear that the semantic irrealist needs to endorse a pluralism about *truth* in order to answer the self-refutation charge—it is sufficient for the semantic irrealist to operate with a plurality of truth *predicates*. If the sceptic is entitled to a plurality of truth *predicates*, the argument could be made much more plausible. That is, the sceptical argument Kripke attributes to Wittgenstein could be understood as follows. Kripke's sceptic argues that *if* we assume that meaning is truth conditional, and if we assume that truth is robust, *then* we find we are faced with the paradoxical conclusion that there can be no (robust) fact of the matter what anybody means. In presenting this argument he must, of course, assume the conception of truth that he claims leads to a paradox. He then urges us to abandon that conception of truth and, in the formulation of the sceptical conclusion, suggests that it ought to be replaced with a deflationary conception.[21] Indeed, when we see it this way, the sceptic really endorses a univocal, *deflationary* conception of

[20] Hale 1997, pp. 377–9.

[21] Kripke 1982, p. 86.

the truth predicate, since the assumption of the robust conception was entirely in service of the *reductio*. Given that this is so, any unfavourable consequences of the robust conception of the truth predicate—such as the tension that Boghossian hits upon—would be welcome to the meaning sceptic, since it only goes to show just how dire the situation is for the truth conditional picture of meaning and the closely allied robust conception of truth.

The above considerations tell against Wright's argument as well. For, even if, assuming a robust conception of the truth predicate, the sceptic is committed to global non-factualism, this does not prevent the sceptic from maintaining that his own conclusion, and the sentences that comprise the arguments leading up to it, are true in a deflationary sense. Indeed, Wright's globalising argument could also be seen as grist to the sceptic's mill. The sceptic argues that the truth conditional picture of meaning and the robust conception of the truth predicate yield the paradoxical conclusion that there is no fact of the matter what any sentence means. Still assuming the truth conditional picture of meaning and the robust conception of truth, Wright shows that non-factualism about meaning entails a self-refuting global non-factualism. All the more reason, the sceptic might say, to jettison the truth conditional picture of meaning and the robust conception of the truth predicate that goes along with it. Instead, the sceptic might continue, we ought to assume only a deflationary conception of the truth predicate and an alternative picture of meaning such as he puts forward. The non-factualist can then go on to say, without risk of self-refutation, that the non-factualist thesis itself is true (in the deflated sense) although no sentence is true (in the robust sense). The non-factualist should be able to avoid the self-refutation charge by endorsing a thoroughgoing deflationism about truth.

SEMANTIC NON-FACTUALISM AND DEFLATIONISM

Is the marriage between semantic non-factualism and deflationism as rosy as it is beginning to look? I will argue that it is not, because there is a fundamental tension between deflationism and semantic non-factualism: the deflationist seems to be committed to some kind of meaning equivalence—that is, between an arbitrary sentence 's' and the statement that 's' is true. This commitment seems to stand in tension with the non-factualist's denial that there is such a thing as meaning. For if no sentence has a meaning, two sentences can hardly be equivalent in meaning. Furthermore, the deflationist seems to be committed to the view that we can use names or descriptions to designate sentences—which we do when we say, for example, 'Fermat's last theorem is true' or 'the first sentence in Wittgenstein's *Tractatus* is true'. This stands in tension with the non-factualist's denial of reference. If there is no fact of the matter what any expression refers to, there is no fact of the matter whether 'Fermat's last theorem' refers to Fermat's last theorem, rather than something else or nothing at all.

The non-factualist's most obvious reply to the objection presented above is to argue that a deflationary conception of truth need not presuppose meaning and reference in the semantic realist's sense, but instead to appeal to some variant of meaning and reference such as is supplied by the 'sceptical solution'. Although the non-factualist denies that sentences have *truth conditions* and words or expressions have *correctness conditions*, nevertheless, sentences and words can be said to be meaningful in some weaker sense. In accordance with Kripke's sceptical solution, for instance, Jones can say that when Smith uses 'Fermat's last theorem' it refers to Fermat's last theorem if Jones judges that Smith uses 'Fermat's last theorem' as she is inclined to do. In the next section I will argue that this account of meaningfulness is untenable. In the remainder of this section I will argue that we cannon give a satisfactory deflationary characterization of truth if we assume a non-factualism about truth conditions. Although I cannot discuss every deflationary conception of truth, I will discuss those that seem most likely to combine well with semantic non-factualism. I will argue that each of these stands in tension with semantic non-factualism.

Consider, first, the redundancy theory of truth. This is the form of deflationism which, according to Kripke, Wittgenstein endorsed. The redundancy theory was initially put forward by Frank Ramsey, who says:

> Truth and falsity are ascribed primarily to propositions. The proposition to which they are ascribed may be either explicitly given or described. Suppose first that it is explicitly given; then it is evident that 'It is true that Caesar was murdered' means no more than that Caesar was murdered, and 'It is false that Caesar was murdered' means that Caesar was not murdered . . . . So also we can say 'It is a fact that he was murdered' or 'That he was murdered is contrary to fact'.[22]

---

[22] Ramsey 1965, p. 142.

The first point to note is that, on Ramsey's characterization, truth and falsity are primarily ascribed to propositions, and the proposition *that p* is said to be equivalent in content to the proposition *that it is true that p*. Is this consistent with semantic non-factualism? It could be argued that it is. The semantic non-factualist does not deny that there are propositions, nor that two propositions can be equivalent in content, only that there is no fact of the matter what proposition is expressed by any token sentence.[23] So, there seems to be no tension between the redundancy theory of truth and semantic non-factualism.

Nevertheless, the redundancy theory of truth cannot help the semantic non-factualist to avoid the self-refutation charge. The initial charge was that, if the semantic non-factualist holds that no sentence has truth conditions, it seems to follow that no sentence has truth value, including the sentence stating the non-factualist thesis itself. The deflationary conception of truth was meant to allow the semantic non-factualist to claim that his own assertions are true even though no sentence has truth conditions. However, if truth and falsity are ascribed primarily to propositions, and there is no fact of the matter what proposition is expressed by a given sentence, then there is no fact of the matter whether any given sentence is true, including the sentences uttered or written by the semantic non-factualist.

This suggests that the semantic non-factualist's natural ally will be a deflationist who takes sentences as the primary bearers of truth. This fits well with Kripke's suggestion, when he presents Wittgenstein's view, that 'we *call* something a proposition, and hence true or false, when in our language we apply the calculus of truth functions to it. That is, it is just a primitive part of our language game, not susceptible of deeper explanation, that truth functions are applied to certain sentences.'[24] The thought is that we apply 'true' and 'false' to certain sorts of *sentences*—declarative sentences, as opposed to questions or commands. We *call* such sentences propositions *because* we apply 'true' and 'false' to them, and because we combine such sentences in truth functional compounds. Talk of propositions—or what we *call* propositions—just is talk of sentences to which we typically ascribe truth and falsity, and which we embed in conditionals, conjunctions and other truth functional compounds. To avoid the confusion that might arise if we think of a proposition not just as a special kind of sentence, but as an abstract object, it would make sense to consider forms of deflationism

[23] Thanks to an anonymous referee for this point.

[24] Kripke 1982, p. 86.

which take sentences to be bearers of truth and falsity, as opposed to propositions.[25]

Consider Quine's deflationary conception of truth, according to which 'true' and 'false' are ascribed to sentences, as opposed to propositions or sentence meanings. Like most contemporary deflationists, Quine claimed that Alfred Tarski's schema (T) tells us pretty much all there is to know about the predicate 'true'. Tarski's (T) schema is:

(T) *X is true if, and only if, p*.

Where *X* is the name of a sentence and *p* is the sentence itself or a translation of it. So, for example, let *p* be the sentence:

Bananas are yellow

And let us construct the name of that sentence by putting quotation marks around it, as in 'bananas are yellow'. Substituting into the (T) schema, we get:

'bananas are yellow' is true if, and only if, bananas are yellow.

The important point, for Quine, is the rule for substituting into the T-schema: *X* is the name of a sentence, and *p* is the sentence named or a translation of it. According to Quine, this captures the function of the truth predicate almost entirely. As Quine puts it, 'ascription of truth just cancels the quotation marks. Truth is disquotation.'[26] To say that 'bananas are yellow is true' is another way of saying that bananas are yellow—it is not to ascribe truth to a sentence, but yellowness to bananas. This rules out certain combinations of substitutions into the T-schema. Given that 'bananas are yellow' is the name of the sentence,

bananas are yellow,

and given that the sentence

oranges are orange

is not the sentence named nor a translation of it, then

'bananas are yellow' is true if and only if oranges are orange

is an incorrect substitution instance of the T-schema (though it may nevertheless be true).

---

[25] Horwich's (1990*a*) version of deflationism, called 'minimalism', is well-known, but I will not discuss it further here. The reason is that, like Ramsey, Horwich takes truth and falsity to be primarily ascribed to propositions, and this in itself makes minimalism of little use to the semantic non-factualist.

[26] Quine 1992, p. 80.

The disquotational conception of truth is likely to sit better with a non-factualist conception of meaning than a conception which takes propositions as truth bearers. Quine himself claimed that there is no fact that will distinguish between alternative, but empirically adequate translation manuals, and no determinate reference either, so the disquotational conception of truth should be compatible with these claims. However, there is a problem with the suggestion that truth and falsity are primarily applied to sentences. The problem is that some sentences contain indexicals (such as ‘here’ and ‘now’), others are ambiguous, or vague. In these cases, the effect of the predicate ‘is true’ seems not to disquote, but to translate, so that the sentence used on the left-hand side of the biconditional is equivalent in meaning to the sentence mentioned on the right-hand side. For example, suppose that Cathy said ‘I am standing’ in the kitchen, yesterday. Under what conditions is the predicate ‘true’ correctly appended to this sentence? What we want to say is:

(1) The sentence ‘I am standing,’ as uttered by Cathy yesterday in the kitchen is true if and only if Cathy was standing when she uttered the sentence yesterday in the kitchen.

However, if ‘is true’ is merely a device of disquotation, the effect of predicating ‘is true’ is to remove the quotation marks, and we get:

(2) ‘I am standing’, as uttered by Cathy yesterday in the kitchen, is true if and only if I am standing.

How are we to understand (2)? On the right-hand side, the sentence ‘I am standing’ is *used*, and when an indexical sentence is used, we usually interpret it relative to its context of use. In this case, suppose that it is used by Anandi, now, while she is sitting at her computer. So, on the right-hand side, we would interpret ‘I am standing’ to mean that Anandi is now standing (which is false). On the left-hand side, we have stipulated that it is the sentence that was uttered by Cathy yesterday in the kitchen. Usually, we would interpret that token of ‘I am standing’ relative to the context of Cathy as speaker, in the kitchen, yesterday. In this case, it is true. Thus, given the way we usually interpret indexicals, the interpretive context for the token of ‘I am standing’ that appears on the right-hand side in (2) differs from the interpretive context for the token of ‘I am standing’ that appears on the left. Since these interpretive contexts differ, (2) is an intuitively incorrect instance of the (T) schema.[27] How is it to be ruled out?

---

[27] If both speakers are standing at the times of their respective utterances, the biconditional would come out true. However, there is still the point that although ‘I am

The natural assumption is that when Cathy says 'I am standing' in the kitchen, she *refers* to Cathy, and says, of Cathy, *that she is standing*. We want to say that even if 'true' is predicated of sentences, it is predicated of sentences in virtue of what they mean, and the conditions under which a sentence is true or false is determined by what the sentence means in its context of utterance. However, this account of why (2) is an incorrect use of 'true' is not available to the semantic non-factualist. It must be ruled out some other way.

Quine has suggested a solution.[28] Quine's solution to the problems posed by indexicals for his account is to say that 'true' applies to 'eternal sentences' only, and sentences containing indexicals can be transformed into eternal sentences by paraphrase to sentences with determinate times, dates, locations, and persons, or descriptions of persons.[29] That is, in the paraphrase of the sentence 'I am standing' as uttered by Cathy, yesterday, in the kitchen, 'I' could be replaced with 'Cathy', and facts about spatio-temporal location are added to yield the resulting paraphrase: 'Cathy is standing on May 2, 2006, at 5:00pm, in her kitchen.' It is this sentence that can be ascribed a truth value, not the sentence 'I am standing'.

One might think, immediately, that a thoroughgoing semantic non-factualist cannot help himself to Quine's notion of a paraphrase of the sentence 'I am standing', since paraphrase, intuitively, implies equivalence of meaning between the paraphrase and the sentence it paraphrases. If there is no fact of the matter what any given sentence means, there can be no fact of the matter whether two sentences have the same meaning—which is our ordinary notion of paraphrase. It may even seem odd that Quine helps himself to this notion, given his acceptance of the indeterminacy of translation. However, Quine argues that a perfectly adequate notion of paraphrase can be given in terms of the behavioural dispositions of a speaker, at least for what he calls 'occasion sentences'. He says, 'What relates such a sentence to its equivalent is simply a coinciding of dispositions: we are disposed to assent to both sentences in the same circumstances.'[30] If we take them to be occasion sentences, then on this view, the sentences 'I am standing', as uttered by Cathy in the kitchen yesterday, and 'Cathy is standing on May 2, 2006, at 5:00pm, in her kitchen' are equivalent because I am disposed to assent to the two in the same circumstances.

---

standing' is the same sentence, it refers to different people in different contexts, and this is what gives rise to failures of equivalence.

[28] Quine 1969. [29] Quine 2001. [30] Quine 1993, p. 86

The question is: am I disposed to assent to both sentences in the same circumstances? It seems not, because, if I am ignorant of the time or the date, while in the kitchen with a standing Cathy, I would assent to the sentence, 'I am standing' were she to utter it, but not to the sentence, 'Cathy is standing on May 2, 2006, at 5:00pm, in her kitchen.' Indeed, I would be prepared to assent to the sentence 'I am standing' as uttered by the speaker of that sentence if the speaker is standing. I need no information regarding who the speaker is in order to assent to the sentence. In contrast, I would not be prepared to assent to 'Cathy is standing on May 2, 2006, at 5:00pm, in her kitchen' unless I had evidence that the person before me is Cathy, as well as evidence of the date, the time, and the location.

A defender of Quine might suggest instead that the sentence 'I am standing' is equivalent to 'the speaker who uttered the sentence "I am standing" was standing at the time of utterance'. Arguably, we would be disposed to respond to these under exactly the same circumstances. However, even for this pair of sentences, the circumstances under which we would assent to each come apart. For example, suppose that Krister, while seated, says 'the speaker who uttered the sentence "I am standing" was standing at the time of utterance'. I am disposed to assent to this sentence because I trust Krister to tell me the truth. However, no matter how much I trust Krister, I will not be prepared to assent if he says 'I am standing' under precisely the same conditions—that is while he is seated. The two sentences are not stimulus equivalent.

Leaving Quine, for the moment, a semantic non-factualist such as Kripke's Wittgenstein might suggest that what makes one sentence a paraphrase of another is that we agree to treat them as alike in meaning. Thus, the non-factualist would say, we agree to treat 'I am standing' and 'Cathy is standing on May 2, 2004, at 5:00pm, in her kitchen' as the same, and this agreement is what makes the sentences paraphrases of one another. But what does this agreement consist in? If it is our dispositions to accept or assent to both under the same conditions, then the two sentences will not come out as equivalent, as I have argued above. The alternative, which is to say that we *judge* these two sentences to be equivalent in meaning or that we *believe* that they are equivalent in meaning appeals directly to our beliefs and judgements which, if they are to constitute our agreement, must have content. For our judgement that the two sentences are equivalent to constitute their equivalence, we must be correctly said to judge *that the two sentences are equivalent* rather than *that they are both rather nice*. This presupposes that our judgements have determinate content—something which the semantic non-factualist denies. The notion of paraphrase seems to be a non-starter for the non-factualist.

Hartry Field[31] has also suggested a solution to the problems posed by indexical and ambiguous sentences in his defence of deflationism about both truth and truth conditions. On Field's deflationism, any sentence *s*, as it is understood by a speaker, is 'cognitively equivalent' to the sentence '*s* is true.' '*s*' and '*s* is true' are cognitively equivalent in the sense that a speaker's inferential practices permit indefeasible inference from one sentence to the other. Field maintains that we are entitled to talk about truth conditions, on such a view, without having to suppose that they are 'robust'. I take it that what he means is that we can say that the truth condition of the sentence 'snow is white' is that snow is white, so long as we take this merely to be a consequence of the 'cognitive equivalence' between 'snow is white' and ' "snow is white" is true'. What Field eschews, however, is the thought that truth conditions play an essential role in a theory of what it is for a person to understand a sentence. Thus, he suggests that what constitutes a person's understanding of a sentence might include the following: (1) verification conditions of the sentence, or conditions under which the speaker would take the sentence to be verified; (2) the conceptual role of the sentence, that is its inferential relations, relations to other beliefs and thoughts, and its causal role in producing actions; and (3) correlations between the person's assent to the sentence and objects or states of affairs in the external world. Like Kripke's Wittgenstein, Field maintains that none of this, even taken together, can determine a unique truth condition for the sentences a speaker understands. All the facts about the verification conditions a person associates with a sentence, its conceptual role and its correlation to external reality are admissible as part of an account of what it is to understand a sentence, but all of these facts do not determine a truth condition for the sentence, as it is understood by the speaker.

Field argues that indexicals and ambiguity pose no problem for a deflationism formulated in the way that he prefers. This is because, he argues, his concept of truth is a concept of 'true-as-a-speaker-understands-it' where the predicate 'true' applies to a sentence (or utterance) *as it is understood by a speaker*. Take the sentence, 'she is going to Pisa'. According to Field, if I hear that sentence uttered, there is a particular way that I process it, such that I associate it with a name, such as, Sheila, or a description, such as 'the woman I met on the airplane'.

---

[31] Field 1994.

At any rate, there is a way that I process this sentence such that the indexical in 'she is going to Pisa" is associated with the value 'Sheila', and hence is understood as 'Sheila is going to Pisa'. The 'inner' sentence, which is then free of the indexical, is cognitively equivalent to ' "Sheila is going to Pisa" is true', hence, by my inferential practices, I am entitled to assert that 'She is going to Pisa' is true if and only if Sheila is going to Pisa. Ambiguous or indexical words have 'inner subscripts' such that the 'inner analogue' of an ambiguous or indexical sentence is itself neither indexical nor ambiguous.[32]

This is not a satisfactory solution to the problem posed by indexicals and ambiguity, because the solution requires that the way in which an agent 'processes' a sentence will uniquely determine an unambiguous, non-indexical 'inner analogue' to the sentence that is correctly ascribed by a that-clause such as *that Sheila is going to Pisa*. That is, not only must there be a determinate value that I associate with the 'she' in 'she is going to Pisa' but my way of understanding the sentence must legitimate the ascription of the belief *that Sheila is going to Pisa*. The burning question, then, is: what makes it the case that the way a speaker understands a sentence licenses us to say that she understands it to mean *that Sheila is going to Pisa* rather than something else, or nothing at all? Insofar as the non-factualist has accepted that there is no fact of the matter what correctness conditions any of my words or thoughts have, it is not open to the non-factualist to suppose that 'she' as I understand it refers determinately to Sheila.

Field might argue that some causal relations between purely formally specified elements of the speaker's mind determine a unique description of what the speaker understands. He suggests that when I hear a sentence such as 'she is going to Pisa', it causes me to access my mental 'file drawer' on Sheila. It is not necessary for this 'file drawer' to determine a unique referent for 'she' on this occasion, but only to allow us to give an account of who I *regard* the 'she' in 'she is going to Pisa' as referring to. Let us suppose, then, that we have a purely syntactic description of the 'file drawers' in my mind, and a purely formal description of the inferential relations that take me from one file drawer to the next, or from hearing an utterance to opening a file drawer, and which determine which file drawer supplies the value of 'she' when I hear the indexical sentence. Then, upon hearing the utterance 'she is going to Pisa', given the formal description of my mind, we should be able to say that what pops up is some syntactically describable thing, 'squiggle, squoggle', if you will, but not the thought *that Sheila is going to Pisa*. Given the speaker's formally specified inferential practices, we can then say that 'squiggle, squoggle' is T if and only if squiggle squoggle squiggle, but we have not got anywhere near being able to say that the speaker is entitled to think *that She is going to Pisa is true if and only if Sheila is going to Pisa*. Perhaps Field is confident that a full formal description of my mind, given all of the inferential relations, determines uniquely that my thought is *that Sheila is going to Pisa*. Similarly, Brandom maintains that the inferential relations between syntactically described elements of a speaker's mind can determine which ascriptions of 'that-clauses' are correct.[33] However, if such an inferentialist semantics were successful, we would no longer have reason to think that *no* facts determine a unique truth condition for what I mean. For, if an inferential semantics were successful, we would have some reason to think that *inferential* relations determine a unique ascription of truth conditions to my utterances and beliefs.

---

[32] Field 1994, p. 279.

Field seems to want to have it both ways, I suspect, because he thinks that the problem of assigning a correctness condition to a sentence or a belief only arises if we assume that there is a correct way to understand the sentence in a public language, and that the problem does not arise for the 'private' language—that is, the language as it is understood by the speaker.[34] That is, he seems to think that we can uniquely describe the way that a speaker understands a sentence without assigning a truth condition to the sentence as it is understood by the speaker. Yet, this is untenable. If there is a determinate way that I understand the sentence 'she is going to Pisa', then there is a determinate truth condition for the sentence-as-I-understand it. If I understand 'she' as referring to Sheila, then as I understand the sentence, it is true if and only if Sheila is going to Pisa. However, this is incompatible with the view that truth conditions play no part in a theory of understanding. On that view, nothing about verification conditions, inferential relations, causal roles, and so forth, determine a truth condition for a sentence as I understand it. If that is true, then nothing will determine that I understand the word 'she' in 'she is going to Pisa' to refer to Sheila, as opposed to Sheena or Sheba or no one at all, and nothing will determine whether, as I understand the sentence, 'she is going to Pisa' is true if and only if Sheila is going to Pisa.

Furthermore, Field rests his deflationary conception of truth on the conception of cognitive equivalence, where two sentences (as a speaker understands them) are cognitively equivalent if the speaker's inferential practices license her to infer, directly and indefeasibly, from one sentence to the other. However, a speaker's inferential practices can themselves be incorrect—by licensing, directly and indefeasibly, an inference from one sentence to another which is not, intuitively, cognitively equivalent with the first. For example, suppose that I understand the 'she' in 'she is going to Pisa' to refer to Sheila. Now, suppose that my inferential practices license me to infer, indefeasibly and directly, that 'she is going to Pisa' is true if and only if John is going to Milan, although I understand 'John is going to Milan' in the ordinary way. If this is the situation, what we would naturally wish to say is that I don't quite grasp the concept of truth. However, this natural thought is not open to Field. If my wayward inferences reveal that I do not grasp the concept of truth, then we must suppose that there are inferences involving 'true' which are correct and others with are incorrect, given the *content* of 'true'. But, according to Field, no fact about me determines what correctness condition I attach to 'true', so 'true-as-I-understand-it' must be as indeterminate in content as 'Sheila'. Either conceptual role does not determine the way in which I understand a sentence, in which case, Field's solution to the problem of indexicals fails, or conceptual role does determine the way in which I understand a sentence, in which case, conceptual role determines truth conditions.

---

[33] Brandom 1994*a*.

[34] Field 1994, p. 278.

Finally, I would like to consider one last form of deflationism: the prosentential theory of truth. The prosentential theory has been endorsed by a number of people, including Belnap, Camp and Grover (together), and Brandom.[35] Like the disquotationalist, the prosentential theorist holds that, for any arbitrary sentence, 's':

's' is true if and only if s

The difference between the prosentential theory and disquotationalism lies in the account of the role of the truth predicate. For the disquotationalist, 'true' is a device of disquotation, whereas for the prosententialist, it is a prosentence forming operator. A prosentence is analogous to a pronoun, such as 'he' or 'she', which inherits its content from a preceding noun. For example, in the sentence 'Elton John wears wigs because he is bald', the word 'he' refers to Elton John, it inherits its referent from the preceding noun. Out of context, 'he wears wigs' does not refer to anyone. The prosententialist claims that sentences of the form ' "s" is true' are similarly devoid of content out of context and must inherit their content from the sentence 's', of which 'is true' is apparently predicated.

---

[35] Brandom 1994*a*; Belnap, Camp, and Grover 1975; Grover 1992.

Now, the prosententialist's talk of 'inheriting content' should suggest a tension with semantic non-factualism. If 'Elton John' has no content—if there is no fact of the matter what 'Elton John' refers to—then the subsequent occurrence of 'he' has no content to inherit. If we think that the sentence 'Phlogiston is smelly' fails to refer to anything, then a subsequent sentence such as 'it is everywhere' likewise fails to refer. Similarly, if there is no fact of the matter what 'bananas are yellow' means, then there is no meaning for ' "bananas are yellow" is true' to inherit. Indeed, Brandom concurs on this point. He says that 'redundancy theories of "true" presuppose the contentfulness of the nonsemantic sentences on which semantic claims are redundant, in order to explain how 'true' ought to be used.'[36]

Nevertheless, it could be argued that something less than truth conditional content renders sentences and words meaningful, and it is this lesser significance that is inherited by sentences containing 'is true'. So, for example, imagine that Field's conditions for understanding held, then ' "s" is true' would inherit the verification conditions, causal relations, conceptual role (and so forth) from 's'. However, if we accept, with the semantic non-factualist, that whatever we might say about verification conditions, conceptual role, and so forth, it will be insufficient to determine the truth conditions of a given sentence, the prosententialist will be unable to rule out bizarre uses of the predicate 'true'. For example, consider the sentence 'grass is green'. According to the semantic non-factualist, there is no fact of the matter whether 'grass is green' means *that grass is green* or *that grass is grue* or nothing at all. All the facts about verification conditions, conceptual role, and so forth are insufficient to determine this. If the sentence ' "grass is green" is true' inherits its verification conditions and so forth from the sentence 'grass is green', what it inherits does not determine whether the sentence means *that grass is green* or *that grass is grue*. Thus, the following statement turns out to be a correct use of 'true':

> 'grass is green' is true if and only if grass is grue

How would Kripke's Wittgenstein respond to this objection? He could argue that although there is no fact of the matter what anybody means, a speaker of English would never say ' "grass is green" is true if and only if grass is grue'. This is just not a permitted move in our language game. And the reason it is not permitted is that—unless we are in the marginal occupation of assessing sceptical arguments—no one would say that 'grass is green' means that *grass is grue*. If you try to say it, others will say that you don't understand the meaning of 'green'. In contrast, I am entitled to say 'grass is green' means that *grass is green* because I would encounter no objections from other members of my linguistic community. Since their approval is precisely what entitles me to use the sentences that I do, then I am entitled to say 'grass is green' means that *grass is green*. Even this may not suffice to entitle me to say ' "grass is green" is true if and only if grass is green', for this requires that I use the word 'true'. However, I have been doing philosophy of language long enough to have read, heard, and used the sentence:

---

[36] Brandom 1994*a*, p. 329.

(3) 'grass is green' is true if and only if grass is green

more times than I care to enumerate, and have encountered no objections to the use of that sentence either. The sceptic could thus argue that my linguistic competence, by the lights of my linguistic community, entitles me to use (3) and other instances of the (T) schema, and thus allows me to make use of a deflationary conception of truth without at any point having to suppose that there is a fact of the matter what any word or sentence means.

It is worth noting, moreover, that this account will yield the correct instances of the (T) schema even for sentences that contain indexicals and ambiguous words. You, my linguistic community, would never let me get away with saying that 'I am standing' (as uttered by Cathy in the kitchen yesterday) is true if and only if I am standing (as uttered by me, now). The instances of the (T) schema to which I am entitled are only those that are acceptable to my linguistic community, and you will make sure that I get it right—or at least, you will make sure that I say what you would say, which, for the Wittgensteinian, is just what it is to get it right.

Despite its promise, however, this suggestion is untenable in conjunction with a non-factualist conception of content. The problem is that according to Kripke's Wittgensteinian solution, we can say that an utterance is 'correct' if other people *judge* that it accords with what they would do, or that they *approve* of it, and we can say that an utterance is 'incorrect' if other people *judge* that it is at odds with what they would do, or they *disapprove* of it. The problem is that the sceptical solution itself seems to presuppose that people can have judgements, beliefs, or attitudes with determinate content. Since the above response to difficulties concerning the use of the predicate 'true' is of a piece with Kripke's sceptical solution, it will be untenable if the sceptical solution is untenable. What remains is to argue that the sceptical solution is itself untenable. This is the purpose of the next section.

THE SOCIAL CONSTRUCTION OF MEANING

Kripke's sceptical solution is a form of social constructivism about meaning and representation. As I am using the term, a social constructivist about representation is someone who claims that what makes it legitimate to call a particular use of an expression 'correct' is that a relevant group of people agree in its use. Take, for example, the word 'square'. According to the social constructivist, if Jones utters the sentence 'this room is square', her use of the word 'square' can be called 'correct' now if and only if we *agree* in her use of the word 'square'. Whether the room is in fact square is irrelevant to whether her use of the word 'square' can be called 'correct'. Moreover, the social constructivist I have described does not hold that agreement in the use of a term determines which future uses *are* correct, as a communal dispositionalist would hold. Rather, the view is that we are justified in *saying* that Jones's use is correct if and only if her use agrees with ours. In contrast, to the realist, there is a reference relation between words and the world: the word 'square' refers to all and only squares. If I say 'origami paper is square', my use of the word 'square' *is* correct if and only if origami paper is in fact square.

Martin Kusch has suggested a nice way to capture the difference between realist and social constructivist accounts of representation, by analogy to the medieval and early modern distinction between two views of divine creation: the 'deistic' and the 'continued-creation' conceptions. On the deistic conception, God simply sets the world in motion and then becomes idle. On the 'continued-creation' conception, God is never idle—He continually meddles in our affairs. The semantic realist is a bit like a Deist about divine creation: after I have learned the meaning of a word from my linguistic community, what I mean by the word determines whether my uses of it are correct, and the community becomes idle. The social constructivist is a bit like a continued-creationist: whether my utterance can be called 'correct' is determined *as the utterance is made* and the community is never idle. Agreement from other competent speakers determines, *case-by-case*, whether my utterances can be called 'correct'.

Kripke argues that if we adopt a picture of representation based on case-by-case communal agreement, our ascriptions of meaning need nothing more than agreement to be justified. That is, if I say 'origami paper is square' and you agree, then you are entitled to say that I grasp the meaning of the word 'square'. This is because what it is for me to grasp the meaning of the word 'square' is not for me to have a rule in my head which I follow, but simply for me to use 'square' in such a way that it secures the agreement of others. Thus, Kripke argues, given communal agreement in the use of words, all of the statements of our commonsense discourse about what people say and what they believe can be legitimately asserted.

The foundation of this picture of representation is the notion of communal agreement in responses or utterances. But what is it for two people's utterances to agree? Intuitively, we think of two people as agreeing in their utterances when what they say has the same *meaning*. So, if I say 'Lo, a bachelor' and you say 'Lo, an unmarried man', what we have said agrees because the word 'bachelor' applies correctly to all and only unmarried men; our utterances have the same correctness conditions, the same meaning. However, the social constructivist cannot assume this intuitive notion of agreement because if she did, whether or not two utterances can be said to agree would depend on what they mean, on what their correctness conditions are. And this, for the social constructivist, gets things back to front: for the constructivist, what an utterance can be said to mean, and whether it can be said to be used correctly, depends on whether two or more people agree in its use.

Thus, the social constructivist faces a challenge: to specify the conditions under which we are entitled to say that two people *agree*. And whatever it is that entitles us to say that two people agree cannot presuppose any representations with determinate correctness conditions—for that would be to concede to the semantic realist that what someone represents is established *prior* to any agreement. In what follows, I will consider what I take to be the most plausible attempts to meet this challenge, and argue that they all fail.

First, consider Kripke's own account. In his description of the assertion conditions for our meaning ascriptions, such as 'Jones means *addition* by "plus",' Kripke makes illicit appeal to *judgements*. For instance, Kripke says that '*Smith* will judge Jones to mean addition by 'plus' only if he judges that Jones's answers to particular addition problems agree with those *he* is inclined to give.'[37] Here, Smith is said to judge *that Jones means* addition *by 'plus'* or *that Jones's uses of 'plus' agree with his own*. Thus, Smith's judgements, we are encouraged to assume, have content. Yet, according to the non-factualist thesis embraced by the sceptic, there is no fact of the matter what anybody means. So, what content are we supposed to take Smith's judgements to have? The sceptic might wish to say that we are entitled to assign a meaning to Smith's judgements if we judge that Smith uses the relevant words as we do. Yet this would hardly do, as he would then make an illicit appeal to *our* judgements. If there is no fact of the matter what anybody means by any word, then there is no fact of the matter what Smith means, and there is no fact of the matter what we mean. Somehow, the sceptic must be able to legitimate meaning ascriptions without an illicit appeal to judgements, beliefs, or any other contentful states.

How is the sceptic to respond? Perhaps, as Scott Soames has suggested,[38] the sceptic could make use of some of the insights of ethical non-cognitivism, which shares with meaning scepticism a non-factualism about a particular region of discourse. A non-cognitivist thesis about our ethical discourse is one according to which the declarative sentences of our ethical discourse do not express judgements or beliefs (cognitive states) but emotions or evaluative attitudes (conative states). According to the non-cognitivist, if I say, 'child abuse is wrong', I do not express the judgement that child abuse is wrong, but a strong negative attitude towards child abuse. Unlike beliefs or judgements, attitudes do not purport to say the way things are; hence we do not normally take them to be the sorts of sentences that can be true or false. If our ethical sentences express attitudes rather than judgements, the non-cognitivist argues, we should also realise that they are not the sorts of sentences that can be true or false.

In a similar vein, the sceptic might say that meaning ascriptions are neither true nor false because they are only masquerading as judgements when they are really something else. The judgements that Jones means *addition* by 'plus' and that Jones's uses agree with our own are both expressions of approval. When we see Jones adding, if he does so in a way that we like, we might say 'fantastic!', 'wonderful!', 'well done!', or we might say 'Jones means *addition* by "plus"'. The latter takes the form of a declarative sentence, but it is just another way of expressing our approval of what Jones does. One expression of approval (a.k.a. the

---

[37] Kripke 1982, p. 91.

[38] Soames 1998a.

'judgement' that Jones's uses of 'plus' agree with our own) could then be said to legitimate another expression of approval (a.k.a. the 'assertion' that Jones means *addition* by 'plus').

Unfortunately, this too presupposes meaning. The sceptical argument is particularly lethal; its conclusion is not just that sentences and beliefs have no truth conditions, but more importantly, that words and concepts have no correctness conditions. Thus it is not just meaning and belief ascriptions that are bereft of content, but any word or sentence whatsoever—including sentences that putatively express approval or disapproval. If the sentence, 'Jones's uses of "plus" agree with mine' is to express approval, it must be meaningful. If 'Jones' does not refer to Jones, then it cannot be used to express approval of what *Jones* does, as opposed to Smith or Baker. If the phrase 'uses of "plus" ' has no meaning, it cannot be used to express approval of Jones's *uses of 'plus'*, as opposed to her gestures or her taste in music. Once again, we need to suppose that some words have meanings in order to legitimate meaning ascriptions. Yet, semantic non-factualism makes such appeal to meanings illicit.

Another of Soames' suggestions is that meaning ascriptions can play the role of bringing people into the community because they are performatives.[39] A performative utterance is one that does not state what is the case, but which brings it about that something is the case. For instance, if I say, 'I promise to return your book tomorrow', I thereby *make* a promise. On some accounts, performatives do not assert propositions, so they do not have truth conditions. Similarly, imperatives have no truth conditions. For instance, an imperative such as 'eat your eggplant!' does not assert a proposition, so 'eat your eggplant!' can be neither true nor false. The sceptic might argue that meaning ascriptions are non-factual because, like imperatives, they lack truth conditions, but that nevertheless they can serve the function of bringing people into the community because they are performatives that bring it about that people are included in the community when they are uttered. Scott Soames, for instance, suggests that meaning ascriptions,

> Could be thought of as performatives, along the lines of 'I hereby name this barge Hilary', or 'I hereby pronounce you husband and wife'. … With this in mind, one might suggest that saying *Jones means addition by* '+', is performative in the following way: to utter such a sentence in appropriate circumstances is to

---

[39] Soames 1998a.

> take Jones into one's linguistic community, to certify him as a competent user of '+', and to license him to use '+' to do what we call 'adding'.[40]

Treating meaning ascriptions as performatives does not, however, obviate the need for correctness conditions to give the meanings of the *words* used in the performatives. Performatives cannot perform unless sub-sentential expressions have some kind of content. Suppose I say 'I hereby name this barge "Hilary".' If this is going to constitute my naming the barge 'Hilary', I will have to mean *barge* by 'barge'—I must refer to the barge as opposed to the sea or my trousers—and I will have to have an intention with the content *that I name this barge 'Hilary'*. None of this is available to the non-factualist about meaning, since on that view, there is no fact of the matter whether 'barge' means *barge* and no fact of the matter whether I intend *that I name this barge 'Hilary'*.

One way the sceptic might avoid these difficulties would be to retreat from the view that it is Smith's *judgement* or attitude which entitles him to ascribe a meaning to Jones, but merely hold that it is a brute similarity between Jones's use of 'plus' and Smith's which makes their uses 'correct' and entitles them to participate in the language game. This version of the sceptical solution is suggested, for instance, by the following passage:

> The entire 'game' we have described—that the community attributes a concept to an individual so long as he exhibits sufficient conformity, under test circumstances, to the behavior of the community—would lose its point outside a community that generally agrees in its practices ... if there was no general agreement in the community responses, the game of attributing concepts to individuals—as we have described it—could not exist. In fact of course there is considerable agreement, and deviant quus-like behavior occurs rarely. ... On Wittgenstein's conception, such agreement is essential for our game of ascribing rules and concepts to each other.[41]

There are two different ways in which the picture of communal agreement could look, one of which Kripke explicitly rejects. The rejected picture is one according to which the correctness of some use of an expression is determined by facts about the dispositions of all of the members of some community. On this view, Jones's use of the word 'dog' at some time, *t*, will be correct at *t*, if and only if everyone in the community (or almost everyone in the community) would use the word 'dog' under the same conditions. This picture is rejected by Kripke on the grounds that it is merely a community wide version of the dispositional view, and would thus succumb to many of the same criticisms as the individual version. He says, 'such a theory would be a theory of the *truth* conditions of such assertions. … The theory would assert that 125 is the value of the function meant … if and only if "125" is the response nearly everyone would give.'[42] Similarly, Wright explains that, according to Wittgenstein, 'it is wrong to think of our understanding of an expression as something determinate, something which beyond a certain point does not grow and which is then applied.'[43] On the communal version of the dispositional theory, the meaning of any given expression would, at any given time, be determinate—it would be constituted by facts about our dispositions at that time.

---

[40] Soames 1998a, p. 322. Soames subsequently criticises this suggestion in Soames 1998a, pp. 323 ff.

[41] Kripke 1982, p. 96.

In contrast, one might think that our agreements and disagreements do not establish a rule for the use of our expressions that extends to cases not yet encountered. Rather, as we go along, we 'extend' the meaning of the word to cover new cases. On this picture, the things that come to fall within the extension of a term—those to which the term is applied—need not themselves be alike in any objective sense, independently of our applying the same term to them. It is not that, as we go along, we apply 'square' to all and only those things that are, objectively, square. Even if there is some pattern discernible in the past applications of an expression, even if we have only applied 'square' to square things in the past, this pattern does not extend to cases that are yet to be encountered. Just because 'square' has been applied only to square things in the past does not determine a rule for the future application of 'square'. As Wright has put it, we do not, as a community, track some property or set of properties such that there is an 'objective' pattern in our use, that is a pattern which obtains quite independently of whether we judge it to obtain. Rather, we are urged to suppose that the community's verdict *decides*, case by case, whether any given term correctly applies. As Wright explains:

> 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 … as something we cannot but know. Wittgenstein's response is to urge … that we should … reject the idea

---

[42] Kripke 1982, p. 111.

[43] Wright 1980, p. 38.

> that … the community goes right or wrong in accepting a particular verdict on a decidable question; rather, it just goes.[44]

To a first approximation, on this account, the Wittgensteinian picture of content goes as follows. We each have dispositions to use expressions in certain determinate ways, unhesitatingly, but blindly—our dispositions do not track properties in such a way that there is an objective pattern discernible in our use, that is patterns that obtain independently of what our judgements would be if we were to investigate fully. Any given individual's use of an expression is correct only if it is acceptable to the rest of the community. If the individual's use is unacceptable to the rest of the community, that use is incorrect. But the dispositions of the community taken together do not track an investigation-independent property either. Therefore, there is no possibility of mistake for the community as a whole. We may all be disposed to call some non-square things 'square'. But so long as we are all so inclined, we have made no mistake. Communal agreement generates 'extensions' of a kind, but never a rule that would determine correct application to novel cases. Even if there is, fortuitously, some discernable investigation-independent pattern in the past applications of 'square', this cannot determine a correctness condition for our subsequent uses of 'square'. It is not just that we are at any time free to change what we mean by 'square', but that so long as we regard ourselves as using 'square' as we did in the past, whatever we choose to call square is square.

This picture of content crucially employs the notion of communal agreement and disagreement as a surrogate for the notions of correctness and incorrectness. Something is included in the extension of a term if (at least) two people agree in calling it by the same name. For instance, some x is included in the extension of 'apple' when Jones and Smith both agree to call x 'apple'. The trouble is, the notion of agreement is as suspect as that of a rule. As Wittgenstein said: 'It is no use, for example, to go back to the concept of agreement, for it is no more certain that one proceeding is in agreement with another, than that it has happened in accordance with a rule.'[45]

The question is: what is it for two responses to *agree*? Clearly, in order for two utterances to agree, they must both be tokens of the same type of utterance, and they must be made under the same circumstances. If

---

[44] Ibid., p. 220.

[45] Wittgenstein 1956, Part VII, 26. I am grateful to Peter Pagin for supplying this quote.

Jones and Smith both say ‘lo, a swan!’ while undergoing precisely the same retinal stimulation, or while confronted with the same birds, then we would say that their utterances agree, whereas if Jones were to say ‘lo, a swan!’ while Smith were to say, ‘lo, a duck’, we would say that their utterances do not agree. Moreover, whatever it is that determines whether Smith’s and Jones’s utterances are tokens of the same type of utterance, it must not simply be that they are both uttered on the same, particular, occasion. If what *makes* Jones’ and Smith’s utterances the same type of utterance is that they are uttered at the same time then *whatever* Smith and Jones were to say on that occasion, they would agree—there would be no prospect of disagreement whatsoever.

The sceptic thus needs a way of settling whether two token utterances are of the same type—utterances need to be assigned to equivalence classes on some basis. There are, of course, several options. One way in which two utterances could agree is in having the same meaning. Intuitively, when we think of Jones and Smith being in agreement, this is the sort of agreement we have in mind. Thus, if Jones says, ‘I am standing’ and Smith says ‘Jones is standing’, we tend to think their utterances agree, even though they are not of the same syntactic type. Similarly, if Jones says ‘lo, a baby swan’ and Smith says, ‘lo, a cygnet’, we would say that their utterances agree. However, in making these judgements of agreement, we assume what Jones and Smith mean by their utterances—we attach a particular interpretation to them. The sceptic can hardly make use of this notion of agreement, because it takes an individual’s meaning something by an utterance as *prior* to whether two utterances agree. (If we assumed that Jones meant *elephant* by ‘swan’, then we would not think that his utterance of ‘lo, a baby swan’ agreed with Smith’s utterance of ‘lo, a cygnet’). That is, we assign a meaning to the utterances *first* before we can say that they agree. In so doing, we assume that the words in those utterances have conditions of correct use—we assume that Jones means *swan* by ‘swan’ and that Smith uses ‘cygnet’ to refer to baby swans. However, according to the sceptic, such assumptions about what Jones and Smith mean are illicit, for if Jones meant *swan* by ‘swan’, then the question whether his use of ‘swan’ is correct on this occasion would depend only on what he means and the way the world is in relevant respects; it would hardly depend on whether Smith agreed. If agreement is to determine which utterances are correct and which incorrect, utterances cannot be assigned to equivalence classes on the basis of what they mean.

The obvious alternative way to assign utterances to equivalence classes is on the basis of syntax. That is, utterances that *sound* the same, such as/pat/and/pat/but not/pat/and/cat/would be said to agree. Similarly, written words could be said to agree if they have the same shape. The advantage of this option is that it treats words and utterances as objects. At the same time, however, this is a disadvantage. For, if two utterances or two words can be said to be the same types of objects, we need to suppose that they conform to some objective pattern; that they have some properties of sound or shape in common. But this seems, on the face of it, not to be available to the sceptic. If there is no sense to be made of the idea that the word 'goose' tracks some objective (i.e. judgement independent) properties of individual birds, what sense can be made of the idea that our assignment of utterances to equivalence classes tracks objective properties utterances? Suppose, for instance, that the sceptic were to say that my disposition to treat/pat/and/pat/as the same sound determines that they are the same. If the sceptical criticism of dispositionalism is correct, however, the same criticism will apply here. For, I might mistake an utterance of/pat/for an utterance of/bat/so that it would be indeterminate whether every utterance of/pat/is an utterance of the same sound.

Nevertheless, it is tempting to think that the sceptic can appeal to speakers' dispositions in order to provide an account of agreement between speakers.[46] The sceptic need not assume that the individuals' dispositions themselves determine what the speaker means, but that agreement between two people's dispositions determines the equivalence class of their utterances. We need to be careful to steer clear of the community version of the dispositional view, according to which the dispositions of others yield a correctness condition. That is, the disposition should not be said to determine which uses of an utterance are correct, but only which utterances are the same. If Jones is disposed to say s in circumstances $C_1 \ldots C_n$ and Smith is disposed to say s_1 in the same circumstances, then we can say that s and s_1 count as the same utterance. Then, if Jones says s and Smith says s_1 in some novel circumstance, we can say that they agree in that circumstance.

The first problem with this suggestion is that it collapses into the communal version of dispositionalism, which Kripke rejects. On that view, what makes an utterance correct is that we agree in our dispositions to make it. On the suggestion we are considering, agreement in dispositions generates equivalence classes and it is agreement in utterances which determine correctness. The two views are indistinguishable in their verdicts on whether a use of an expression is correct.

---

[46] Cf. Zalabardo 1989.

The second problem with this suggestion is that it will become impossible for Jones and Smith to disagree. If there is a case in which Jones would say s, but Smith would not say s_1, then s and s_1 cannot be counted as the same utterance when we come to the novel case. But the fact that Jones and Smith say different things in this novel case cannot be taken to mean that they disagree. 'That is an apple' and 'apples are fruits' are different sentences, but they do not disagree. What we need for disagreement is the same sentence uttered in different circumstances, but that is something that would be unavailable on this account.

I have been assuming that genuine agreement is necessary for meaning, yet the sceptical picture of meaning could be interpreted differently. Instead of saying that Jones's response is correct when it *is the same as* Smith's response, the sceptic could say that Jones's response is correct when Smith *is inclined to assent to it*, or when he *takes it* to agree with what he would say. Wright, for instance, sometimes talks about a 'community of assent', which suggests that some behaviour will be correct if other members of the community assent to it, which only requires that they take it to agree with what they do, not that it actually does.

For instance, the sceptic might say that Jones's utterance is correct in C if Smith takes Jones to have said what Smith himself would have said; so long as it *seems* to Smith that Jones's utterance agrees with his own. It does not matter whether Jones's utterance is *really* in agreement with what Smith would have said, it only matters that it seems that way to Smith. Moreover, it does not matter in what respect their utterances agree, only that it seems to Smith that they agree in whatever respect Smith cares about. If this approach is adopted, the problems raised above might seem to disappear since it is not the fact that two responses agree that makes them correct, but merely a perceived agreement.

Unfortunately, this amendment makes no improvement over the original theory. In order for it to seem to Smith that Jones's uses agree with his own, he must *believe* that Jones's uses agree with his own. If the correctness of Jones's utterance depends on what Smith *believes* about Jones's utterance, the sceptic must presuppose what he is trying to explain. For Jones's utterance to be correct, Smith will have to have a belief that has content independently of this agreement with Jones. If

Smith's belief is to acquire its content from some prior agreement, say, with Brown, then Brown will have to have a belief which has content prior to any agreement with Smith. No matter how many agreements there are, at some point, somebody has to have a belief that has content independently of agreement with anyone else. Thus, to say that for Jones's use to be correct, it must seem to Smith that Jones's uses agree with his is to concede the possibility that Smith's thoughts have content independently of the agreement with Jones.

The sceptic might try to avoid this by saying that Smith need not *believe* that Jones's uses agree with his own. It is enough that Smith has a pro-attitude towards Jones's responses, that Smith 'assents' to what Jones assents to. This offers no escape either. First, as was pointed out previously, pro-attitudes have contents. A pro-attitude is always a pro-attitude *towards* some thing or some state of affairs. What makes it the case that Smith has a pro-attitude towards Jones's response and not to his hairstyle or his shoes? If Smith says 'yahoo!', he must *mean* 'yahoo!: Jones's uses of "plus" ' and not 'yahoo!: Jones's hairstyle', for his approval to be approval of Jones's uses of plus. Second, pro-attitudes imply beliefs. If I have a pro-attitude towards the boots you are wearing, I must believe that you are wearing boots. If Smith has a pro-attitude towards Jones's uses of the word 'plus', Smith must believe that Jones has used the word 'plus', and moreover, that he has used 'plus' in a certain way. Third, if Smith must express his approval to Jones (for, otherwise, how would Jones ever have the confidence to go on by himself), then Smith's utterance must have a content. If Smith says, 'I approve of your responses', or 'well done, Jones', these will only be expressions of approval if Smith means *I approve* by 'I approve' and 'well-done', as opposed to something else or nothing at all. Meaningless sounds express neither approval nor disapproval, neither assent nor dissent. The same considerations arise even if we assume that Smith expresses his disapproval or approval wordlessly: when Smith disapproves, he smacks Jones; when he approves, he gives Jones a kiss. But, if Smith gives Jones a kiss, what has he given Jones a kiss for? What is it that Smith approves of? Jones has no way of knowing and neither, for that matter, does the sceptic. If Smith's utterances have no meaning apart from communal agreement, then surely his actions can have no 'private' meaning either.

Now consider a suggestion due to Quine.[47] Although Quine was no social constructivist, he attempted to specify the conditions under which it would be reasonable to correlate a word of an otherwise unintelligible foreign language with the English word 'yes'. This would satisfy the social constructivist's requirement that no representations with determinate content are presupposed at the outset. Here is how the story goes.

---

[47] Quine 1960, p. 29.

Imagine that you go to Mars and without initially understanding a word of Martian, set out to determine which of the Martian's words expresses assent. On Quine's suggestion, you do the following. Every time the Martian seems to be sincerely asserting something, you repeat the very same sounds and see what he does. Now, suppose that you follow Quine's advice and the Martian frequently says 'Yok' in response to your repetition. On Quine's view, you would be encouraged to assume that 'Yok' is to be translated in English as 'Yes'.

Unlike Quine, however, the social constructivist needs to be able to say that under circumstances such as the above, it would be *legitimate* or *reasonable* for you to take 'Yok' to be equivalent to the English word 'Yes'. However, it is not clear that this is so. For, it is possible that you have systematically mispronounced the Martian's words, and what he means by 'Yok' is equivalent to the English 'What?'. Alternatively, it is possible that what the Martian keeps saying contains an indexical, as in 'I am the king of the Martians', 'here is my home', and 'here is my wife'. If you repeat these sounds, even if impeccably pronounced, and the Martian says 'Yok, yok, yok', this would most reasonably be correlated with the English 'No'. Indeed, the only conditions under which it would be reasonable to assume that the Martian's 'Yok' is an expression of assent is if you are already sure that your repetitions of the Martian's sentences are 'the same as' the Martian's original sentences in the relevant respects. And, as we have already seen, this is not to be obtained from objective similarity relations. What is needed is a rule that determines what *counts* as the same, and a rule is precisely what the social constructivist does not have.

Next, consider another popular suggestion: that assent and dissent or agreement and disagreement can be expressed by sanctioning behaviour or dispositions to sanction. Several people, such as John Haugeland[48] and Bob Brandom,[49] have made this suggestion. I will focus on Brandom here.

Like Kripke's Wittgenstein, Brandom holds that we can build up a picture of meaning and content out of collective agreement and disagreement without presupposing realist correctness conditions. Brandom adds that *inferences* establish the content of expressions. For example, an inference constitutive of the meaning of the word 'red' might be expressed as follows: someone who says '*x* is red' thereby undertakes a commitment to say or assent to '*x* is coloured', '*x* is extended', '*x* is not blue', and so on. These inferences are instituted by our practices of taking and undertaking the relevant commitments. Moreover, what it is to take or undertake a commitment is to be disposed to punish behaviour that is out of line and reward behaviour that is in line. The question is, do dispositions to sanction really amount to attributions of determinate commitments, as Brandom suggests?

---

[48] Haugeland 1998.

[49] Brandom 1994a.

The answer is 'no', if you accept the sceptical arguments against dispositionalism outlined previously. Imagine that Jones says 'that is red' and Smith then becomes disposed to sanction Jones under some conditions rather than others. And now suppose that Jones says 'that is blue' and Smith punishes her. The question is, what does this tell us about the commitment Smith has attributed to Jones? We want to say that the commitment she attributed was to *not* say 'that is blue' after saying 'that is red'. But Smith's behaviour is consistent with his having attributed the commitment to say 'that is schmue' where something is schmue if and only if it is either red or blue—and Jones did not say *this*, which was why she was punished. The problem is that a disposition to sanction is not sufficient to determine *which* of several candidate commitments Smith has attributed to Jones. For, if all of Kripke's dispositions to use the word 'plus' are consistent with his having meant some bizarre function, such as *quaddition* by 'plus', so too are Smith's dispositions to sanction consistent with her having attributed some bizarre commitment to Jones. However, if Smith's sanctioning behaviour does not determine which commitment he has attributed, then their practice does not institute one inference pattern as opposed to another, and so does not determine the correctness or incorrectness of uses of expressions.

Maybe if we forget about the complexities introduced by inferentialism, sanctions can just be used as a way of expressing assent or dissent. Reward and punishment seem to be a natural way to express assent or dissent—simply because rewards are pleasant and punishments are unpleasant. Moreover, we might think that when Smith punishes Jones, it is Jones's behaviour that causes Smith to punish her, not her shoes or taste in music, so the object of Smith's punishment is Jones's behaviour as opposed to her shoes or her taste in music. However, even the decision to treat Jones's behaviour as the object of Smith's punishment requires making an arbitrary choice. For, there is a causal history to

Jones's behaviour, all of which is the part of the total cause of Smith's punishment, and there are numerous intermediate causal events between Jones's behaviour and Smith's punishment. Why should we pick *one* stage in this causal history rather than another as *the* cause and therefore object of Smith's punishment? If Jones's behaviour was caused by her having a headache, for example, then we could equally say that Smith disapproved of Jones's headache. Once again, either it is indeterminate whether we are entitled to say that Jones and Smith agree, or we must assume realism.

Finally, consider Donald Davidson's version of the Private Language Argument.[50] Like Kripke, Davidson holds that no substance can be given to the idea of an individual using his or her words 'correctly', in the absence of any interpreter of the individual's words. Unlike Kripke, Davidson denies that we can only legitimately say that someone means something by a word if we see that he does as we do. Rather, for Davidson, 'the intention of the speaker to be interpreted in a certain way provides the "norm"; the speaker falls short of his intention if he fails to speak in such a way as to be understood as he intended.'[51] On Davidson's view, then, a speaker's utterance can only be said to be 'correct' if it is understood (or can be understood) in the way that the speaker intended it to be understood.

Now, there are some immediate problems with this suggestion, which will no doubt be familiar from the foregoing considerations. Suppose Jones says 'emeralds are green' and intends to be interpreted as saying something that is true if and only if emeralds are green. The first question is, what is it for a speaker to intend to be interpreted in a certain way? What makes it the case that Jones intends that Smith interprets her utterance of 'emeralds are green' to be true if and only if emeralds are green? If Jones intends that Smith understands her to use 'green' to refer to the green things, Jones must at the very least have a mental representation that refers to Smith (and only Smith), a mental representation of the word 'green' and a mental representation of green things. Otherwise, given the rule following considerations, there will be no fact of the matter what Jones intends. And even if we can assume that Jones has an intention with determinate content, we can then ask what it is for Smith to interpret Jones as Jones intended to be interpreted? It would seem that for Smith to interpret Jones as she intended to be interpreted is for Smith to *believe that* Jones said something that is true

[50] Davidson 2001.

[51] Ibid., p.116.

if and only if emeralds are green. For Smith to do this, he must have a mental representation that refers to Jones and only Jones, a mental representation of Jones's utterance, of emeralds and greenness.

These worries are not alleviated when we consider Davidson's argument for the essentially public character of language. The argument goes as follows. Davidson asks us to imagine a primitive learning situation—that of a child learning the English word 'cow'. The child babbles, and when he produces a sound like 'cow' in the conspicuous presence of a cow, he is rewarded. This process is repeated until the child repeatedly says 'cow' in the presence of cows. That is, the child generalizes from the first cow to further cows, on the basis of a perceived similarity between them. Although this story may seem straightforward, there is a problem: the causal chain ending in the child's utterance of 'cow' goes from prehistoric ancestors to cows right up to the stimulation of the child's optic nerve and events in his brain. Why choose the cow as *the* stimulus of the child's response? Davidson claims that there can be *no answer to this question*, in the absence of an interpreter. However, if the mother is interpreting the child, then there is an answer to the question. The child finds cows similar, his mother finds cows similar, and his mother finds the child's responses to cows similar. There is a triangle—one line goes from the cow to the child, another from the cow to his mother, and the third from the mother to the child. Under these conditions, Davidson claims, it is reasonable to call the child's responses responses to cows.[52]

There may be a triangle here, but there is also a circle. As I have already mentioned, the rule following considerations put pressure on the idea that objective similarity relations between any two things can ground what *counts* as 'the same'. That is, between each utterance of the word 'cow', there are both relations of similarity and relations of difference; between each cow, there are both relations of similarity and relations of difference. Davidson's suggestion seems to be that what makes it legitimate to say that the child's sounds are the same is that the mother responds to them in the same way. But, of course, just as there are both relations of similarity and difference between two utterances of 'cow', so too are there both relations of similarity and difference between any two of the mother's responses. We cannot simply take for granted that the mother's responses count as 'the same' without any further explanation. To make the problem more vivid, consider what, in my experience, is a more accurate picture of what happens. First, the child says something like ‘ga!’ in the presence of a picture of a cow. The mother responds by clapping her hands, and saying ‘yes, “cow!” well done!’. On the next occasion, the child says something a bit different, such as ‘goo!’, now in the presence of a plastic replica, of a cow. This time, the mother smiles, gives her child a hug, and says ‘clever boy!’. In this picture, for any two things that are supposed to come out as ‘the same’—such as the two cows, the child’s two utterances, and the mother’s two responses—there are very clearly differences as well. If what makes the child’s utterances count as utterances of the same sound, in response to the same kinds of stimuli, is that the mother’s responses are the same, we need independent grounds for saying that the mother’s responses count as the same.

---

[52] Davidson 2001.

Of course, Davidson, or Kripke, might like to introduce a third person. But it is clear that this will turn the circle into a vicious regress. For, if we need to appeal to the father’s responses to the mother’s responses in order to claim that the mother’s responses are ‘the same’, we need some independent grounds for claiming that the father’s responses to the mother’s responses are ‘the same’. In general, if the second person is to make a difference in determining which of the first person’s utterances count as ‘the same’, there must be some way of answering the question of what makes the second person’s utterances count as the same prior to and independently of the introduction of a third person. But this, of course, means that if there is to be any such thing as a public language, there must already be such a thing as a private one.

Is there any way for the sceptic to wriggle free of these difficulties? It would seem not, because the sceptical solution cannot do without appeal to contents altogether. The reason is that the sceptical solution aims to rehabilitate our talk of meaning and truth; it purports to show that our talk of meaning and truth is *legitimate*, even if non-factual. Smith is supposedly justified in saying that Jones means *addition* by ‘plus’ because he judges that Jones’s uses of ‘plus’ agree with his own. Presumably, if Smith were to judge that Jones’s uses of ‘plus’ diverged radically from his own, he would not be entitled to assert that Jones means *addition* by ‘plus’. The relevant difference between these two judgements is, of course, their contents; hence, it is in virtue of their contents that certain judgements can legitimate meaning ascriptions. Furthermore, Smith’s judgement that Jones’s uses agree with his own allegedly justifies his judgement *that Jones means* addition *by 'plus'*—not, for instance, the judgement *that Jones means* multiplication *by 'plus'*. At both ends of this justification relation are judgements, and it is in virtue of the contents of those judgements that they stand in the appropriate relation. The non-factualist seems to be caught in a bind. Appeal to judgements with contents is both unavoidable and yet conflicts directly with non-factualism about meaning.

CONCLUSION

The social constructivist faces a dilemma: the choice is between semantic realism on the one hand, and the self-refuting thesis that nobody ever means anything by any word on the other. If realism is rejected, then it is indeterminate whether any two people can be said to agree. The social constructivist cannot appeal to judgements, expressions of approval or assent, because to do so is to presuppose representations with determinate content constituted *prior* to the agreement. The social constructivist cannot appeal to inclinations or dispositions to sanction, because to do so leaves it completely indeterminate, in any given case, whether two people can be said to agree. And if it is indeterminate whether two people can be said to agree, it is indeterminate whether they can be said to use their words 'correctly', and indeterminate whether they can be said to mean anything by them. Of course, the sceptic can say that there really is no such thing as meaning, just brute inclinations to respond unhesitatingly, but blindly. However, this way lies self-refutation: if this claim is true, then it is itself meaningless, and therefore cannot be true. The only remaining option is to concede to the realist that the meaning of our representations is determined *prior* to collective agreement. And this, unsurprisingly, is what I urge the social constructivist to do.

Furthermore, semantic non-factualism is irremediably incoherent. Any attempt to rehabilitate meaning talk or truth talk seems to conflict with the no fact thesis. There are, no doubt, other forms of semantic non-factualism which I have not had the space to consider. However, the arguments given above seem to yield a general problem that will be faced by any form of semantic non-factualism whatsoever. The problem, as I see it, is that any attempt to rehabilitate our practices of semantic discourse in the face of non-factualism will need to show that the judgements or statements of our semantic discourse are legitimate in some sense—either true in a weakly deflationary sense or justified.

But it is only the contents of those judgements and the statements that are expressed by the sentences of our discourse that can properly be said to be true or justified in even the most rarefied, deflationary sense. Any attempt to show that we are justified in our talk about meanings or our use of the predicate 'true' requires that we ascribe content to the sentences or judgements before we can proclaim them to be true or justified. Uninterpreted sentences are nothing but noise, and noise cannot be true or false, justified or unjustified, legitimate or illegitimate.

If non-factualism about meaning is irremediably incoherent, the sceptical argument, which leads to that thesis, must go wrong somewhere. The aim of the following chapters is to discover where.

5

Reductionism

Kripke's sceptic argues that no reductive theory is able to accommodate the normativity of meaning. As I suggested in the previous chapter, this argument seems to presuppose that meaning is normative in the strong sense of being categorically prescriptive, since it is only if meaning is normative in this sense that reductive theories would face Moore's Open Question Argument. In this chapter, I will examine more carefully further reductionist responses that have been made to Kripke's sceptic. Of course, I cannot consider all of the proposals which have been made, but I discuss what I take to be the most prominent and plausible proposals in the literature. The vast majority of these have been more or less sophisticated versions of the dispositional theory. All of these, it turns out, fail to solve the sceptical problem.

Reductive theories, broadly speaking, aim to reduce representation to causal covariance—my 'horse' thoughts refer to horses because they are typically caused by horses. Of course, mere causal covariance is not enough for representation. Fire typically causes smoke, but we do not think that smoke refers to or represents fire. At least one further condition must be met: a causal account of representation must allow for *mis*representation. My 'horse' thoughts are about horses even when they are not caused by horses, and even when these thoughts are false. The capacity to misrepresent is essential to the capacity to represent. Hence, it is a condition of adequacy for theories of representation that they can account for misrepresentation. As it turns out, this condition of adequacy is not met by any of the extant naturalistic theories that follow.

DISPOSITIONALISM REALISM

The dispositionalist theory, as Kripke characterises it, is obviously very crude. Hence many have sought to improve on the dispositionalist theory in the hope that it can thereby meet the sceptic's objections. I have already argued that dispositionalism, even in its crudest form, can meet the objection from the infinitude of the addition function. The question now is whether dispositionalism can be so modified that my meaning *addition* by 'plus' can be entirely reduced to facts about what I am disposed to do.

The problem that dispositionalism still needs to solve is the problem of error. If I am both disposed to add, under some circumstances, and disposed to make mistakes, under other circumstances, then is there some principled way to legislate that my disposition to add determines what I mean by 'plus'? One of the most sophisticated attempts to modify the dispositional view has been advanced by C. B. Martin and John Heil.[1] Martin and Heil argue that Kripke makes controversial assumptions about the metaphysics of dispositions: Kripke assumes an empiricist account of dispositions, as opposed to a realist one. Moreover, they argue that if we embrace a realism about dispositions, Kripke's objections miss their target. Empiricism about dispositions, as they characterise it, consists of two main claims, one concerning the semantics of disposition ascriptions, and the other concerning the metaphysics of dispositions. The empiricist's semantic thesis is that ascriptions of dispositions can be analysed in terms of true conditionals, where the antecedent of the conditional specifies a 'trigger' or a 'stimulus' and the consequent specifies some behavioural response. For example, on a very simple empiricist account, to say that something is fragile is to say that if it were sufficiently and lightly stressed, it would break. Given the conditional analysis of disposition ascriptions, the empiricist advances the metaphysical thesis that dispositions are nothing more than these conditional facts relating stimulus to behaviour.

Martin and Heil argue that Kripke assumes an empiricist account of dispositionalism. According to Kripke, the dispositionalist cannot account for error because whatever a speaker does, she must have been disposed to do. If what the speaker is disposed to do is correct, then everything a speaker does will be correct. Thus, it seems as though Kripke implicitly assumes that a disposition ascription can be analysed in terms of what a speaker would do. The realist about dispositions rejects this assumption. According to the realist, we cannot tell from the fact that a speaker fails to add, on some occasion, that the speaker lacks the disposition to add. On some occasions, although the speaker is disposed to behave in a certain way, she may fail to behave in that

[1] Martin and Heil 1998. See also Martin 1994 and Mellor 2000.

way because something has interfered with the manifestation of her disposition. Likewise, for instance, a water soluble salt ordinarily fails to dissolve in water when surrounded by a strong electromagnetic field. But we do not say that it is not the case that the salt is water soluble or that it is indeterminate whether the salt is or is not water soluble.

Realists about dispositionality can allegedly avoid the objections raised by the sceptic. Take the problem of infinitude. Kripke's sceptic argued that a speaker's dispositions could not determine the infinite addition function because there will always be some numbers too large for the speaker to grasp. The dispositionalist would have to appeal to a radical idealisation of the speaker's capacities to make it come out true that she *would* respond with the sum of any two numbers when asked. By rejecting the conditional analysis of dispositionality, the realist apparently avoids this difficulty. That I have the disposition to add does not imply that I would respond with the sum of any two numbers when asked. Given large numbers, I cannot manifest my disposition to add, but I am disposed to add nonetheless. The constraint on my disposition to add comes from my capacity to grasp numbers, which is limited. Martin and Heil call my capacity to grasp numbers a 'disposition partner' of my disposition to add. My disposition to add determines the correct answers to sums involving all numbers, no matter how large, but it will only be manifest in conjunction with a capacity to manipulate the given numbers. What Kripke fails to notice is that:

> [some speaker] *S*'s finitude with respect to addition results from limits on the numbers *S* can consider or manipulate at a given time. This is a limitation, not on [her disposition] *P*, but on *P*'s *manifestations* owing to limitations in [her capacity] *C*, one of *P*'s reciprocal disposition partners. This is not a limit on magnitudes of numbers *S* (in virtue of possession of *P*) is prepared to add; the dispositional readiness encompassed by *P* is for *any* magnitude, and is in that sense infinite.[2]

The realist about dispositions maintains that a speaker can have the disposition to add even if she does not manifest that disposition under some, putatively appropriate, conditions. The disposition to add is itself infinite, although her capacity to manipulate numbers might be limited. Martin and Heil suggest a similar solution to the problem of error. The speaker who means addition by 'plus' is one who is disposed to add, even though this disposition may not be manifest under certain circumstances. Since the realist does not reduce dispositions to conditionals, the fact that a speaker does not add under certain putatively appropriate conditions, does not mean that she does not have the disposition to add, and so means *addition* by 'plus'.

---

[2] Martin and Heil 1998, p. 302.

The problem with Martin and Heil's putative solution to the sceptical problem is that it fails to provide a way of non-circularly distinguishing the meaning-constituting dispositions from the error-producing ones. Capacities that interfere with the manifestation of my disposition to add are themselves dispositions. Not only do I have the disposition to respond with the sum of any two numbers when asked, I am also disposed *not* to do so when I am tired, or the numbers are very large. Martin and Heil want to be able to say that it is my disposition to add that determines what I mean by addition. But a wily sceptic can easily claim that it is my disposition to give deviant responses that determines what I mean. If we want the dispositional theory to account for my meaning *addition* by 'plus', there must be some principled way to uniquely identify the disposition to respond with sums of numbers as meaning-constituting, so that other 'interfering' dispositions can legitimately be ruled out as error-producing. Realism about dispositions does not in itself help to solve the problem since if my disposition to add is real, then so is my disposition to be confused by large numbers. As an imperfect adder, I have two dispositions:

D_1: I am disposed to respond with o, such that $m + n = o$.

D_2: I am disposed to respond with o, such that it is not the case that $m + n = o$.

Here, D_2 could be assumed to specify the admittedly bizarre function whose correct applications are coextensive with all of my putative errors. For instance, if I systematically forget to carry, my disposition to forget to carry will determine a function that will yield the result that $68 + 57 = 115$. The function coextensive with the output of D_2 will no doubt be bizarre, but it is a perfectly well defined function nonetheless. Call that function *Schmaddition*. The sceptic can now ask, what makes it the case that I mean *addition* rather than *shmaddition* by 'plus'.

We would like, of course, to be able to say that D_1 constitutes what I mean by 'plus', whereas D_2 interferes with my disposition to add, and so constitutes my errors. The sceptic, however, suggests that the opposite is true: D_2 constitutes what I mean by 'plus', whereas D_1 constitutes my errors. That is:

S means *addition*		S means *schmaddition*	
D_1	Meaning-constituting	D_1	Error-producing
D_2	Error producing	D_2	Meaning-constituting

Martin and Heil's response to this difficulty is somewhat vague. They claim that a complex dispositional state answers to someone's meaning *plus*, and that this complex state supplies the difference. They ask us to imagine Lilian, who is counting by twos but gets tired and therefore makes a mistake.

> The basis of Lilian's error—what makes it the case that she has erred and not merely changed her mind—lies in fine-grained details of her dispositional make-up. In the simplest terms, the manifesting of one disposition, that underlying her mastery of the two-rule, has been blocked; some other disposition is manifested in its stead. This constitutes a mistake in Lilian's case because she has deviated from the route she set out to follow and still takes herself to be following.[3]

This is unsatisfactory because it appeals to what Lilian *intended* to do—what rule she set out to follow and still takes herself to be following. The sceptic can always ask what makes it the case that Lilian intended to follow the rule for counting by twos? Perhaps, when she told herself what she was to do, she meant *quount* by 'count', where to quount by twos is to count by twos until a certain point and then count by fours. Martin and Heil could respond that what Lilian means by 'count' is also determined by her dispositions. She has a disposition to count and that disposition can be interfered with. But then, of course, the same difficulty of ruling out error-producing dispositions will emerge. Although Lilian may have the disposition to count, she also clearly has the disposition to make 'mistakes'. To answer the sceptic, we need a way to specify that the disposition to count is meaning constituting that does not appeal to Lilian's intentions or what she takes herself to be doing.

DISPOSITIONS AND CONDITIONALS

Hugh Mellor advanced another of the more compelling dispositionalist responses to Kripke's sceptic, and Mellor's response is radically different from Martin and Heil's. Whereas Martin and Heil attempt to solve the sceptical problem by rejecting the conditional analysis of dispositional predicates, Mellor suggests that it can be solved if we adopt the correct conditional analysis of dispositional predicates.

---

[3] Martin and Heil 1998, p. 306.

Nevertheless, Mellor's response to the problem of infinitude is of a piece with Martin and Heil's. On a typical conditional analysis of, say, '*x* is fragile', it means 'if *x* were dropped, it would break'. Mellor claims, however, that to argue that I do not mean addition because there are some numbers too large for me to grasp seems to be on a par with saying that 'a glass *a* that cannot be dropped cannot be fragile, an absurdity that no conditional account of fragility need entail.'[4] He concedes that if it is metaphysically impossible to drop a glass, then 'if *a* were dropped, it would break' and 'if *a* were dropped, it would not break' would both be vacuously true, since their antecedents are impossible, and '*a* is fragile' vacuous. However, he says, 'even if there must be some upper limit to the numbers we can add, it can still be contingent that, for any given *n* and *m*, we cannot add them.'[5] Mellor's conclusion, then, is that the contingent fact that I am incapable of grasping numbers greater than *n* does not entail that I do not mean *addition* by 'plus', just as the contingent fact that some glass *a* cannot be dropped does not entail that *a* cannot be fragile. This is because it remains true that if glass *a* were dropped, it would break, and it allegedly remains true that if I *could* grasp numbers greater than *n*, I would respond with the sum if asked to add them.

Does this meet Kripke's objection? I think not. First, because although it is true that if I could grasp extremely large numbers, I would respond with the sum if asked to add them, it is not clear what this says about my current dispositions. One of Kripke's objections to this line of reasoning is to point out that, since the addition function is infinite, for it to be true that I *would* give correct answers to all sums, I must be capable of grasping infinitely large numbers. Even if it is true that if my mental capacities and life span were considerably enhanced, I would answer with the sum of any two numbers, if asked, what would this tell us about my current dispositions? We have to make an assumption about what I would do if much more intelligent than I currently am, which, in effect, is to make an assumption about what I would do if my dispositions were enhanced, not what I would do given the dispositions that I have now.

In response to this, Mellor might argue that Kripke seems to presuppose that you need to know *exactly* how something would behave under an idealisation in order to say what it is disposed to do.[6] Of course, this is far too strong a requirement. As I have argued in Kripke's defence, however, it seems that in the case of the disposition to add any two numbers if asked, the necessary idealisation is just far too radical. I am a finite being, with finite capacities, whereas the addition function is defined for the infinite series of positive integers. Thus, the only suitable idealisation would have to be one that credited me with an infinite mental capacity and immortality. This idealisation seems less on a par with the ideal gas laws than with saying that the tortoise is faster than the hare because the tortoise would beat the hare if only the tortoise could move its stubby, little legs much, much faster.[7]

---

[4] Mellor 2000, p. 764.

[5] Ibid.

Second, this response seems to miss its target because Kripke does not simply claim that I cannot mean *addition* by 'plus' because there are some numbers that are too large for me to grasp. He adds that, were I to try to add very large numbers, I would not succeed; I would give answers that do not accord with the addition function. This is more on a par with saying that in certain situations, if glass *a* were dropped, it would not break. The real problem is that there are some numbers that are so large that I would make mistakes if I tried to add them, or would get frustrated when trying to do so and give bizarre, *quus* like responses. Hence, it is not the case that I would give an answer which accords with the addition function if asked to add such large numbers. Given that this is so, the sceptic can always invent a bizarre *quus*-like function with which my errors do accord, and then argue that it is this bizarre function that I meant. Indeed, I do not even need particularly large numbers to induce error—sometimes I give answers that do not accord with the addition function because I am tired or inattentive or confused or feeling perverse or have written the numbers down illegibly, or for any of a great number of other reasons.

Mellor's response to this problem is that it can be assimilated to the problem of 'finkish' dispositions.[8] A finkish disposition is one that disappears when its stimulus conditions obtain. For instance, if dropping a fragile glass would cause it to cease to be fragile, then it is 'finkishly fragile' so that '*a* is fragile' would be true, but 'if *a* were dropped, it would break' would be false. Mellor argues that we can amend the antecedent to the conditional that specifies the meaning of '*x* is fragile' to read 'if *x* were dropped without ceasing to be fragile, *x* would break'.[9]

---

[6] Blackburn 1984.
[7] Boghossian 1989.
[8] Martin 1994.
[9] Mellor 2000, p. 763.

This is what Mellor calls a 'reduction sentence account' of the semantics of dispositional predicates. In this form, it looks circular. However, he says that we could remedy the ignorance of someone who does not know what 'fragile' applies to by saying that 'by definition, all and only those things that remain or become fragile when (relatively suddenly and lightly) stressed will then break.'[10] With this amendment to the conditional account of 'fragile', a finkishly fragile glass that was dropped and ceased to be fragile would not satisfy the antecedent of the conditional that specifies the meaning of '*x* is fragile' so that the fact that it does not break would not entail that it is not fragile. Mellor claims that a similar amendment can be used in response to Kripke's argument against the dispositionalist. Mellor holds that what Kripke shows is that:

> Some numbers make our disposition to follow the addition rule finkish: that is, trying to add those numbers would cause us to lose this disposition and hence to add them wrongly or not at all. But that, as we saw in the case of fragility, is no problem for the reduction sentence account of dispositional predicates. For on that account, if "x means plus by '+' " ascribes a disposition to *x*, this means that, for any two numbers *n* and *m*, if *x* were to apply "+" to them while having this disposition, *x* would get the answer $n + m$.'[11]

On Mellor's view, then, if some numbers are so large that they cause me to lose the disposition to add them, then my disposition to add is finkish. And on Mellor's reduction sentence account, finkishness can be dealt with. When faced with numbers that are so large that they cause me to lose the disposition to add them, I do not satisfy the antecedent of the conditional that specifies the meaning of '*a* means *addition*'. Hence, the fact that I fail to respond with the sum of these numbers when I lack the disposition to add them does not entail that I do not mean *addition* by 'plus'.

This answer to the sceptic is unsatisfactory. Suppose that being faced with numbers greater than *n* causes me to lose my disposition to add. Since the dispositionalist answer to the sceptic is that to say that I mean *addition* by 'plus', is to ascribe to me a disposition to add any two numbers when asked, then when I lose that disposition, it would seem that I can no longer mean *addition* by 'plus'. That is, although the fact that I fail to add numbers greater than *n* when asked may not entail that I do not mean *addition* by 'plus', the fact that I lack the disposition to add those numbers surely entails that, when faced with those numbers, I do not mean *addition* by 'plus'. Suppose that $n = 57$, so that when faced with the sum '57 + 68', I would lose the disposition to add and I would say '5'. If to say that I mean *addition* by 'plus' is to ascribe to me the disposition to add any two numbers when asked, and if I lose the disposition that determines what I mean when I am faced with a sum involving numbers greater than 57, then I do not mean *addition* by 'plus'. When faced with numbers greater than 57, I manifest another disposition, the disposition whose output accords with some bizarre function, we might as well call *quaddition*. The sceptic could now go on to argue that since the quaddition function can be defined in such a way that it converges with the addition function for values less than 57, but only diverges for values greater than 57, what I really mean by 'plus' is *quaddition* and to say 'x means *quaddition* by "plus" ' is to attribute a disposition to quadd which means that 'for any two numbers n and m, if x were to apply "+" to them while having this disposition, x would get the answer o, where n *quus* $m = o$.' The problem is that we want to be able to say that when I am faced with extremely large numbers, even though I might make mistakes, I do not cease to mean *addition* by 'plus'. Yet, all the facts about how I would behave are impotent to rule out the sceptical suggestion that I really mean *quaddition* rather than *addition* by 'plus'.

---

[10] Mellor 2000, p. 763.

[11] Ibid., pp. 764–5.

Thus, what the dispositionalist needs to say is not that I lose the disposition to add when faced with large numbers, but that I can make mistakes even though I retain the disposition to add. In a similar vein, we would say that a water soluble salt that fails to dissolve in a sufficient quantity of water because it is surrounded by a strong electromagnetic field is still water soluble—it still has the disposition to dissolve in water, even though, in this circumstance, it fails to manifest that disposition. In the case of adding, it also seems as though I can fail to give answers that accord with the addition function under certain circumstances—when tired, confused, not paying attention, etc.—without losing the disposition to add. Now, to accommodate this intuition, Mellor might amend the conditional which specifies the meaning of 'x is disposed to add' to include the relevant circumstances—that is, circumstances in which x is not tired, confused, etc. Indeed, this is the line of response Mellor suggests to Bird's problem of 'antidotes'—that is, circumstances that prevent the manifestation of a disposition.[12] Thus, Mellor might say that 'x is disposed to add' means 'for any two numbers n and m, if x were to apply "+" to them in circumstances of a kind C, x would get the answer $n + m$.'

---

[12] Bird 1998.

Once again, this turns out to be an unsatisfactory response to the sceptic. For, the sceptic can argue that if x is in circumstances that are not C, but C', x would get the answer *n quus m*, where the *quaddition* function is defined in such a way that the answers x would give in either C or C' would be correct. Thus, the sceptic could argue that x really means *quaddition* by 'plus', which is to ascribe a disposition that means 'for any two numbers n and m, if x were to apply "+" to them in C or C', x would get the answer *n quus m*.' This is not to say that I mean *addition* by 'plus' under some conditions, and *quaddition* by 'plus' under others, but that I only ever mean *quaddition* by 'plus'.

These difficulties for Mellor's view are related. The fundamental problem is that regularities of behaviour are caused by dispositions. And a speaker who is disposed to add may nevertheless also be disposed to give answers that do not accord with the addition function. Remember the person who means *addition* by 'plus', but who systematically fails to carry when doing complex sums. What we have here is (at least) two dispositions, manifested on different occasions, one to produce answers that accord with the addition function and one to produce answers that do not accord with the addition function. What we want to say is that such a speaker means *addition* by 'plus', but makes mistakes. But to say this we need to be able to non-circularly specify that one of the speaker's dispositions is *meaning-constituting* and the other *error-producing*.[13] That is, the following conditionals are true of the imperfect adder:

(1) If asked in circumstances C, S would respond with o, such that $m + n = o$.

(2) If asked in circumstances C', S would respond with o, such that it is not the case that $m + n = o$.

For any English speaker for which these conditionals are true (which includes, I would wager, all of us), we *want* to say that (1) specifies a disposition that is meaning constituting, whereas (2) specifies a disposition that is not meaning-constituting. But in response to this inclination, the sceptic has two options. He can either say that the reverse is true—that is, that the disposition specified by (1) is meaning constituting whereas that specified by (2) is error producing—or he can say that a single disposition, specified by both (1) and (2), is the meaning-constituting disposition and the speaker simply never errs. The facts about what a speaker would do are thus consistent with both the hypothesis that the speaker means *addition* by ‘plus’ and sometimes errs, and with the hypothesis that the speaker means something non-standard by ‘plus’.

---

[13] Boghossian 1989.

I have been assuming that Mellor’s proposal is in the spirit of most dispositional responses to Kripke’s sceptic—that is, that it is intended to be reductive. However, this may not be so. Indeed, given Mellor’s views on the poverty of physicalism,[14] it might make sense to suppose that he is not attempting a reduction of semantic content to dispositions, but that his aim is to elucidate what we mean when we say such things as ‘Anandi means *addition* by “plus” ’. On this interpretation, Mellor’s aim is simply to give the application conditions of meaning ascriptions. That is, when we say ‘x means plus by “$+$” ’, what we mean is that x is disposed to add, which in turn means that ‘for any two numbers n and m, if x were to apply “$+$” to them while having this disposition, x would get the answer $n + m$’.[15]

If Mellor could specify the application conditions of ‘x means plus’ in such a way that ‘x means plus’ applies to us, but ‘x means quus’ does not, then a refutation of the semantic sceptic would clearly be in the offing. Unfortunately, however, Mellor’s specification of the application conditions of ‘x means plus’ falls short of this target. On Mellor’s view, to say that ‘x means plus by “$+$” ’ means that x is disposed to add any two numbers when asked, but, by the same token, to say that ‘x means quus by “$+$” ’ means that x is disposed to quadd any two numbers when asked. The trouble is that so long as x is fallible, the sceptic can always define ‘quus’ and ‘quadd’ in such a way that whenever ‘x means plus by “$+$” ’ applies, so too does ‘x means quus by “$+$” ’. All of the true conditionals about how a speaker would behave when faced with sums are impotent to discriminate between whether the disposition to add applies to the speaker or the disposition to quadd.

Thus, it seems that Kripke’s sceptical problem remains pressing. Assimilating the sceptical problem to finkishness does not help because if the speaker loses her disposition to add when faced with some numbers then, when faced with those numbers, she no longer means *addition* by ‘plus’. Thus, she will be disposed to add when faced with small enough sums, but disposed to quadd when faced with larger numbers or more complex sums. Indeed, the sceptic can even construct a ‘quaddition’ predicate that applies to her even in circumstances when she gives answers that accord with the addition function as well. Introducing circumstances of kind C is of no help because we still need some way to justify the claim that what the speaker says in circumstances of kind *C* is *correct*, whereas what she says in circumstances of kind *not-C* may be incorrect. We need some non-circular way of specifying which of the speaker's dispositions are *meaning-constituting* and which are *error-producing*. Even if we treat Mellor's proposal as a non-reductive response to the sceptic, on his specification of the application conditions of '*x* means plus', the predicate '*x* means quus' can be constructed so as to apply to any fallible speaker to whom '*x* means plus' applies. Hence, Mellor's reduction sentence account of the semantics of dispositional predicates does not, as he suggests, supply a solution to Kripke's sceptic.

---

[14] Crane and Mellor 1990.

[15] Mellor 2000, pp. 764–5.

EXTENDED DISPOSITIONALISM

Simon Blackburn's response to Kripke's sceptic is also a version of the dispositional theory. Instead of arguing that dispositions must be real, however, he argues that Kripke's dispositionalist restricts his vision to too small a range of dispositions. Like most dispositionalists, Blackburn maintains that the problem of infinitude is not, in itself, pressing. As I have argued previously, Blackburn's solution to the problem of infinitude is compelling. Dispositions are infinite—the fragile glass is disposed to break from an infinite number of possible strikings in an infinite number of possible places at an infinite number of possible times. A fragile glass is one that is disposed to break anywhere—even in places, such as Alpha Centauri, which the glass would never reach. Since dispositions are 'infinite', my dispositions to add small numbers can be used to determine a procedure which, according to Blackburn, specifies the answer to sums that I would accept. The procedure itself could be the one that I have characterised dispositionally in Chapter 2— I am disposed to add single digits in a certain way; when faced with multiple digit figures, I align them, add single digits and 'carry'. I may not be disposed to carry out the procedure enough times to yield an answer to every possible addition sum, but since my dispositions are 'infinite', the fact that I cannot add very large numbers is like the fact that the glass cannot get to Alpha Centauri. Just because I am unable to add large numbers does not imply that my dispositions to add small numbers fail to determine an infinite function covering those large numbers as well. There is no reason why the infinitude of the addition function poses a problem for dispositionalism.

Blackburn's solution to the infinitude problem jibes with those presented by other dispositionalists. Unfortunately, his solution to the problem of error also jibes with those presented by other dispositionalists, and so faces precisely the same difficulties. I may be disposed to add numbers, and this disposition may determine a function that covers all cases, but I also have dispositions to make mistakes. The question is: which of my dispositions determines what I mean? Blackburn's suggestion is that if we look at an extended set of dispositions—possibly all of my dispositions taken together—we will be able to distinguish those dispositions that constitute what I mean from those that do not. In addition to the disposition to add, we need to also consider the disposition to retain an answer after further investigation. Blackburn's intuitive thought is that although I may systematically fail to carry, and give mistaken answers to large addition sums, I will retract my mistakes if, for instance, I employ an independent procedure to check my calculations and find that they were mistaken. Thus, we might say that someone who means *addition* by 'plus' is someone who is disposed to carry out a procedure which yields answers that accord with the addition function and is prepared to retain only those answers that, upon further investigation, do accord with the addition function. Thus, Blackburn says,

> putting the errant disposition into a context of general dispositions of this sort supplies the criterion for which function is meant. The equation would be: By '+' I mean that function ϕ that accords with my extended dispositions. An answer $z = \phi(x, y)$ accords with my extended dispositions if and only if (i) it is the answer I am disposed to give and retain after investigation, or (ii) it is the answer I would accept if I repeated a number of times procedures that I am disposed to use, this being independent of whether I am disposed to repeat those procedures that number of times.[16]

The condition (i) is designed to address the problem of error. The answer that I am not disposed to retain after investigation is a mistake. Does this solve the sceptical problem? Unfortunately not—because under some conditions, I am disposed to wrongly revise or retain an answer that does in fact accord with the addition function. Suppose that in response to a lengthy addition sum, I give an answer that accords with the addition function. Because the sum is lengthy, I investigate further, and in carrying out this investigation, I make a mistake and arrive at an answer that does not accord with the addition function. Now, on Blackburn's suggestion, it would remain indeterminate whether I mean *addition* or some other bizarre function by 'plus'. Of course, we might add further dispositions. Perhaps I would have revised the second judgement upon further investigation. Or perhaps not. There is no guarantee that if I investigate as much as I can, I will come up with the answer that accords with the addition function. There is thus no guarantee that at the end of a long string of ordered dispositions, the highest order disposition to accept an answer will yield the answer that accords with the function that I meant.

---

[16] Blackburn 1984, p. 290.

In response to this worry, Blackburn suggests that there are still more dispositions to be taken into account—in particular, dispositions to act in ways that are indirectly related to the rule. Imagine, for example, a bricklayer, who is disposed to count by twos until 1,000, and then by fours, and who takes this to be the rule for counting by twos. If the bricklayer is told to move 2,000 bricks, two at a time, she might protest because four bricks are too heavy to carry at once. The point of this example is to show that dispositions to use expressions (as well as dispositions to revise prior uses) are integrated within practical *projects*. These projects and pursuits provide us with a measure of success or failure. I would revise my judgement that '$68 + 57 = 5$' if I tried to use that calculation while trying to build a barn, or when drawing up a guest list for a party. If I add a list of 68 of my closest friends to another list of 57 relations, and then set the table for five, I will no doubt have trouble when 125 people show up. When concepts are used in the execution of my projects, I will not be disposed to revise my correct judgements and retain my mistakes because success or failure in the project itself provides an independent measure of correctness.

Paul Coates has usefully expanded on Blackburn's suggestion. He argues that we need to consider the extended dispositions in light of not just belief states but of desire states as well. The content of a concept that I grasp, such as *addition* or *horse* is determined by those dispositions that lead to success in satisfying desires with the same content. The intuition is this: if I mean *horse* by 'horse', then my application of 'horse' to a cow is mistaken because I would revise it if, for example, I were to try to saddle and ride the cow, or if I were to try to comb its mane. In general, Coates says:

> The first … cases, which are relevant for fixing the content of [a belief] B, involve cases where B acts in conjunction with a congruent desire D, so that the conjunction of the two inner states leads to an action making use of a particular object X. … The content of B and D (which together make up the complex inner state C) will be fixed by that class F of items which could be substituted for X without changing the Complete Cycle of behaviour which would occur

> in the context: the content of C is determined in part by how the subject would discriminate objects in that context. So, for example, if my reason for reaching out and eating a given apple X is because I believe it is an apple and desire to eat an apple, the behaviour controlled by C would be unchanged if a different apple were substituted for X.[17]

A Complete Cycle of behaviour involves a belief and a desire with appropriately related contents, as well as an action aimed at satisfying the desire on the basis of the belief. Given an adequate specification of a Complete Cycle of behaviour, Coates can give a counterfactual account of error. I have made a mistaken judgement if I would revise that judgement after carrying out a Complete Cycle of behaviour. Thus, for instance, I would revise the judgement that there is a horse yonder if, were I to try to act on my belief that there is a horse yonder and my desire to ride the horse, my action would fail.

Blackburn's and Coates' 'pragmatist' account of content is certainly promising. However, as a reduction of meaning to dispositions, it is still unsatisfactory. First, a mistaken belief can lead to successful action under fortuitous circumstances, thus obviating the need to revise the original judgement. For example, suppose I believe that there is an apple in the bowl by my elbow, but it is really a pear. If I were to act on the belief that there is an apple in the bowl, and the desire to eat the apple, I should discover the apple to be a pear. However, suppose that as I reach out for the pear, someone else very quietly, and unbeknownst to me, removes the pear and places an apple in my hand. In that case, my belief that there was an apple in the bowl was mistaken although my desire to eat an apple was satisfied. Under those circumstances, even if I were to carry out this cycle of behaviour, I would fail to notice my mistake and revise my belief. The sceptic could then argue that what I really mean by 'apple' is *apple or pear*.

Second, the failure to satisfy a desire may not lead me to revise the right belief, the one that contributes to the failure. Suppose that I mistakenly believe that the nearby pear is an apple. Suppose further that I strongly believe that all the nearby fruits are apples. Then, even if I were to carry out a Complete Cycle of behaviour, I would not be disposed to revise my judgement that the pear is an apple. I may instead revise my beliefs about how apples taste, and end up with the additional false belief that I am eating an odd tasting apple. For these reasons, it is not clear how an account such as this one will ultimately be able to discriminate those applications of a concept that accord with its content from those which do not.

---

[17] Coates 1997, p. 177.

SUCCESS SEMANTICS

Success semantics is another example of a pragmatist account of meaning and content, as it too attempts to make mileage out of the truism that true beliefs are more likely to lead to successful action than false beliefs. Originally attributed to Ramsey, success semantics has gained a number of contemporary defenders, including Blackburn, Dokic and Engel, and Whyte.[18] On Whyte's view, success semantic is an augmentation of functionalism. According to Functionalism, every psychological state has a distinctive causal role, and beliefs and desires are no exception. For any given belief or desire, there are certain conditions that typically cause the belief/desire to be tokened, and there are certain actions that the belief/desire typically cause. Of course, the causal roles of psychological states are complex, because beliefs and desires only cause actions in conjunction with other beliefs and desires, and the action a particular belief causes will depend on the other beliefs and desires of the agent. For instance, suppose I believe that there are cookies in the cupboard. This belief will cause me to go to the cupboard if I want to eat the cookies, but it will not cause me to do so if I want to go on a diet. The same belief may cause me to buy a cookie jar, if I also believe that there are mice about. Thus, instead of saying simply that the belief that there are cookies in the cupboard typically causes me to go and eat them, we will have to say what actions the belief will typically cause in conjunction with other combinations of beliefs and desires. That is, the functionalist will specify the causal role of belief, B_1, roughly as follows: if combined with beliefs $B_2, B_3, B_4 \ldots B_n$, and desires $D_1, D_2, D_3 \ldots D_n, B_1$ would cause action A_1. If combined with beliefs $B_2^*, B_3^*, B_4^* \ldots B_n^*$, and $D_1^*, D_2^*, D_3^* \ldots D_n^*, B_1$ would cause action A_2. And so forth, for every complex of beliefs and desires.[19]

This makes it look as though Functionalism is circular: a description of the causal role of one belief must make reference to other beliefs and desires. However, this is not an insurmountable difficulty. One way around it is to specify the causal role of beliefs with the help of Ramsey sentences. To construct the Ramsey sentence for the causal role of a belief involves replacing all reference to other beliefs and desires by a description of *their* causal roles. If those causal roles make reference to further beliefs and desires, then those references are replaced by causal role descriptions. This continues until there are no references to beliefs or desires in the ascription of a causal role.[20]

---

[18] Blackburn 2005; Dokic and Engel 2002; Whyte 1990.

[19] Whyte 1990.

Functionalism is an important part of Success Semantics, but it does not tell us enough about the contents of beliefs and desires. An obvious way to augment the Functionalist story is, of course, to identify the contents of beliefs with the conditions that cause them to be tokened. However, this move will turn Functionalism into another version of Dispositionalism, and it will therefore suffer the same difficulties. Because my belief that there are horses about can be tokened not only by horses, but by pictures of horses, or worse still, horsey looking cows, I cannot identify the content of my belief with the conditions that cause its tokenings. As I have argued above, all of the more promising Dispositionalist theories failed to supply a principled distinction between the 'correct' causes of a speaker's tokenings, and 'incorrect' ones. Thus, even if the Functionalist specifies the typical causes of belief, this still will not amount to an adequate account of their contents.

The distinctive contribution of Success Semantics is to exploit the pragmatist insight that true beliefs are useful. Suppose that I want cookies, believe that there are cookies in the cupboard, and act successfully on that belief/desire pair. The pragmatist thought is simply this: if my belief is true, when I act on it, I will typically get what I want; when my belief is false, when I act on it, I will typically fail to get what I want. It is this insight that Whyte thinks can help supply a theory of meaning.

The question is, how, exactly, should this insight be formulated? According to Whyte, the truth condition of a belief is that condition which *guarantees* the success of any action caused by the belief.[21] This yields the central principle of Success Semantics, which Whyte calls (R), in deference to Frank Ramsey, to whom he attributes the insight:

> (R) A belief's truth condition is that which guarantees the fulfilment of any desire by the action which that belief and desire would combine to cause.

Whyte's thought is this: given all of the possible worlds at which some belief causes a successful action, the truth condition of the belief is that which guarantees the success of all of the actions it would cause. The truth

[20] Ibid. [21] Ibid.

condition of a particular belief is thus 'factored out', from the truth conditions of the conjunction of the beliefs that cause an action at any one possible world. The fact that there are cookies in the cupboard will guarantee the success of every action caused by the belief that there are cookies in the cupboard; this fact will play no role in guaranteeing the success of actions caused by the belief that there are cookies in the bedroom. Moreover, the fact that there are cookies in the bedroom will guarantee the success of actions caused by the belief that there are cookies in the bedroom, but it will not guarantee the success of actions caused by the belief that there are cookies in the cupboard. Thus, Whyte argues, the truth conditions of a given belief must be common to all those conjunctions of conditions that guarantee the success of actions caused by the belief.

To say that the truth condition is that which 'guarantees' success, according to Whyte, is to say that the truth condition is that condition which is necessary and sufficient for success. The problem with this suggestion, however, is that the truth is neither necessary nor sufficient for success.

In a critical discussion of Success Semantics, Robert Brandom pointed out that the truth is not sufficient for success because sometimes we can fail due to ignorance, rather than false belief.[22] Suppose that I truly believe that there are cookies in the cupboard, that the cupboard is in the pantry, and that the pantry can be reached through that door over there. Although all these beliefs are true, I could still fail to get the cookies if the cookie jar is stuck fast, or if the cupboard door has so swelled from humidity that it cannot be opened. If I simply have no beliefs about the state of the cookie jar, or the cupboard door, my ignorance will have undermined my success. But this seems like a counterinstance to (R).

In response, Whyte argues that ignorance really is just a kind of error. In order to act, he says, an agent must believe that there are no physical impediments to his action, and this belief will entail that the cupboard door is not swollen shut. If the agent believed that there were physical impediments to her action she would not act. Since an agent would not act at all if she believed that there were physical impediments to her action, the belief that there are no physical impediments must be necessary for action. If the cupboard door is swollen shut, then, the 'no impediments' belief will be to blame.[23] Thus, our failures are never due to ignorance, but always due to the falsity of one of our beliefs.

---

[22] Brandom 1994*b*.

[23] Whyte 1997, p. 85.

This suggestion does not help. Even if we grant that every agent has a 'no impediments' belief, it is not clear that the belief will entail what Whyte wants it to entail. Because the 'no impediments' belief is supposed to be necessary for action, it must be one of the causes of all actions. As such, its truth conditions will be those states of affairs that guarantee the success of all actions—the molecular cohesion of the agent, the fact that the agent remains conscious throughout the performance of the action, and so forth. According to Whyte, the truth condition of a belief is what is common to the truth conditions of all the conjunctions of beliefs with which it would cause actions. Thus, the fact that the cupboard door will open smoothly will not come out as part of the truth conditions of the 'no impediments' belief because it will not be common to *all* conjunctions of conditions that guarantee success, only some of them. Clearly, the state of the cupboard door is irrelevant to the success of many actions. Nor, for that matter, will the state of the cupboard door be entailed by a belief whose truth conditions are all those states of affairs that are necessary for action. Thus, even if all of my beliefs are true, and even if one of my beliefs is a 'no impediments' belief, if I have no belief about the state of the cupboard door, my ignorance can undermine my success.

Moreover, the 'no impediments' belief is not, as Whyte argues, necessary for action. As Brandom points out, it is sufficient for an agent to act that he *does not believe* that there are physical impediments to his action; and the fact that he does not believe that there are physical impediments does not imply that he believes that there are no physical impediments. Unfortunately, this means that (R) is in trouble. Truth is not sufficient for success.

Truth is not necessary for success, either. Just as we can fail through ignorance, we can also succeed through good fortune. Suppose I falsely believe that there are cookies in the cupboard. Suppose, however, that Krister knows that the cookies are on the table. When I amble over to the cupboard, this act causes Krister to realise my mistake. Since Krister is a nice sort of fellow, his realising my mistake causes him to get the cookies from the table and give them to me. In this case, Krister's action brings it about that my desire is satisfied. My action is successful, although my belief is false.

In response to a similar objection, raised by Peter Godfrey-Smith,[24] Whyte admits that true beliefs are not necessary for the success of each action a belief would cause, since false beliefs can sometimes cause successful actions. Nevertheless, he claims, there is a necessary connection between truth and success because, 'only when a belief is true is the success of *every* action it would cause guaranteed'.[25] However, this seems like the claim that only true beliefs are *sufficient* for successful action; which, as I argued above, is false. Alternatively, Whyte may mean that although true beliefs are not necessary for success, true beliefs are necessary for *guaranteeing* success, where 'guarantees' means something other than 'necessary and sufficient'. But, now, we are owed an alternative interpretation of 'guarantees'. The suggestion that truth is necessary and sufficient for success was initially presented as a clarification of the claim that truth guarantees success. If truth is neither necessary nor sufficient for success, then what, after all, does it mean to say that the truth condition of a belief is that which guarantees the success of actions the belief would cause? (R), if not strictly false, is truly baffling.

---

[24] Godfrey-Smith 1994.

Another, serious difficulty with success semantics is the notion of 'success' which is used to generate truth conditions for beliefs. What is it for an action to be successful? Intuitively, we would say that my action succeeds when I get what I want, but what, exactly do I want? Desires, just like beliefs, have content; only desires have satisfaction conditions, rather than truth conditions. Hence, it looks as if Success Semantics is mired in circularity. It presupposes contentful desires, and so presupposes precisely what it seeks to explain.[26]

Whyte's solution to this difficulty is to augment (R) with a reductive account of desire satisfaction. This account is not supposed to analyse the content of all desires, but at least of some basic desires, so that we can break into the intentional circle. So long as the reductive account gives some content to some desires, other, more complex ones, can be built up from simpler ones. Even given this proviso, if the reductive account is to succeed, it must supply a determinate content for at least some desires, such that if *D* is the desire that p, it will be satisfied iff p. Whyte, then, rehabilitates Bertrand Russell's suggestion that satisfied desires are also psychologically fulfilling; when you get what you want, your desire ceases to plague you, it 'goes away'. Since we are thinking of desires as dispositions to behave, when a desire 'goes away', it simply ceases to be manifest and ceases to cause actions. On Russell's view, and on Whyte's, the satisfaction condition of a desire, *D*, is *p* if and only if *D* 'goes away' when *p* obtains. Unfortunately, Russell's suggestion is open to an obvious objection, famously made by Wittgenstein, that some conditions will make my desire go away although they are not, intuitively, the satisfaction conditions of the desire. For instance, my desire for an apple will go away if you punch me in the stomach, but it is not the case that, all along, I desired a punch in the stomach. Whyte argues that this difficulty can be avoided if we add another condition: that the desire goes away in the *appropriate way*. More specifically, *p* is the satisfaction condition of *D* iff *p* makes *D* go away and, in so doing, reinforces the disposition to act in the way that caused *D* to go away. If I want to eat an apple and I get a punch in the stomach, although my desire goes away, I will not try to get punched in the stomach the next time I want an apple. In contrast, if I want an apple, and I get an apple, then this will reinforce my disposition to do what I did to get the apple the next time I have the desire.

---

[25] Whyte 1997, p. 87.

[26] Teichmann 1992.

The trouble with this suggestion is that it fails to yield determinate satisfaction conditions, and therefore determinate content, for desires. For example, suppose that I want some chocolate, I believe that I can buy chocolate at the Health Food Store, but in fact, I buy myself some ingenious carob substitute for chocolate. Fooled by this fake chocolate, my desire for chocolate goes away and my disposition to buy this foodstuff when I want chocolate is reinforced. On Whyte's account, this implies that what I wanted all along was carob, rather than chocolate. But this is not a happy conclusion to draw—Whyte needs to be able to say that what I wanted all along was, indeed, chocolate, although I was fooled by the carob substitute. If my desire for chocolate is sometimes appropriately extinguished by chocolate and at other times appropriately extinguished by carob, why should we not say that my desire was, after all, the desire for chocolate or carob?

Whyte's response to this problem is to invoke the notion of 'normal conditions', so that, under 'normal conditions', my desire for chocolate will be satisfied by chocolate and chocolate alone. However, Whyte's 'normal conditions' are far from normal. He says that of the conditions that would satisfy my desire for chocolate, only those that would remain after an arbitrarily large improvement to my perceptual abilities would count. The satisfaction condition of the desire would be that condition that extinguishes my desire given an arbitrarily large improvement of my perceptual faculties. But this obviously falls prey to objections raised against similar views previously. In order to decide what counts as an 'improvement' of my perceptual abilities, assumptions have to be made about what I want, which is ultimately circular.

TELEOSEMANTICS

One temptation, at this stage, is to go teleological. Crucially, on the teleosemantic theory, biological purposes, or biological needs play the role played by desires in the success semanticist's theory. Since teleosemantics replaces desires with biological purposes, and builds biological purposes out of natural functions, it seems to avoid the circularity problem that success semantics faces. The teleosemanticist answers the sceptic by giving a Darwinian account of natural functions, and argues that those functions determine representational content for linguistic or mental terms. Moreover, teleosemantics is often presented as a satisfactory reduction of the normativity of meaning, since normative assessments seem to be appropriate to ascriptions of function. Something that has the function of φ-ing is *supposed to* φ, it *ought* to φ, and if it fails to φ, there must be a malfunction, some kind of mistake. Thus, it might seem as though teleosemantics is best placed to deal with Kripke's sceptical problem.

The immediate difficulty with functions is that when we think about things that have functions, the clearest cases are artefacts. The function of a radio is to transmit sounds, the function of a telephone is to enable telecommunication, and the function of a pencil is to write. Radios and telephones that transmit only static are *mal*functioning; broken pencils are pointless. But in each of these cases, the function of the artefact is derived from the intentions of the designer or user of the artefact. And there is no hope of a solution to the sceptical problem if we need to appeal to intentions to respond to the sceptic. This is allegedly where Darwin comes in. He provides the basis for an account of how, for instance, an organ can acquire a function, how it can be *designed* to do something, without our ever having to invoke the intentions of a designer. All you need for a wholly non-intentional, non-semantic account of function are natural facts and natural selection. For example, the heart is an organ whose proper function is to pump blood, not to produce a funky beat. The function of the heart is determined by its selectional history—pumping blood is the heart's function because that is what it was selected for, whereas producing a funky beat was irrelevant to this selection process. This is to say that creatures with hearts that were good at pumping blood out-competed similar creatures whose hearts were not so good at pumping blood (even if their hearts produced funkier beats). The function of the heart is also partly determined by the contribution it makes to the fitness of the organism that has it, insofar as that contribution explains why the trait of having a heart is preserved in subsequent generations. As Millikan explains:

> Proper functions are determined by the histories of the items possessing them; functions that were 'selected for' are paradigm cases. The notions of 'function' and 'design' should not be read, however, as referring only to origin. Natural selection does not slack after the emergence of a structure but actively preserves it by acting against the later emergence of less fit structures. And structures that can be preserved due to performance of new functions unrelated to the forces that originally shaped them. Such functions are 'proper functions' too and are 'performed in accordance with design.'[27]

Because the function of the heart is to pump blood, pumping blood is what the heart is for, and thus what it is *supposed* to do, what it *ought* to do. If it is possible to give a naturalistic account of functions, the teleosemanticist suggests, it will be equally possible to give a naturalistic account of meaning and content, in terms of what a system is *supposed* to do. The teleosemanticist tries to apply this kind of account of natural functions to the explanation of intentionality.

In her response to Kripke, Millikan claims to provide an account of rule-following and meaning that adapts the teleological account of the proper functions of such things as hearts to language.[28] What is somewhat peculiar about her response, however, is that she focuses on the question whether there can be a naturalistic account of rule-following in general, rather than, more specifically, a naturalistic account of following meaning-determining rules. She argues that hoverflies—who are clearly too inarticulate to follow explicit rules—can nevertheless be described as following implicit rules. Moreover, Millikan argues that *which* rules the hoverflies are following is fully determinate: they follow the rule that best explains their reproductive success. While this may be so, however, it is difficult to see how it will help to solve Kripke's problem. There may be a direct correlation between a male hoverfly's reproductive success and his following a rule that allows him to intercept the flight path of potential mates, rather than to fly right past them. But surely there is no direct correlation between my reproductive success and my following the rule for *addition*, at least when we compare *addition* with some other sceptical function that diverges from the *addition* function for numbers too large for me to grasp. Millikan only gestures at how the analysis of proper function can carry over to human rule following. She says:

---

[27] Millikan 1990, p. 86.

[28] Ibid.

> Ordinary human purposes, ordinary intentions, can only be a *species* of biological purpose. To suppose otherwise would be to suppose that the whole mechanism of human belief, desire, inference, concept formation, etc. ... [is] an accidental by-product of systems that nature designed for other purposes. ... The reasonable conclusion seems to be that ordinary explicit intending rests on biological purposing—biologically purposing to be guided by, or to react this way rather than that to, one's representations. Whether this biological purposing is innate ... or whether it is derived from learning, mechanisms of concept formation, etc., it must *ultimately* derive its content from the details of our evolutionary history.[29]

This claim is hardly objectionable, but it fails to deal directly with Kripke's sceptical problem. The sceptic wanted an account of what I mean that determines which of my uses accord with what I mean and which do not. To claim that ultimately our capacity to represent must have evolved along with us does not begin to touch the sceptical problem. For Kripke's sceptic can just as well say that we evolved to be good at getting around in the world without the capacity to 'represent' in the way that is assumed by the truth conditional picture of representation. Fortunately, Millikan and others have elsewhere defended teleosemantic theories of representation in detail, which might be pressed into service in response to the sceptic.[30]

Millikan argues that Darwinian evolution explains why representations generally have the function of supplying us with information about our surroundings, information that helps us succeed in the pursuit of our biological purposes. Concepts, however, need to be learned, and the process of learning can be understood in terms of an analogy with the process of natural selection, by which evolution occurs. By analogy, then, the proper function of an acquired characteristic or of an acquired concept is given by what it contributes to the functioning of the organism. Even in the case of a concept, it is what the organism needs, and how the concept allows the organism to satisfy its needs, that determines the proper function of the concept. Millikan suggests that she can specify which applications accord with the content of a concept and which do not on the basis of the organism's needs.

---

[29] Millikan 1990, pp. 86–7.

[30] Millikan 1984, 1995. Cf. Neander 1995; Papineau 1987; Stampe 1979.

Take an example (of Fred Dretske's).[31] Some marine bacteria have internal magnets, called magnetosomes, which point towards geomagnetic north (in the northern hemisphere) or to geomagnetic south (in the southern hemisphere). Oxygen is toxic to these bacteria which, in the northern hemisphere, propel themselves downwards towards oxygen-free water. Bacteria in the southern hemisphere have their magnetosomes reversed. If you put a bar magnet near these bacteria, they can be lured into an oxygen rich environment, where they die. On Millikan's view, the bacteria 'represent' the direction of oxygen, and they 'misrepresent' the direction of oxygen in the presence of the bar magnet. However, a sceptic could just as well argue that the bacteria represent the direction of the local magnetic field, so that they are not 'mistaken' when they move towards deadly, oxygenated water in the presence of the bar magnet. If this indeterminacy cannot be eradicated, then the sceptical problem seems to arise even for such primitive representations as these. What makes it the case that the bacteria represent the direction of high oxygen concentration rather than the local magnetic field? Millikan claims that one and only one answer is correct: the bacteria represent the direction of oxygen, not the orientation of the local magnetic field:

> What the magnetosome represents is only what its *consumers* require that it correspond to in order to perform *their* tasks. … What they need is only that the pull be in the direction of oxygen-free water at the time. … What a magnetosome represents, then, is univocal; it represents only the direction of oxygen-free water. For this is the only thing that corresponds (by a compositional rule) to it, the absence of which would matter, the absence of which would disrupt the function of those mechanisms that rely on the magnetosome for guidance.[32]

Millikan comes down quite heavily in favour of the needs of the organism determining what some structure in the organism represents. If concepts are taken to be analogous to such structures as magnetosomes, then we must assume that what determines a concept is similarly determined by the needs of the organism using the concept.

Unfortunately, it is not always clear that Millikan's hard line yields a theory that matches our intuitions with respect to concepts. Although we may want to say that the bacteria represent the direction of the oxygen, and so *mis*represent the direction of the oxygen in the presence of a bar magnet, there seem to be other cases in which we would not want the needs of the organism to determine the content of its representations. As Dretske has pointed out:

---

[31] Dretske 1986.

[32] Millikan 1995, p. 93.

> If I need vitamin C, my perceptual-cognitive system should not automatically be credited with the capacity for recognizing objects *as* containing vitamin C (as [meaning] that they contain vitamin C) just because it supplies me with the information required to satisfy this need. Representing things as oranges and lemons will do quite nicely.[33]

The point is that teleosemantic theories face a general dilemma in adapting the biological notion of proper function (for organs and the like) to concepts or representations in general. If *F*s and *G*s are correlated in nature, and I need *F*s, I have two options. Either I can represent the *F*s, and thus get what I need, or I can represent the *G*s and rely on the correlation to do the rest. In response to the problem of magnetosomes, Millikan suggests that the former strategy will always work, but this gives a peculiar result for the case of vitamin C.

Millikan might argue that new teleofunctions acquired during learning would determine the representational content of 'orange' and 'lemon'. That is, she might argue that it is oranges and lemons that I want (or need), and these needs are merely derivative of my biological needs. However a decision as to which needs determine the content of a concept starts to look arbitrary. The problem is that the needs of an organism are, so to speak, 'nested'.[34] Organisms do not merely have one need in a given situation, or a given type of situation, but many. I need to eat oranges in order to get vitamin C, in order to avoid getting ill, in order to avoid a premature death, in order to procreate and raise my children to childbearing age, and so on. Which of these needs determines the proper function of some representational item, a concept, such that it has the content *orange*, not *vitamin C*, or *health*, or *longevity*, or *procreation*? There seems to be no principled way to decide; indeed, we are pulled in different directions in different cases. So, any principled way of deciding is liable to contravene some of our intuitions. In general, organisms have diverse purposes and diverse functions, so there seems to be no good way to use purposes and functions to determine univocal, determinate representational content.

In response to this worry, Dretske argues that the emphasis on biological functions is necessary only to explain the emergence of creatures that have the ability to represent. However, if we are to give an account of the content of the concepts of representational creatures, what occurs during learning is crucial. He says:

---

[33] Dretske 1986, p. 32.

[34] Neander 1995.

> Suppose ... that we have a system capable of some form of associative learning. Suppose, in other words, that through repeated exposures to a *cs* (conditioned stimulus) in the presence of *F*, a change takes place. [Response] *R* (and, hence, avoidance behaviour) can now be triggered by the occurrence of *cs* alone. Furthermore, it becomes clear that there is virtually no limit to the kind of stimulus that can acquire this 'displaced' effectiveness in triggering *R* and subsequent avoidance behaviour. Almost any *s* can become a *cs*, thereby assuming 'control' over *R*, by functioning (in the 'experience' of the organism) as a sign of *F*.[35]

For example, imagine, as Dretske later suggests, a generic animal called 'Buster'.[36] Buster lives in an environment where there are furry (F) worms (W), or 'furms' of uniform size but varying colours. Initially, Buster has no concept of *furm*; although he sees furry worms, he does not see them *as* furms. One day, Buster touches a red furm and gets stung on his nose. Later, he tries to play with a green furm and once again, feels an unpleasant stinging sensation. Eventually, Buster acquires a conditioned response to furms: he retreats whenever he sees a furm approaching. Dretske argues that once Buster has acquired this discriminative response, he has acquired the concept of a *furm*. More precisely, Buster has acquired the concept of something that is furry (F), wormy (W), and stinging (S)—the concept of an *FWS-er*. Now, if you put Buster in a new environment, where there are fake furms, he will still respond as if they were FWS-ers, that is, he will retreat and avoid them. The best explanation of Buster's behaviour, according to Dretske, is that he believes himself to be in the presence of an FWS-er. Moreover, Buster is *mistaken* in thinking he is in the presence of an FWS-er because he is really in the presence of a caterpillar, which will not sting him if he goes near.

Fodor has argued that Dretske's proposed reduction of meaning to such facts about behaviour during the learning period presupposes that there is a sharp distinction between when a concept is being acquired and the subsequent period when the concept is applied.[37]Buster's acquiring the concept *furm* is dependent on the fact that his conditioned response was caused by furms, and is explained by the causal properties of furms.

---

[35] Dretske 1986, p. 35. [36] Ibid. 2000. [37] Fodor 1990.

Whatever it is that causes and explains Buster's conditioned response must determine his meaning. Second, Fodor argues, Dretske suggests that during the learning period, a speaker is immune to error. But this is not so. Suppose, for instance, that I am learning the concept of *horse*. Perhaps I acquire a conditioned response to horses—for instance, in the presence of horses, I learn to behave in certain determinate ways. Now, as Fodor complains, there is no way to guarantee that my *horse* thoughts will not be caused by distant, horsey-looking cows *while* I am learning the concept. If I can make a mistake after I have acquired the concept, why am I immune to error while I am learning it? Of course, in some sense I cannot make a mistake in applying the concept *horse* while I am learning it simply because I have not yet *acquired* a concept with a determinate content. The trouble is that if whatever causes my tokenings of the concept *horse* during the learning period determines what I mean by it, and if I am caused to think *horse* thoughts by horsey looking cows, then what I mean by *horse* will be *horse or cow*. That is, if there is a problem of ruling out errors from constituting what I mean once I have learned a concept, there will be a problem of error before the learning period is over.

Dretske's response is that he does not presuppose a sharp distinction between the learning period and the application period. Buster can acquire a new conditioned response and thereby acquire new concepts or alter his existing ones. For example, suppose that he chances upon a harmless caterpillar and so learns to distinguish furms from caterpillars. But here, the sceptic will argue that before Buster acquired a new conditioned response to caterpillars, it was indeterminate whether 'furm' would extend to caterpillars or not. When confronted with the novel case of the caterpillar, Buster could either decide that caterpillars are not furms or that some furms do not sting. To decide which would be the correct extension of the concept *furm* requires that we be able to distinguish between Buster having the concept of a furry wormy thing that sometimes stings from the concept of a furry wormy stinging thing. And nothing about Buster's conditioned response seems to make that distinction. So, there is nothing in Buster's conditioned response that seems to determine whether *furm* applies correctly to caterpillars or not.

Thus, it seems that the difficulty with the teleosemantic theory, even when augmented by a story about how concepts are learned, fails to eliminate the indeterminacy that fuels the sceptic's fire.

ASYMMETRIC DEPENDENCE

Fodor's theory of mental representation is another of the family of broadly dispositional theories. That is, Fodor takes the content of a given representation to be reducible to facts about what would cause the representation. On such a theory, what makes it the case that the concept 'cow' means *cow* and refers to all and only cows is the fact that 'cow' thoughts would be caused (fundamentally, normally or ideally) by cows. The difficulty for such a theory will by now be familiar: 'cow' thoughts are sometimes caused by non-cows, including buffalo when seen on dark nights, thoughts about milk, hearing the command 'name the first farm animal that comes to your mind', and so forth. In order to rule all of these causes out of the extension of 'cow', it must be possible to specify, in non-intentional terms, and without begging the question, what makes it the case that the fact that cows cause 'cow' thoughts determines the representational content of 'cow'. If any given representation refers to all the things that cause it and if what causes a representation determines what it represents, then we can never *mis*represent.

Fodor's theory starts with the intuition that for something to represent, say, cows, is for it to carry information about cows. Since the capacity for something to carry information is a function of its causal covariance with what it carries information about, Fodor claims that the fact that 'cow' represents cows is (roughly) reducible to the fact that thoughts involving the concept 'cow' are caused by cows. But this claim needs to be augmented, in order to solve the problem of error. Fodor argues that for 'cow' to mean *cow*, the fact that cows cause 'cow' thoughts must be fundamental; if 'cow' means *cow* then the fact that non-cows cause 'cow' thoughts *depends* on cows causing 'cow' thoughts, but not vice-versa. This is the 'asymmetric dependence thesis', which Fodor formulates (to a first approximation) as follows:

> Cows cause 'cow' tokens, and (let's suppose) cats cause 'cow' tokens. But 'cow' means *cow* and not *cat* or *cow or cat* because *there being cat-caused 'cow' tokens depends on there being cow-caused 'cow' tokens, but not the other way around.* 'Cow' means *cow* because, as I shall henceforth put it, noncow-caused 'cow' tokens are *asymmetrically dependent upon* cow-caused 'cow' tokens. 'Cow' means *cow* because *but that 'cow' tokens carry information about cows, they wouldn't carry information about anything.*[38]

---

[38] Fodor 1990, p. 91.

Fodor's solution to the problem of error is this. If *any* non-cow caused 'cow' thought is dependent on there being cow-caused 'cow' thoughts, then 'cow' thoughts are about cows even when they are caused by buffalos, or by thoughts about milk, or by expressions of the desire to see cows and so forth. That is, Fodor's theory, as it is presented, aims to *exclude* non-cows from the extension of 'cow'. However, Fodor's theory faces a problem of *inclusion* as well—that his theory cannot ensure that *all* cows are in the extension of 'cow'.

In responses to objections, Fodor discusses the Old Paint objection (which he attributes independently to Steven Wagner, Tim Maudlin, and Scott Weinstein). The Old Paint objection goes as follows. Consider the horse called Old Paint. It is plausible that Old Paint would not cause 'horse' thoughts except that horses other than Old Paint do. And it is also plausible that horses other than Old Paint would cause 'horse' thoughts even if Old Paint would not. So, Old Paint's causing 'horse' thoughts is asymmetrically dependent on horses other than Old Paint causing 'horse' thoughts. So, 'horse' means *all the horses except Old Paint*.[39]

Fodor's response to this problem of inclusion is to insist that it is the *property* of *being a horse* that is causally responsible for 'horse' thoughts—it is in virtue of instantiating the property of *being a horse* that horses cause 'horse' thoughts, not in virtue of instantiating the property of *being a horse other than Old Paint*. Although it is true that if the property of *being a horse* did not cause 'horse' thoughts, then Old Paint would not cause 'horse' thoughts, it is also true that if the property of *being a horse* did not cause 'horse' thoughts, then the horses other than Old Paint would not cause 'horse' thoughts. Thus, it is not the case that the fact that Old Paint causes 'horse' thoughts asymmetrically depends on other horses causing 'horse' thoughts.[40]

Perhaps the Old Paint objection can be reformulated so as to have more bite. Suppose that Old Paint is a very funny looking horse; such a funny looking horse, in fact, that most of us (and maybe even other horses) would not think he is a horse. By some genetic mishap, Old Paint, although born of a mare, looks just like a zebra. The point is that despite exemplifying the property of being a horse, Old Paint would not cause 'horse' thoughts, although horses other than Old Paint would. Horses other than Old Paint, it seems, cause 'horse' thoughts in virtue

[39] Fodor 1990, p. 102.

[40] Ibid., p. 103.

of exemplifying the property of *looking like horses*, a property that Old Paint does not exemplify. So, if Old Paint would not cause 'horse' thoughts, and horses other than Old Paint would, then on Fodor's view, 'horse' means *horsey looking horse* and refers to all and only horses other than Old Paint.

Fodor might argue that although Old Paint does not look like a horse, if he is a horse, then it is likely that we would discover him to be a horse on closer examination. Because Old Paint would cause 'horse' thoughts, at least upon further investigation, then it must be the property of *being a horse* that causes 'horse' thoughts, not the property of being a *horsey-looking horse*. However, consider the case more closely. The thought is that if we investigated further—if, for example, we took a gene sample of Old Paint—we would discover it to be a horse. However, if we were to take a gene sample of Old Paint, and then tokened 'horse' thoughts, it seems correct to say that it is the *gene sample* that would cause the 'horse' thought even though Old Paint would not cause 'horse' thoughts. But 'horse' does not refer to horses and gene samples. On Fodor's view, if 'horse' means *horse*, then the fact that gene samples cause 'horse' thoughts must asymmetrically depend on the fact that horses cause 'horse' thoughts. So, even if gene samples taken from Old Paint would cause 'horse' thoughts, if Old Paint would not cause 'horse' thoughts, then 'horse' means *horsey-looking horse* and refers to all the horses except Old Paint.

Furthermore, for the response to the Old Paint objection outlined above to work, it must be the case that any horse would cause 'horse' thoughts if investigated sufficiently thoroughly. However, for anyone who is a realist about properties, it is in principle possible for something to instantiate the property of being an F even if no one would recognise it as an F.[41] Thus there might be horses that would not cause 'horse' thoughts, protons that would not cause 'proton' thoughts, square things that would not cause 'square' thoughts, and so forth. Moreover, it is difficult to see how Fodor could abandon realism about properties, since it is the properties in virtue of which representations are caused that determine the content of the representations. If Fodor were to abandon realism, what we call a horse or what we would call a horse could not in turn determine what things instantiate the property of being a horse. If Fodor is a realist about properties, which he seems to be,[42] then it is in principle possible for Old Paint to exemplify the property of being a horse even though (for whatever reason) no one would recognise Old Paint to be a horse. The upshot, once again, is that 'horse' does not mean *horse*, but *horsey-looking horse* and does not include Old Paint in its extension.

---

[41] Cf. Boghossian 1989.

[42] Fodor 1990, p. 93.

In Fodor's response to the original Old Paint objection, he stresses that he merely *stipulates* that the property of being a horse is the property causally responsible for 'horse' thoughts.[43] Fodor merely stipulates this because all he wants to say is that *if* the property of being a horse is the property causally responsible for 'horse' thoughts, *then* 'horse' means *horse* and refers to all (and only) the horses. And it is only *if* all non-horse caused 'horse' thoughts asymmetrically depend on horse-caused 'horse' thoughts that 'horse' means *horse* and refers to all and only horses. So, in response to the revised Old Paint objection, Fodor might argue that *if* the property that is causally responsible for 'horse' thoughts is the property of *being a horsey looking horse*, then 'horse' does mean *horsey looking horse*, not *horse*. On the flipside, if the property that is causally responsible for 'horse' thoughts is the property of *being a horse* then 'horse' means *horse* and all is well.

This answer might be satisfactory if we had an adequate account of what would make the property of *being a horse* causally responsible for 'horse' thoughts, but not the property of *being a horsey looking horse* (or some other suitable candidate). However, causally responsible properties often come in groups—properties are jointly causally responsible, but not without their team mates. For the meaning of 'horse' to be determined by just one of those properties—such as the property of *being a horse*—we need some principle that determines which of the causally responsible properties is the one (or ones) that determine representational content. Without such a principle, Fodor's theory has the counterintuitive consequence that our concept of 'horse' is either indeterminate in content or means something bizarre like *horsey looking horse in sufficient light* …

According to Fodor, 'P is a causally responsible property if it is a property in virtue of the instantiation of which the occurrence of one event is nomologically sufficient for the occurrence of another.'[44] This suggests that there is always just one property that is nomologically sufficient for an effect and thus causally responsible for it; but very often, we find that several properties are jointly nomologically sufficient for their effects, and so only jointly causally responsible. For example, consider the causal law that, ceteris paribus, the striking of a match will cause it to light. At first this law reads as though it is the property of being a (sufficiently hard) striking of a match that is nomologically sufficient and so causally responsible for its lighting. But if we fill out the 'ceteris paribus' clause, things look different. If a match is wet or if the striking occurs in insufficient oxygen, the match will not light. With the 'ceteris paribus' clause partly spelled out, the law should read: 'Ceteris paribus, if you strike a dry match in sufficient oxygen, the match will light'. The striking of the match is not nomologically sufficient for its lighting on its own; it is the properties of being a striking of a dry match in sufficient oxygen that are nomologically sufficient for the lighting of the match. Indeed, since none of the properties could be sufficient for the lighting of the match on its own, it seems that the three properties together are jointly nomologically sufficient for the effect, and so only jointly causally responsible.

---

[43] Ibid, p. 102.

[44] Fodor 1990, p. 143.

Let us return to horses. Could the occurrence of the instantiation of the property of being a horse be nomologically sufficient for the occurrence of 'horse' thoughts? Intuitively, it would seem not. Horses do not cause 'horse' thoughts if they look like Old Paint, if they are seen on dark nights, or when very far away, or hiding behind barns, or when we are simply not paying attention (because preoccupied by weighty philosophical problems, of course). That Black Velvet instantiates the property of being a horse is hardly sufficient for Black Velvet to cause 'horse' thoughts. At the very least, a horse needs to be in my perceptual field in order to cause 'horse' thoughts in me. What Fodor might like to say is that the law on which all the others asymmetrically depend is this: in circumstances of kind *C*, horses cause 'horse' thoughts, where circumstances of kind *C* include such conditions as not being hidden, being in sufficient light and close enough to be seen. That is, he might want to say that there are circumstances or background conditions under which the law that horses cause 'horse' thoughts holds (and other circumstances in which it does not hold), but it is the property of being a horse that is really the only causally responsible property.

This argument requires that there be a sharp metaphysical difference between causally responsible properties and 'background' conditions. However, the distinction between causally responsible properties and those that are in the background seems to be epistemological not metaphysical. If we want to explain why this match lit, but the other match did not (given that both were dry and in sufficient oxygen), we will naturally cite the striking of the match that lit. If we want to explain why one match lit when struck but the other did not, we will naturally cite the wetness of the second match. What is a 'background' condition in the first explanation is a causally responsible property in the second. Our explanatory methods may be reliable guides to discovering what the causes are, but they do not distinguish between the causes that are 'really' in the background and those that are 'really' causally responsible.

Similarly, if we want to explain why Jones tokened a 'horse' thought, but Smith did not, we might cite the fact that there was a horse in front of Jones but not in front of Smith. However, if there is a horse in front of both of them, but Jones is in the light and Smith in the dark, we will cite the fact that the horse in front of Jones was well lit. In that case, it looks like it is the light rays bouncing off the horse and into Jones's eyes that caused the 'horse' thought, since in the absence of that condition, in Smith's case, a 'horse' thought was not caused. The point is that what is relevant to a particular explanation will single out one or a few of the causally responsible conditions, but will not distinguish from among the causally responsible conditions those that are 'really' causally responsible. Thus, it seems that Fodor cannot claim that the property of being a horse is *the* causally responsible property and so determines the content of 'horse' thoughts.

Another option Fodor might try is applying Mill's Method of Agreement, which involves looking at several cases in which an effect is caused to see which conditions are always present.[45] Applying the Method of Agreement, we might find that the property of being a horse is the one property present in *all* of the conjunctions of causal conditions that have 'horse' thoughts as an effect. That is, although it is true that instantiating the property of being a horse is not sufficient to cause 'horse' thoughts, the property of being a horse is common to all the conjunctions of properties in virtue of which horses cause 'horse' thoughts. A horse would cause 'horse' thoughts even if not sufficiently well lit so long as it is audible. A horse that is inaudible would cause a 'horse' thought so long as it is sufficiently well lit. What we have really is a disjunction of causally responsible properties—being a horse and being sufficiently well lit or being a horse and being audible or being a horse and being close enough to touch and so forth. What is common to each disjunct is the property of being a horse, so it is the property of being a horse that can be singled out as the property that determines the content of 'horse', because it is the only property that is common to all the conditions under which 'horse' thoughts are caused.

---

[45] Mill 1904, pp. 253–6.

This would be a nice response to the problem, were it not for the familiar but nevertheless irritating problem of misrepresentation. It is not the case that the property of being a horse is common to all conjunctions of properties in virtue of which 'horse' thoughts are caused since horsey looking cows cause 'horse' thoughts, as do pictures of cowboys and thoughts about farm animals. Fodor rules these out on the grounds that non-horse caused 'horse' thoughts asymmetrically depend on horse-caused 'horse' thoughts. However, if it is plausible to suppose that non-horses would not cause 'horse' thoughts if horses did not cause 'horse' thoughts, it is equally plausible to suppose that sounds of neighing in the dark would not cause 'horse' thoughts if well-lit horses, or neighing horses did not cause 'horse' thoughts. It is also plausible that unusual looking horses (such as Old Paint) would not cause 'horse' thoughts (upon further investigation) were it not that horsey-looking horses cause 'horse' thoughts. Thus, the fundamental law, on which the others asymmetrically depend, is that things that have the properties of being horses, of being well lit, of looking, sounding, smelling, and feeling horsey are jointly causally responsible for causing 'horse' thoughts. If this is so, then 'horse' means *horsey-looking*, *horsey-smelling*, *horsey-sounding*, *well lit (etc.) horse* and does not refer to Old Paint, perfumed horses, mute horses, and horses in the dark.

There is a further difficulty, which is to specify, in non-intentional terms and without begging the question, whether *horse* refers to horses, the property of *being a horse*, or horsey-events. Fodor's response to the original Old Paint objection was to claim that it is the property of *being a horse* in virtue of which horses cause 'horse' thoughts and that this is why 'horse' refers to all and only horses. That is, on Fodor's view, horses (individuals) cause 'horse' thoughts and so 'horse' refers to the individuals that cause the 'horse' thoughts. The implicit assumption is that objects are causes of 'horse' thoughts, as opposed to events, situations, or facts. However, according to some philosophers,[46] causal relations hold between *events*. Every time I have a 'horse' thought, it is caused by an event, such as my catching sight of a horse running by or my becoming aware of the sound of hooves or neighing, or, events such as that of horses appearing in my visual or auditory (or more generally,

[46] e.g. Davidson 1980.

perceptual) field. If we take, for example, my catching sight of a horse as a cause of my 'horse' thoughts, then it looks as though, on Fodor's theory, what 'horse' means is *horse sighting*—the term refers to types of *events*, rather than types of objects. Even if we take a more distal example, such as horses entering my perceptual field, still 'horse' will mean *horse entering my perceptual field* or, as it were, *horsey event*, as opposed to *horse*.

Furthermore, since Fodor claims that it is the property of being a horse in virtue of which horses cause 'horse' thoughts, then arguably, 'horse' means not *horse* but *horsiness* and refers to the property of being horsey, rather than to all and only horses. Thus, even if we grant that it is the property of being a horse in virtue of which horses cause 'horse' thoughts, we have several suitable candidates for the representational content of 'horse': horsey events, the property of horsiness and horses. What reason could we have for choosing one over the other?

Perhaps Fodor would say that events are reducible to properties of objects or changes in properties of objects, so that objects are the primary causal relata, although they cause in virtue of their properties. However, if a causal theory of representation is to apply across the board, it cannot choose only objects as causal *relata* and therefore the sole determinants of representational content. For, we have mental representations that refer to properties (e.g. redness, solidity) and mental representations that refer to events (e.g. explosion, collision) as well as mental representations that refer to objects. If we are to give an account of the representational content of, say, 'explosion', we will want to say that 'explosion' thoughts are caused by explosions, that it is the property of being an explosion in virtue of which 'explosion' thoughts are caused. But if we decide in favour of objects as causal *relata* and therefore determinants of meaning, then 'explosion' refers not to explosions (events) but exploding things (objects). Since exploding bombs cause 'explosion' thoughts, we will be forced to conclude that 'explosion' means *bomb* as opposed to 'explosion'.

Thus, whether the real *relata* of causal relations are objects or events we will not have a suitable answer to the question of what determines the content of a representation. Rather, irrespective of what the causal *relata* ultimately are, we need to be able to non-arbitrarily say that 'explosion' thoughts refer to events and 'horse' thoughts refer to objects. This problem, as well as the problem of causal responsibility, besets the very meaning-determining causal laws on which all of the others asymmetrically depend.

THE CAUSAL THEORY OF REFERENCE

Kripke's arguments against naturalism focus on dispositionalism. However, some have suggested that an externalist account of reference—of the kind Kripke has himself elsewhere propounded[47]—might be more suitable in response to Kripke's sceptic. An externalist theory of reference, unlike an internalist theory of reference, is one according to which 'meanings ain't in the head'.[48] What makes it the case that I mean, say, *water* by 'water' is the fact that 'water' is used, by English speakers, to refer to a clear, colourless liquid found in abundance on Earth. What 'water' refers to is not determined by some finite state in our heads, but in the fact that someone, in the past, dubbed this clear, colourless, thirst quenching liquid 'water'. This dubbing event determined that 'water' refers to the stuff in the initial sample, the natural kind of stuff, and we have subsequently 'borrowed' reference to this natural kind. What we mean by 'water' is just that stuff that the erstwhile dubber referred to with the use of that word. If we imagine that we were all on a planet called Twin Earth, which is exactly like earth, but where the clear, colourless, thirst quenching liquid in lakes and ponds has the chemical composition XYZ rather than H_2O, then 'water' would refer to XYZ, rather than H_2O.[49] For the internalist, switching worlds like this should make no difference to what we mean; for the externalist, it makes all the difference in the world.

A causal theory of reference, of the kind just sketched, seems to promise certain advantages over the dispositional theory, as well as other theories criticised by Kripke. For one thing, a causal theorist should have no difficulty dealing with the potential infinitude of extensions—if meanings 'ain't in the head', then there should be no *prima facie* reason why a finite mind with finite capacities could not use an expression with an infinite extension. Externalism may also seem to promise a solution to the problem of error. If the content of 'water' is determined by a dubbing relation to the external, mind-independent stuff to which 'water' was initially used to refer, then this supplies an external, mind-independent fact of the matter as to which uses of 'water' are correct. If water refers to H_2O, then if I say that there is water in a glass full of vodka, then I have made a mistake.

---

[47] Kripke 1972.

[48] Putnam 1975.

[49] Ibid. 1975.

Colin McGinn has defended just such a Kripkean response to Kripke.[50] McGinn argues that in order to answer the sceptic, we need to be able to specify some relation, *R*, between a name and an object, such that if the name does not so relate to a given object, then any application of the name to that object will not accord with what the name means. He continues:

> [W]hy not say that *R* is the (or a) causal relation, and hence that a use of a name is correct just if it involves applying the name to the object which lies at the origin of the causal chain leading up to that use? In the case of a natural kind predicate we can likewise say that such a predicate is correctly applied to an object just if that object is of the same kind as the original sample which initiated the causal chain leading up to that use. Does this type of theory exclude the kinds of non-standard extension contrived by the semantic sceptic? It seems to me that it does, since it will not be true that the non-standard extension figures as the *causal origin* of the use of the name or predicate.[51]

At first pass, this seems highly compelling. According to the theory, there is an initial dubbing event, or baptism, which determines the reference of a name. For instance, a sample of gold was initially dubbed 'gold', I was named 'Anandi', water was dubbed 'water'. Thenceforth, 'Anandi' has been correctly applied to somebody if and only if that somebody is me (or someone else called 'Anandi') and 'gold' has been correctly applied to something if and only if it is gold. If you call a copper pot 'gold', you have made a mistake because the stuff you are calling 'gold' is of a different natural kind from the sample initially dubbed 'gold'.

The strength of the causal theory of reference lies in its specification of a single instance which seems to determine the whole extension of a name. This strength is also the theory's weakness. The trouble is that a sample that is dubbed can instantiate numerous properties: a sample of water is also a member of the kind, liquid; a snail is a mollusk and an invertebrate; a sample of gold is a member of the kind, gold, as well as the kind, metal. The question now becomes how the causal theorist can fix just *which* of those properties constitutes the natural kind of which the sample is to be taken as a paradigmatic member. Thus, the sceptic can ask what makes it the case that, at the initial dubbing event, it was the natural kind H_2O that was dubbed 'water' and not the natural kind liquid, or the individual sample (the pond, for instance). This is called the 'qua' problem—since it seems to be indeterminate whether, at the initial dubbing event, the sample water was dubbed *qua* member of the natural kind H_2O or *qua* member of some other natural kind, such as liquid. Unsurprisingly, a similar difficulty affects the causal theory of proper names. When I dub my cat 'Felix', do I name the whole cat 'Felix' or merely the side of the cat that is closest to me? Do I name the whole universe plus the cat, or do I name everything but the cat? Or, for that matter, do I designate the natural kind, feline?

---

[50] McGinn 1984. See also Maddy 1984.

[51] Ibid., pp. 165 ff.

The solution that is often presented to the 'qua' problem is that an intention under some description must be brought into play to fix the referent of a term coined at the initial dubbing event.[52]Thus, H_2O is dubbed 'water' in the initial event, because the initial dubber had intended to dub the *substance* of which the sample was a sample; it was qua substance that the initial dubber named water 'water'. Although this answer may satisfy some, it is unlikely to satisfy the sceptic. By introducing an intention in determining the referent of 'water', the causal theorist invites the sceptic to ask what constitutes the content of the initial dubber's intentions. The sceptic's line will no doubt be familiar: perhaps when she said to herself, I hereby dub this qua substance, she really meant *schmubstance* by 'substance', and if she meant *schmubstance* by 'substance', then 'water' applies correctly to H_2O or gold. Of course, one might appeal to the dubber's dispositions to determine what she meant by 'water' in the initial dubbing event, but this will face the very problems beset by the dispositional theory, and which we had hoped that the causal theory would solve.[53] The sceptic would thus argue that it is wholly indeterminate what the initial dubber called 'water' when it was initially dubbed, and thus wholly indeterminate which circumstances make the use of the word 'water' correct.

An alternative solution to the 'qua' problem is to rely on the dispositions of speakers subsequent to the initial dubbing event. If, after some gold has been dubbed 'gold', speakers are disposed to apply 'gold' to gold, but not other metals, then 'gold' refers to gold, as opposed to metal. However, this solution to the 'qua' problem is obviously susceptible to the criticisms raised against the pure dispositional theory. Although a person is disposed to apply 'gold' to gold things, she will also be disposed to apply 'gold' to non-gold metals, under unfavourable circumstances. What is needed is a non-circular specification of which disposition is meaning-constituting. This can no longer come from the initial dubbing event, where it is indeterminate whether 'gold' refers to the property of being gold or the property of being metal. Invoking the speaker's disposition can hardly solve the problem with the dispositional theory. Thus, the causal theory seems to be no better able to solve the sceptical problem than dispositionalism.

---

[52] Devitt and Sterelny 1993

[53] Kitcher and Stanford 2000.

COLLECTIVISM

Most of the theories I have so far considered have been broadly *individualist* theories. One exception might be the direct causal theory, since it appeals to communal 'reference borrowing' going back to an initial dubbing event. Several of the best-known proponents of externalism have combined causal/physical features of the speaker's environment, and the dispositions of her peers, in their theories of meaning. Putnam, for instance, who invented the Twin Earth thought experiment, also suggests that expert knowledge in a community can be an essential determinant of individual meaning.[54] Teleological theories are also not strictly individualistic because it is the evolutionary history of the species to which the speaker belongs that determines her functions, and thus (partially) determines her meaning. Nevertheless, these theories seem only to be 'weakly' collectivist: in the initial dubbing event it is presumably the speaker's intention that determines meaning; for the teleosemanticist, concepts can be acquired by individuals through learning.

Tyler Burge, another of the main proponents of externalism, more directly defends the idea that the community plays an essential role in determining meaning. Burge introduces thought experiments similar to the Twin Earth case, except that what differed between the actual and counterfactual cases is the communal practice in the use of the relevant word. Thus, he says that a person's

> mental contents differ while his entire physical and non-intentional mental histories, considered in isolation from their social context, remain the same. . . . The differences seem to stem from differences 'outside' the [person] considered as an isolated physical organism, causal mechanism, or seat of consciousness. The difference in his mental contents is attributable to differences in his social environment.[55]

---

[54] Putban 1975.

[55] Burge 1998, p. 28.

Because he takes mental content to be partly determined by a person's 'social context' or 'social environment', Burge can be called a 'collectivist'. Nevertheless, Burge does not give an account of *what it is* in the social context or environment that determines an individual's meaning. Burge suggests that a 'social practice' is the primary determinant of meaning, but this is broadly speaking what any collectivist takes to be essential. The sceptic is going to be concerned with how the idea of a social practice is to provide an adequate account of what the speaker means.

Some collectivists appeal to primitive intentional, semantic, or normative facts; I will discuss these versions of collectivism in the following chapter.[56] Kripke's sceptic is also a collectivist insofar as he argues that agreement between two or more members of a community justifies our discourse of ascribing meanings, however his sceptical solution also embraces a non-factualism about meaning. In this chapter I restrict the discussion to *naturalistic* forms of collectivism—collectivists who appeal only to *natural* (i.e. non-intentional, non-semantic, non-normative) facts about the members of a linguistic community. The only naturalist form of collectivism that has been explicitly defended is *communal dispositionalism*. Other natural facts, external to the individual, will also be external to the other members of the community, and so an appeal to those facts does not make membership in a *community* essential to meaning.

A communal dispositionalist maintains that the dispositions of all the members of a linguistic community to use an expression, x, determines the correct uses of x for each individual, and thus determines what each speaker means. Whether or not a particular use of x is correct on some occasion, depends on what others in the relevant community would do:

> Making a step in following a rule counts as a 'right' step, i.e. a genuine and successful piece of rule-following, if it is aligned with the steps everyone else, or nearly everyone else, takes. ... To make a 'wrong' move is ultimately to make a move that leads the individual along a divergent path. To be wrong is to be deviant.[57]

Thus, the communal dispositionalist purports to supply a clear account of error. If the individual is out of synch with the community, if she uses a word in a way that others would not, she has made a mistake.

---

[56] Cf. Brandom 1994*a*, McDowell 1993, 1998.

[57] Bloor 1997, p. 16. Bloor arguably does not defend communal dispositionalism, but something more like non-factualism. Nevertheless, this passage provides an accurate account of the theory.

The problem with this view is that if individual dispositions cannot fix meanings, it is difficult to see how meanings could be fixed by bringing individuals together to form a collective.[58] The very same arguments that were addressed to the individual dispositionalist can be repeated at the level of communal dispositionalism, and to similar effect. To begin with, Kripke's argument that there is no way to make a principled distinction between error-producing and meaning-constituting dispositions applies to the communal dispositionalist as well. The communal dispositionalist claims that the correct action is the one that we are all disposed to make. But we might all share a variety of dispositions, some of which are meaning-constituting and others of which are error-producing. If the members of the community share both 'error-producing' and 'meaning-constituting' dispositions, then how are they to be distinguished?

Imagine, for example, that I think *lo, a horse* when I come across a deceptively horsey looking cow on a dark night. If it is sufficiently dark, and the cow looks sufficiently like a horse, there is every reason to suspect that anyone else under those circumstances would have reactions just like mine. Let us just say, for the sake of argument, that there are certain conditions under which everyone *would* think *lo, a horse* in the presence of a particularly horsey looking cow on a particularly dark night—including the experts. Even if this example is far fetched, there certainly are an abundance of examples from history where everyone has been mistaken. Since everyone's dispositions determine facts about correctness, it turns out that we all mean something non-standard by 'horse', and make no mistake when we apply 'horse' to cows on dark nights. Thus, the communal dispositionalist fares no better than the individual dispositionalist in ruling out bizarre, disjunctive meanings for all of our terms.

Collectivists are prone to suggest that this does not matter. The dispositions of the members of the community determine whether an *individual*'s use of an expression is correct, but the community *taken as a whole* is neither correct nor incorrect. Thus, for instance, Wright says:

> What is it for the community to recognise that it here continues a pattern of application of an expression on which it previously embarked? 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? It is unclear how we

---

[58] Blackburn 1984; Boghossian 1989.

> can answer. We are inclined to give new linguistic responses on which there is securable consensus the dignity of 'objective correctness'; but we have, so to speak, only our own word for it.[59]

Wright suggests that the communal dispositionalist fails to determine a single, correct, 'ratification-independent' pattern of application for expressions. However, the communal standard *replaces* the ratification-independent pattern, or what we usually take to be the correctness condition of the term. The natural intuition is that 'horse' applies correctly to some *a* if and only if *a* is a horse, that is, if and only if *a* has some objective properties whereby it qualifies for inclusion in the extension of 'horse'. Wright suggests that the communal dispositionalist replaces this standard of correct application with something like this: 'horse' applies correctly to some *a* (in some community, *C*) if and only if the members of *C* are disposed to apply 'horse' to *a*. This can supply a perfectly *determinate* standard for the individual, although there is no further standard for the community to meet.

While collectivists might be tempted to embrace Wright's suggestion, it is not clear whether a communal dispositionalist can take solace in it. At least for our present purposes, we need to consider whether introducing the community helps us develop a 'straight', reductive response to the sceptic, that is, one that will capture what we take to be the correctness conditions of our terms. This will not be accomplished, however, if the pattern of communal behaviour makes it indeterminate whether the standard generated by the community is the intuitive one or the sceptical one. Rather, the communal standard must yield a correctness condition that adequately reduces the correctness condition we take our expressions to have.

John Haugeland presents a very clear account of how semantic content might emerge when a group of individuals get together to form a community. I should note that although I am referring to this as a form of *reductionism*, Haugeland calls it 'neo-pragmatism'. Nevertheless, I see nothing in his theory that distinguishes it from reductionism—he accounts for meaning in terms of the dispositions of the members of a community, and adverts to nothing more than 'natural' features of individuals to explain the emergence of norms.

Haugeland's central idea is that when certain kinds of individuals get together—individuals with certain dispositions and abilities—they can create standards that supply a distinction between correct and incorrect applications and that reduce semantic oughts. The 'proto-normative' creatures are characterised by Haugeland simply as *conformists*. This means that they have the following three dispositions: (1) the disposition to imitate one another; (2) the disposition to censor one another's behaviour when it fails to conform to their own; and (3) the disposition to alter their own behaviour in response to the censorship of others. These dispositions, Haugeland supposes, are wholly natural, they are 'wired in'. Moreover, he says, this picture only 'presupposes innate capacities to react differentially (via sensory discrimination, say), to adapt behaviourally (that is, to learn, as in conditioning or habit formation), and to influence the adaptation of others (by setting an example, reinforcing, punishing, and such like).'[60]

---

[59] Wright 1980, p. 219.

Haugeland's key idea is that a community so described will develop convergent dispositions to behave, and thus there will begin to be something like a *normal* practice within a community. Furthermore, what is normal is enforced normatively by collective sanctions, making it 'right' to act in a way that conforms and 'wrong' to do otherwise. Each individual simply does as she is disposed to do, sanctioning when she sees fit; but because her dispositions are shared within the group, she is *tacitly* enforcing the group-wide norms. Haugeland explains this as follows:

> The community-wide classes of similar dispositions that coalesce under the force of conformism can be called 'norms'—and not just collections or kinds—precisely because they themselves set the standard for that very censoriousness by which they are generated and maintained. ... Out-of-step behaviour is not just atypical, but abnormal and unacceptable; it is what one is 'not supposed to' do, and in that sense improper. ... *Abiding* by a norm or custom (engaging in a practice) is behaving in the manner required by that norm, and not merely by coincidence, but as the exercise of the dispositions fostered by that norm.[61]

This account has two major weaknesses. I have already mentioned one of these above: that a communal dispositionalism of this form will give non-standard accounts of the meanings of most of our expressions. We may all be disposed to mistake a cow for a horse on a dark enough night, thereby making the meaning of 'horse' non-standard or indeterminate.

---

[60] Haugeland 1998, p. 148.

[61] Ibid., p. 149.

A second worry is that the account might be circular; intentionality seems to have crept in through the back door. In order for the dispositions to coalesce, the individual members of the community need to be able to alter their behaviour in response to sanction. This requires that the individuals can tell how to go on in the same way as their peers, how to alter their behaviour so that it conforms with the behaviour of others. This further requires that when one individual in the community sanctions another member of the community both of them know just which actions are being sanctioned. If Jones sanctions Smith's use of the word 'plus', then Smith can only change his behaviour appropriately if Jones's behaviour expresses his disapproval of Smith's use of the word 'plus', as opposed to Smith's hairstyle or shoes. And not only must Jones express this disapproval, but Smith must come to *believe* that Jones disapproves specifically of his use of the word 'plus' and thereby alter the way that he uses it in the future. If these beliefs and contentful attitudes did not determine the way in which dispositions changed, then it is difficult to see how they would ever coalesce.

CONCLUSION

None of the extant reductionist solutions to the sceptical problem is adequate to the task. Insofar as we can give a principled account of what determines content, that account seems to conflict with what we ordinarily take ourselves to mean by our words. For instance, Teleosemantics makes 'orange' refer to vitamin C, and communal dispositionalism makes 'horse' refer to some deceptively horsey-looking cows. The alternative to non-standard reference is, of course, that it is indeterminate what we mean since there really is no principled way to determine what we mean.

It is worth noting at this point, however, that none of the arguments against the semantic naturalists have required that we invoke Moore's Open Question Argument or anything of that form. Indeed, although all the naturalist proposals considered fail to solve the sceptical problem, they do not seem to fail because they cannot account for categorical semantic prescriptions. This should already teach us to be somewhat suspicious of the thesis that meaning is normative and that it is because meaning is normative that naturalistic theories are destined to fail.

This page intentionally left blank

6

Anti-Reductionism

Anti-reductionists maintain that intentional, semantic, or normative facts are irreducible to natural facts, and therefore must be assumed as primitive. Because Kripke discusses anti-reductionist theories only briefly, and even then, focuses mainly on the implausible *quale* theory, the sceptical argument against anti-reductionism might appear to be its weakest point.[1] Nevertheless, as I argued in Chapter 2, the argument against such theories could be made more powerful. In this chapter, I show that none of the more sophisticated non-naturalist responses to the sceptic is adequate. Each of the putative non-naturalist solutions either evades the sceptical problem, and thus fails to solve it, or begs the sceptical question. Once again, it is notable that the failure of anti-reductionist responses to the sceptical problem have nothing to do with the normativity of meaning or content.

THE CAPACITY VIEW

McGinn argues that Kripke's objections to the *quale*-theory do not generalise to all non-representational theories. There is, he claims, 'a legitimate notion of "introspection" which does not involve what Kripke describes as "attending to the qualitative character of our own experiences"'.[2] Rather, McGinn suggests that we most naturally think of meaning and understanding as *capacities*, and that these capacities are not reducible to a speaker's behaviour or physical state. McGinn explains that 'to possess the concept of *red* or *square* is to have the capacity to discriminate, recognise or identify red or square things and to classify them accordingly'.[3] The capacity could be manifested in various ways—it could be that the speaker calls red objects by the same name ('red'), or groups red objects together when asked to do so. This answers the sceptic for the following reason:

---

[1] Cf. Boghossian 1989, pp. 540–1 and McGinn 1984, p. 161.

[2] McGinn 1984, p. 161.

[3] Ibid., p. 169.

> [W]e can say that a thought has a content involving the concept *red* and not *gred* in virtue of the fact that the thought involves the exercise of a capacity to recognise *red* things and not *gred* things. ... The capacity to discriminate red things is a *different* capacity from the capacity to discriminate objects belonging to some nonstandardly defined class (the extension of the sceptic's alternative concept), and so we can find a basis for the distinction between possessing the concept *red* and possessing some nonstandard concept compatible with all the applications made hitherto ... to mean addition by '+' is to *associate* with '+' the capacity to add, i.e. to exercise *that* capacity in response to questions involving '+'; to mean red by 'red' is to exercise the capacity to discriminate red things when asked to judge whether 'red' applies to a presented object.[4]

Although McGinn treats capacities as irreducibly mental, his proposal is remarkably similar to dispositionalism. Indeed, 'capacity' and 'disposition' are terms that can often be used interchangeably. However, McGinn argues that the capacity view differs from dispositionalism because capacities, as opposed to dispositions, are irreducible to facts about a speaker's actual or counterfactual behaviour. This allegedly improves the ability of the capacity view to accommodate error: while I can have the capacity to classify red things, I may fail to do so in a given circumstance, because, for instance, my vision is impaired in some way. Similarly, I have the capacity to type, but I sometimes make mistakes: 'it is not a necessary condition of possessing an ability that one *always* exercise it correctly—other factors can interfere in such a way as to falsify the counterfactuals that purport to capture what it is to have a capacity.'[5]

Given that McGinn takes capacities to be irreducible to actual or counterfactual behaviour, his view differs from dispositionalism as Kripke characterises it. However, a realist dispositionalism, such as that put forward by Martin and Heil, similarly treats *dispositions* as irreducible to facts about behaviour. Unfortunately, this means that the capacity view will suffer similar difficulties to those suffered by realism about dispositionality. As I argued in the previous chapter, treating dispositions as real and irreducible to facts about behaviour does not help distinguish between someone's meaning *addition* by 'plus' and making a *mistake*, as opposed to meaning something non-standard such as *quaddition* by 'plus'. I might have several (real) dispositions, including the disposition to add, in some conditions, and the disposition to make mistakes in others. The dispositionalist needs to supply a non-circular distinction between those dispositions that constitute meaning and those that produce errors. Treating dispositions as real does not supply the relevant distinction.

---

[4] McGinn 1984, pp. 169–70

[5] Ibid., p. 173.

Capacities do not fare any better. Just as I am both disposed to add in some circumstances and to err in others, I am capable of adding numbers on some occasions, and I am also capable of quadding. That is, given my dispositions, the capacity theorist will want to say that I am capable of adding, although under certain conditions, I am disposed to err. Given precisely the same dispositional make up, however, the sceptic can just as easily argue that I am capable of quadding, although under certain conditions, I am disposed to err. The question is, which of these capacities constitutes what I mean by 'plus'? Just as the dispositionalist needs to distinguish between meaning-constituting and error-producing dispositions, the capacity theorist must specify which of my dispositions amount to a capacity, and which dispositions interfere with the exercise of that capacity. The capacity theorist must also distinguish between error-producing and meaning-constituting dispositions, in order to say that the latter, but not the former, constitute a capacity. Given all the facts about how I use the word 'plus', therefore, it is indeterminate whether I have the capacity to add, and therefore mean *addition* by 'plus' or have the capacity to quad, and therefore mean *quaddition* by 'plus'.

McGinn's suggestion, which is that we treat capacities as irreducible to facts about behaviour, is fundamentally like the realist about dispositionality who treats dispositions as irreducible to facts about behaviour. But as we have already seen in the previous chapter, even if dispositions are not reducible to facts about behaviour, the problem of distinguishing, without circularity, the dispositions that constitute meaning from those that do not remains. It is not as though the disposition or capacity to add is real and irreducible to facts about behaviour whereas the disposition or capacity to err is unreal or is merely a set of facts about behaviour. Especially if someone systematically errs, the sceptic seems to be entitled to say that the so-called 'error' behaviour is the manifestation of a real capacity, which determines what the speaker means.

INTENSIONALISM

According to Katz, the problem with suggestions like McGinn's is that they only appeal to psychological facts. The sceptical problem can be solved, Katz argues, if we can *also* appeal to objective facts about abstract semantic objects, namely, expression types and their senses. Katz's leading idea is that the semantic theory for a natural language correlates structures of expression types with relations between senses. Expression types are individuated by syntactic criteria, and like senses, they are abstract objects (utterances, as token expressions, are concrete). By examining the structure of expression types in a language, Katz argues, we can arrive at a theory of what each English expression *means*, that is, what sense it is correlated with in the English language. The structure of relations between expression types maps onto structures of relations between senses, and so determines which expression types express which senses.

On a given occasion, however, an expression may not be used to express the sense with which it is normally correlated. For example, 'cat' has the lexical sense 'feline animal', but if I say, 'Cary Grant played a cat burglar', I do not use 'cat' to express its literal sense; a cat burglar is not a feline animal who burgles. An expression is used to express the sense with which it is correlated only when it is used *literally*. Katz defines literal sense (RLS) and literal reference (RLR), respectively, as follows:

> (RLS) The senses of the constituent tokens within a linguistic token *T* must be the senses of their expression types, but the sense of *T* may contain information not in the sense of the linguistic type if the information only makes the sense of *T* more specific.[6]

> (RLR) The reference of linguistic token *T* is literal just in case its sense is literal and the object(s) to which *T* refers are in the extension of its sense.[7]

In addition to objective facts about utterances, expression types, and correlations between expression types and senses, Katz appeals to psychological facts about a speaker to explain what *she* means by a given *utterance*. In particular, Katz appeals to the Gricean idea that 'the speaker uses [a term] *T* with the intention of causing the members of the audience to understand *T* to mean *m* in virtue of their recognition of his or her intention that they do so on the basis of their knowledge of the common language and common pragmatic principles.'[8] This is meant to account for 'the way that speakers exploit their knowledge of the norms of the language to use linguistic tokens with a particular sense and reference.'[9] The fact that, for instance, I mean *addition* by 'plus', then, is a composite fact—involving facts about what 'plus' means in

---

[6] Katz 1990, p. 145.

[7] Ibid., p. 146.

[8] Ibid., p. 148. Cf. Grice 1989.

[9] Ibid., p. 147.

English, and psychological facts about my intentions. In his response to Kripke's sceptic, Katz uses the example of 'table' and the contrast between 'table' meaning *table* and meaning *tabair*: the contrast between *table* and *tabair* is presented as follows:

(C) a piece of furniture consisting of a flat surface, to serve as the locus of activities in the use of the artefact, and supports for holding the surface in a position for it to function as a locus for those activities.

(C′) something that is a table not found at the base of the Eiffel Tower, or a chair found there.[10]

The sceptical challenge is, then, to account for someone's meaning *table* by 'table' rather than *tabair*. Katz answers the sceptic as follows: 'We can say that a speaker's literal use of "table" means *table*, rather than, say, *tabair*, in virtue of the fact that "table" means (C) in English, not (C′), and that the speaker has the communicative intention to use the utterance of "table" literally in the sense of RLS.'[11] The fact that 'table' means *table* in English is a fact about the objective relations between the expression types including the expression type 'table', and the correlation between structures of related expression types and structures of senses. The fact that S has the communicative intention to use the utterance of 'table' literally 'in the sense of RLS' is the fact that she intends to cause the members of the audience to understand 'table' to mean *table* in virtue of their recognition of her intention that they do so on the basis of their knowledge of the common language and common pragmatic principles.

The appeal to abstract objects is promising, but it does not ultimately help answer the sceptic. Katz maintains that it is a fact that 'table' means *table* in English, that 'table' is correlated with the sense of (C) rather than with the sense of (C′). Since expression types and senses are both autonomous, abstract objects, these correlations hold timelessly. When I utter an expression of the type 'table', this has the sense of (C), because 'table' means *table* in a timeless English language. It is easy to see how facts about the English language, being autonomous and timeless, will come in handy in response to the sceptic. If it is a fact that 'table' means *table* in English, then the correctness conditions of 'table' are always there to be intuited.

Unfortunately, the situation is more complicated, and Katz's suggestion starts to lose some of its appeal. Because expression types and senses are abstract objects, Katz's view has the bizarre implication that the

---

[10] Ibid., p. 164.

[11] The following is a paraphrase of Katz 1990, p. 166.

English language has never changed.[12] Since this is patently false, Katz distinguishes between the abstract language and the spoken language. There are many variants of the abstract English language, and in each variant, different expression types are correlated with different senses. Katz maintains that when the language that we speak changes, we start to speak a new abstract language, another variant of one of the many abstract English languages. Given this twist, it is no longer a straightforward matter that 'table' means *table* in the English that we speak—the expression type 'table' correlates with the sense *table* in some variants of English, but not others. We need to find out which variant of the English language *we speak*. To discover the appropriate facts, Katz argues:

> we look at the sense of English expressions in which 'table' appears. For example, the compositional meaning of 'gaming table', 'dining table', 'operating table', 'drafting table', 'drawing table', etc. in each case involves the sense of the modifier's specifying the activities for which the surface serves as a locus. ... Or 'Nothing designed to serve as a seat and serving solely and exclusively as a seat is a table' is analytic. If 'table' had the meaning (C′), then a chair that served solely and exclusively as a seat and was located at the base of the Eiffel Tower would be something to which 'table' properly applies, and hence, *per impossible*, it would be a table.[13]

Katz suggests that what the expression type 'table' means is a function of relations between 'table' and other composite expression types, such as 'gaming table', 'operating table', and 'drawing table'. Given information about these relations, we can simply correlate expression types with senses. However, there is a difficulty. We do not want to know what 'table' means in *any* variant of the English language; we want to know what 'table' means *in our language*, in the language we *speak*. And if we want to know which expression types are correlated with which expressions in the language we speak, we need to start with *utterances*, not with expression types. Unfortunately, if we start with utterances, we face a difficulty: expressions can be used both literally and non-literally; hence, token expressions are not always used to express the same sense. For instance, 'Tim was under the table' can have a literal meaning, namely, that Tim was under a flat-topped piece of furniture, or a non-literal meaning, namely, that Tim was intoxicated. Moreover, 'data table' is not, like 'drawing table', a specification of the activities for which a flat-topped piece of furniture might be used. If we start with utterances, there will be multiple correlations to senses to choose between. Gerrymandering these correlations will require that we be able to distinguish, non-arbitrarily, that some uses of ‘table’ are literal, whereas others are not. Unfortunately, Katz’s definition of literal reference does not help because it makes circular reference to the sense of expressions in the public language. According to RLS, an expression is used literally if it is used to express the sense it expresses in the English language. Hence, any distinction between literal and non-literal uses will presuppose what ‘table’ means, that is, what sense ‘table’ is correlated with in the language that we speak. This is viciously circular.

---

[12] Katz 1990, p. 50.

[13] Ibid., p. 166.

It turns out, therefore, that objective facts about English are useless in answering the sceptic. Katz’s thought in appealing to senses is that they will provide determinate correctness conditions for a speaker. However, given that there are numerous languages, in which different senses are correlated with different expression types, we cannot appeal to senses without knowing *which* language we speak, that is, which senses are correlated with which expressions. Moreover, the sceptic wants to know, more specifically, what *I* mean by an expression. Facts about the English language can only help us to answer the sceptic if it is assumed that I speak English, rather than French, Telugu, or Quenglish.

On Katz’s view, we should be able to look at my communicative intentions to find out whether I mean *table* rather than *tabair* by ‘table’. Katz says that I mean *table* by ‘table’, on some occasion of utterance, if I have the communicative intention to use ‘table’ literally, in the sense of RLS and of causing my audience to understand ‘table’ to mean *table* in virtue of their recognition of my intention that they do so on the basis of their knowledge of the common language and common pragmatic principles.[14] This response to the sceptic is either circular or question-begging.

First, it is circular because my meaning *table* by ‘table’ involves my having the intention to use ‘table’ literally in the sense of RLS. This is merely to say that I have the intention to use it to express the sense of the expression type ‘table’, that is, the sense with which ‘table’ is correlated in English. We were hoping, however, that my communicative intentions would determine which language I speak; we cannot simply assume that I speak a particular variant of English—for, which variant should we choose? Hence, my communicative intention to use ‘table’ literally does not determine what I mean independently of some assumption of what ‘table’ literally means in the language that I speak.

---

[14] Ibid., pp. 147–8.

The response is question-begging because it appeals to my intention to cause my audience to understand 'table' to mean *table*. The sceptic asked for an account of what it is for someone to understand the meaning of an expression; what it is for me to understand 'table' to mean *table* rather than *tabair*. But how can I intend to cause my audience to understand 'table' as meaning *table*, unless I understand what it is for 'table' to mean *table* in the first place? In order to form a communicative intention of the appropriate kind, I must already grasp the concept *table*. Thus, at bottom, Katz's response to the sceptic is that I mean *table* by 'table' because I grasp the concept *table* rather than the concept *tabair*, and I use 'table' to mean *table* on the occasion in question. Unfortunately, Katz does not tell us what it is to grasp the concept *table*; he merely assumes that relation as primitive. However, if grasp of the concept *table* is meant to be a primitive relation, then Katz begs the question against the sceptic. What it is to understand (grasp) the concept *table*—to mean *table* by 'table'—is precisely what the sceptic wanted us to explain.[15]

EXTENSION-DETERMINING INTENTIONS

The sceptic divides non-reductive theories into three broad types—those that assume non-representational psychological states, those that assume abstract, platonic objects, and those that assume representational psychological states. I have argued, so far, against more sophisticated theories of the first two types. Kripke maintains that the third type of theory—which appeals to representational psychological states—either yields a vicious regress or begs the question. Because the sceptic raises the general question of what constitutes understanding, it may be difficult to see how the regress could be avoided. Nevertheless, the most popular non-naturalist response to the sceptic is to deny the force of the regress argument. In this, and the following sections, I consider a variety of proposals that purport to show that we can legitimately appeal to representational states, and thus avoid the looming regress. In this section, I consider Wright's proposal that intentions can be taken as primitive.

Wright argues that we can legitimately appeal to representations in answer to the sceptic and thereby answer the sceptical challenge:

---

[15] For further criticism of Katz around these issues, see Boghossian 1994; Zemach 1994; Katz 1994.

> In order, then, to rebut the sceptical argument, it would have sufficed, at the point where the sceptic challenged you to adduce some recalled mental fact in order to discount the grue-interpretations, to recall precisely your former intention with respect to the use of 'green'. … [I]f you are granted the intuitive notion of intention, you can reply that you do not in any case know of the content of an intention via a specification of it; rather, to repeat, you recognise the adequacy of the specification because you know of the content of the intention.[16]

Wright's suggestion—that we can answer the sceptic by simply recollecting what we formerly meant—is an upshot of the thought that the sceptical challenge is at bottom *reductionist*, that it presupposes that whatever it is that makes it the case that I mean *addition* by 'plus', it must be something other than a representational or intentional state. Wright is suggesting that this reductionist bias in the formulation of the sceptical challenge begs the question against the anti-reductionist. However, the sceptical argument is clearly meant to apply to all representational states, including intentions. This is not a product of an unjustified reductionist bias, but of the fundamental nature of the sceptical challenge. Hence, unless we are able to give an account of what makes it the case that I mean addition by 'plus' we cannot simply respond to the sceptic by appeal to the intention to use 'plus' to mean addition. Whether we appeal to 'natural' or 'non-natural' facts, some answer to the sceptic's metaphysical challenge must be given before we can be entitled to appeal to intentions.

Nevertheless, Wright has a further objection to the way the sceptical argument is formulated. He suggests that Kripke insists that my knowledge of my own intention or my knowledge of meaning is inductive. But Wright maintains that this assumption is mistaken. True, if I must rely on correctly recollecting a prior intention with regard to the use of 'plus' my memory might fail. But my knowledge that I now mean *addition* by 'plus' need not require me to recollect a prior intention, since my current intention with respect to the use of the word would suffice. This is because, Wright argues, knowledge of my current intentions is not inductive, but rather, is constitutive of what my intentions are. Roughly speaking, the thought is that knowledge of my own intentions does not admit of any gap (under suitable conditions) between what the content of my intention is and what I judge its content to be. Thus, introspective knowledge of my intentions is not to be understood by analogy with perceptual knowledge of shape. My judgements about shape can be mistaken if they fail to track the properties that are independently there. In contrast, my judgements about my own intentions cannot be mistaken (under suitable conditions), because what I judge my intentions to be *constitutes* what they are. We do not 'look inside' and check our own inner states when we avow our own intentions. More precisely, according to Wright, if John is in certain conditions, C, then it will be *a priori* that his beliefs about his intentions covary with his intentions. That is:

---

[16] Wright 1984, pp. 776–7.

$$C\ (\text{John}) \rightarrow (\text{John believes he intends to } \varphi \leftrightarrow \text{John intends to } \varphi)$$

The C conditions are those in which John is not self-deceived, is thinking clearly, sincere and so forth. These conditions must be substantively specified; it must not turn out to be trivially true that John's judgements will covary with his intentions, simply because the circumstances are specified as those in which he is always right. If we can substantively specify the ideal conditions, Wright maintains, it will turn out to be *a priori* that John's beliefs about his intentions covary with his intentions. John is not merely a good detector of his intentions—his beliefs constitute them. Thus, so long as John is in the appropriate conditions, he will believe that he intends to ϕ if and only if he intends to ϕ.

If my best judgements about my intentions constitute their content, then there can be no sense in asking whether my intentions really are what I judge them to be. How might this suggestion help answer the sceptic? Wright could argue as follows. In answer to the question, 'What constitutes my meaning *green* by "green"?', he could cite my current intention to apply 'green' to something if and only if it is green. This, in effect, is to cite my intention to conform to a rule—the rule for the correct application of 'green'. But the sceptic would no doubt question what constitutes my understanding of this rule. That is, she would ask, 'What makes it true that I intended to call all and only green things "green," and not to call green or blue things "green"?' Wright could argue that the regress stops here: my best opinion about my intention *constitutes* my having a given intention; if I judge that I intended to call all and only green things 'green', then it follows *a priori* that this is what I intended to do.

Wright's suggestion is interesting in its own right, but it does not answer the sceptic satisfactorily. The problem, quite simply, is that Wright appeals to my *judgements*, and the sceptic can always question the contents of those judgements. The sceptic can accept that John's intention to ϕ is constituted by his judgement that he intends to ϕ, she can still ask what makes it true that John judges *that he intends to* ϕ. The sceptic cast a general suspicion over the possibility that we might give an account of any intentional content, including beliefs about our own intentions. Hence, even if Wright is right in thinking that John's intentions are constituted by his best judgements, then his account of what constitutes intentions appeals to the contents of his best judgements, independently constituted. Thus, Wright's account presupposes precisely what it seeks to explain.[17]

It may be that Wright supposes that the anti-reductionist does not need to answer the sceptic, since the sceptic asks for a reduction of semantic statements to statements in non-semantic terms. In this case, the discussion of intention here could be taken as an account of how we can purport to have *knowledge* of what we mean. So interpreted, the account of intentions does not even purport to answer the sceptic who asks what makes it the case that I mean *addition* by 'plus', since that sceptic should have been seen off by the observation that she begs the question against anti-reductionism. If we take it as read that there is a fact of the matter what I mean, then this account does give an interesting story about what *kind* of knowledge I might have of my own intentions. The trouble is that we cannot go directly from this claim about knowledge of intentions back to an adequate account of what constitutes content, since the account of knowledge of intentions presupposes judgements with determinate content. Thus, insofar as Wright's answer to the constitutive question is unsatisfactory, his answer to the question of how we know our own intentions is either question-begging or irrelevant. And if the sceptic is entitled to reject appeal to intentions on the grounds that the sceptical challenge applies to mental representations quite generally, then what we are left with is an answer to the question how we *know* what we mean, without an answer to the question what facts constitute meaning. Since the sceptic will argue that there is no fact of the matter what I mean or what I intend, she will equally complain that the concept of knowledge is misplaced in Wright's discussion, or must be appropriately reinterpreted. At any rate, Wright does not seem to adequately respond to Kripke's argument.

IMPLICIT RULE-FOLLOWING

Brandom maintains that in order to answer the sceptic, we need to assume *normativity* as primitive. Brandom does not treat *semantic* norms as primitive, however; he does not assume that 'green' means *green*.

---

[17] See also Boghossian 1989, pp. 544–7.

Rather, he assumes what he calls ‘normative practices’ as fundamental. He then explains how linguistic and conceptual norms—correctness conditions—can emerge out of a normative practice. In particular, what Brandom calls ‘material inferences’ establish the content of expressions and beliefs.[18] For example, a material inference constitutive of the meaning of ‘red’ might be expressed as follows: someone who says that *x* is red all over thereby undertakes a commitment to the claim that *x* is coloured, *x* is extended, *x* is not blue all over, and so on.

Although Brandom assumes normative practices as primitive, it is not clear exactly how his view differs from a straightforwardly naturalist one. It might seem more charitable to read Brandom as offering a kind of sceptical solution to the sceptical problem—that is a story about the conditions under which we are entitled to *say*, for instance, that ‘green’ means *green* or that Jones means *addition* by ‘plus’. However, Brandom frequently suggests that he seeks to explain how content is constituted from natural elements. For example, Brandom says:

> The natural world does not come with commitments and entitlements in it; they are products of human activity. In particular, they are creatures of the *attitudes* of taking, treating, or responding to someone in practice *as* committed or entitled (for instance, to various further performances).[19]

Moreover, Brandom suggests that *attitudes* are to be understood as behavioural dispositions. This is evident in Brandom’s use of a thought experiment in his explanation of how content gets constituted.[20] This involves imagining a pre-conceptual community, with a normative practice, but no concepts. This story is not meant to be answerable to the facts of human evolution, but is designed to pick out the features that a practice needs to have in order for its practitioners to be talking. The intended result is an account of what constitutes the fact that someone understands the meaning of a word. The starting point is supposed to be a proto-hominid community in which there are norms, but no concepts or contents—neither propositional attitudes, nor explicit thoughts.

---

[18] There are, of course, other important rules of combination and syntax that allow for the complex array of expressions and uses in a natural language. However, material inferences are the most important starting point, according to Brandom, for fixing the semantic content of the kinds of expressions with which we are concerned.

[19] Brandom 1994*a*, p. xiv.

[20] Brandom (1994*a*) does not invent the thought experiment himself, but builds on the one introduced by Haugeland 1982.

Although Brandom uses normative vocabulary to say that the proto-hominids treat each other's performances as 'correct' or 'incorrect', he suggests that they do so only by *behaving* in certain ways:

> The story to be told here assumes only that suitable social creatures can learn to distinguish in their practice between performances that are treated as correct by their fellows (itself a responsive discrimination) and those that are not. ... [I]t should be clear at each stage in the account that the abilities attributed to linguistic practitioners are not magical, mysterious or extraordinary. They are compounded out of reliable dispositions to respond differentially to linguistic and non-linguistic stimuli.[21]

Responsive discrimination is just a reliable disposition to respond differentially to various stimuli. So put, it appears as though Brandom is offering a communal dispositionalist account of the determination of correctness; after all, the starting point includes nothing more than behavioural dispositions. Moreover, when we come to his positive account of the structure of the social practices necessary for conceptual content, nothing is added that would distinguish the account from dispositionalism. Brandom makes liberal use of normative vocabulary in describing the practices of the proto-hominids, but the practices he describes presuppose only that the proto-hominids have dispositions to behave in certain sorts of ways in certain sorts of circumstances. It seems that even for Brandom, sanctioning—whether it is beating with sticks or social exclusion—*constitutes* or *amounts to* the attitude of taking someone as committed, for instance, to saying '*x* is coloured' or entitled, for instance, to saying '*x* is not blue all over'. He says:

> The notion of a normative status, and of the significance of performances that alter normative status, is in turn to be understood in terms of the practical deontic *attitude* of taking or treating someone as committed or entitled. This is in the first instance *attributing* a commitment or entitlement. Adopting this practical attitude can be explained, to begin with, as consisting in the disposition or willingness to impose *sanctions*. ... What counts as punishment may ... be specifiable in nonnormative terms, such as causing pain or otherwise negatively enforcing the punished behaviour.[22]

The structure of Brandom's explanatory strategy is evident here. Normative *status* is explained in terms of *attributions* of status, and attributions are in turn explained in terms of practical *attitudes* to take or treat something as having a certain status. Finally, these attitudes are explained in terms of *dispositions* to *sanction*, to respond in certain ways to certain behaviour. The participants of a communal practice treat one another as committed or entitled, *by* imposing sanctions—by 'causing pain or otherwise negatively reinforcing the punished behaviour'.[23] Hence, treating someone as having a certain status requires only an ability to discriminate features of the environment by sanctioning differentially. Moreover, these dispositions are sufficient for the adoption of attitudes: the disposition to impose sanctions *amounts to* the 'practical deontic attitude of taking or treating someone as committed or entitled'.[24] To adopt a certain attitude is to take a stance, to attribute a normative status. Despite Brandom's use of conspicuously normative language to describe the proto-hominid practices, his picture is still largely, or perhaps even entirely, founded on dispositions: normative statuses are always derivative of normative attitudes, and the latter are explained in terms of responsive discrimination and propensities to sanction. Since these practices institute conceptual content, the ingredients are assembled for the explanation of conceptual content. The question is, how could this possibly avoid the objections against dispositionalism?

---

[21] Brandom 1994*a*, pp. 155–6.

[22] Ibid., p. 166.

Indeed, this account cannot avoid the objections against dispositionalism. The normative vocabulary Brandom uses affords no purchase on the problem. This is because Brandom cannot assume that the pre-conceptual creatures are able to *think*; by hypothesis, they are supposed to have only *pre-conceptual* abilities. Moreover, if it were necessary that the pre-conceptual creatures could *think*, Brandom would face the problem of regress. Brandom's central aim is to explain how it is that the rules that determine conceptual content could be *implicit* in practices. The account of how a practice must be in order for it to institute conceptual content cannot presuppose that the participants of the practice can have explicit, contentful thoughts.

Without assuming the ability to think, however, Brandom's response to the sceptic suffers the same difficulties as dispositionalism. Consider face-to-face interaction between two members of the kind of community described by Brandom; call them 'Jim' and 'Pam'. Jim says to Pam, pointing, 'that's red'. We are supposed to imagine that Jim makes these sounds and gestures, and Pam, taking all of this in, attributes certain commitments and entitlements to Jim. This just means that Pam becomes disposed to sanction Jim—disposed, that is, to punish Jim under some circumstances but not under others. Imagine, further, that at some later time poor Jim *is* punished. The question is, *what* has Jim been punished for? Has Pam attributed the commitment to say 'that's not blue', or has she attributed the commitment to say 'that's not grue'? Which of these commitments has Jim violated?

---

[23] Brandom 1994*a*, p. 166.

[24] Ibid.

When Pam attributes a commitment to Jim, Pam cannot *think* to herself *Jim has now committed himself to saying that that is not blue*. The reason is, simply, that for Pam to have *that* thought in mind, she must be able to think, and thinking contentful thoughts is not one of her abilities. Pam's abilities, as they have been described, are not sufficient for Pam to attribute a *determinate* commitment or entitlement to Jim. Similarly, Jim is unable to undertake any determinate commitments and entitlements, insofar as that requires that he attribute them to himself. Given Brandom's rejection of dispositionalism, he can hardly maintain that Pam's disposition to behave determines that she attributes *these* commitments to Jim rather than *those*. As Brandom puts it, 'there is no such thing as *the* regularity that is being reinforced by a certain set of responses to responses, or even dispositions to respond to responses.'[25] Since there is no such thing as *the* regularity of Pam's sanctioning behaviour, there can be no such thing as *the* commitment she has attributed.

However, if Brandom alludes to facts about the correctness of attributions, we run into problems of regress. To see this, consider the possibility that commitments and entitlements are constituted by their attribution, that is, that the members of our little community *create* deontic status. In this case, when Pam attributes a deontic status to Jim's behaviour, she is making a contentful evaluation or characterisation of what Jim does. This means that she attributes commitments and entitlements on the basis of what Jim has said or done, and she sanctions in the ways that she does *in virtue of* the entitlements and commitments she has attributed. Thus, in order for Pam to attribute a particular set of commitments and entitlements, her attribution of those deontic statuses must be subject to standards—in particular, the correctness conditions supplied by the content-determining material inferences. But how can her behaviour be subject to these rules? Whatever it is that determines the proper application of the rules, it can hardly be the regularity that Pam in fact adheres to. The only alternative seems to be to suppose that she has them in mind, that she *grasps* the requisite rules. But if this is the answer, then we can ask whether she has grasped the rules correctly, and a further rule needs to be invoked. And so on, *ad infinitum*.

---

[25] Brandom 1994*a*, p. 36.

One way Brandom might try to avoid these problems is by adverting to his claim that it is norms all the way down, and that attributions of normative status—taking some behaviour as correct or incorrect—are always prior. He sometimes suggests that *we* attribute deontic status to the behaviour of the members of another community, when we, at the *meta-level*, say that they have a normative practice.[26] That is, he suggests that the *interpreter* attributes normative status and thus conceptual content to the members of the community being *interpreted*:

> The account of deontic scorekeeping on doxastic and practical commitments explains what one must interpret a community as doing in order for it to be *talking* that one is thereby taking them to be doing ... in place of a direct explanation of what commitment and entitlement are, an account of what it is to *take* someone to have such a status was offered.[27]

The idea seems to be that the normative status of the actions of the proto-hominids is a function of *our* attribution of that status: *we*, so to speak, institute their norms by attributing to them the ability to attribute correctly or incorrectly. So one more level of explanation needs to be added to Brandom's picture. As it is, we had normative status explained in terms of attributions of normative status, attributions explained in terms of practical deontic attitudes, and these attitudes explained in terms of dispositions to sanction. Now, we need to explain the practical deontic attitudes in terms, not just of the natural abilities of the members of the community, but *also*, in terms of normative statuses *attributed* explicitly by us, the interpreters in this exercise.

It might help to think of this picture as similar to the Davidsonian view that we employ a principle of charity when interpreting others.[28] For both Davidson and Brandom, we apply *our* standards to others when we interpret their behaviour as, for instance, linguistic. Indeed, this kind of picture makes sense when we imagine an anthropologist or a radical interpreter struggling to understand the doings of a far-flung community.[29] Clearly *some* standards need to be employed by the interpreter, because the behaviour of speakers to be interpreted will radically underdetermine what they mean. So why not invoke *our* standards?

---

[26] These points touch on the issue of Brandom's 'phenomenalism' about norms, which has been criticized by Rosen (1997). Rosen argues that there is no sensible way to understand the claim that attributing a normative status (taking something to be correct) could be prior and institutive of its having that normative status. Of course, this sort of self-reflexive institution does make sense for 'social kinds' such as money. However, in accounts of the social constitution of money, psychological attitudes provide a resource for analysis. In this case, it is the content of psychological attitudes itself that is at stake.

[27] Brandom 1994*a*, p. 637.

[28] Davidson 1984, Essay 2.

[29] For a discussion of the 'anthropological' point, see Rosenberg 1997.

While some level of ethnocentrism might make cross-cultural interpretation possible, it hardly makes sense of the case we were considering—that of the pre-conceptual proto-hominids. I take it that we were not meant to imagine the proto-hominids to be an as yet undiscovered group of primates living in a distant and isolated land. The proto-hominids were a fiction, a model designed to show what it is that a community must be like in order for it to develop conceptual contents. If our attribution of status is necessary to discriminate *which* commitments and entitlements are being undertaken or attributed by the fictional proto-hominids, then it will turn out that *explicit* attribution of normative status (by us) is necessary for the *implicit* practice to be one capable of mere normative sanction. But this has Brandom's order of explanation back to front. Indeed, we may need to use our explicit vocabulary to talk about the proto-hominids, but that talk cannot be a *necessary* condition for practices to institute norms *implicitly*. This conflicts with Brandom's manifest commitment to the idea that an implicitly normative practice is prior to one in which deontic status can be made explicit:

> I am indeed committed to the possibility of norms implicit in prelinguistic (and so nonconceptual) practices. Such implicitly normative practices are prior in the order of explanation. … The picture is that what proto-hominids could do before they could talk is to take or treat each other's performances as correct or incorrect by practically sanctioning them, e.g. by beating each other with sticks as punishment.[30]

Given the sceptical problem, there is no sense to be made of the idea that (1) the proto-hominids are pre-linguistic and pre-conceptual, and that (2) they can attribute deontic status—that is, take or treat each other's performances as correct or incorrect—merely by sanctioning. Without invoking the ability to talk and think, the proto-hominids cannot be said to confer deontic status at all when they practically sanction one another's behaviour. For their treatment of one another to amount to attribution of specific deontic statuses, they must be able to entertain concepts; if their practice does not amount to the attribution of specific norms, then the deontic vocabulary turns out to be another way of describing behavioural dispositions. Thus, given the way that the dilemma was initially set up by Brandom, (1) and (2) above just do not hang together. The upshot seems to be that if Brandom sticks to the bare bones of dispositions, punishment, responsive discrimination, and behaviour—regardless of the vocabulary he uses—he has insufficient resources to evade the objections against dispositionalism. Alternatively, if he tries to bring more into the picture—if he expects status attributions to determine *specific* deontic statuses—he involves himself in a regress.[31]

---

[30] Brandom 1997, p. 201.

BEDROCK NORMS

John McDowell responds to Kripke's sceptic by arguing, first of all, that the sceptical argument founders on a mistaken assumption. If this assumption is rejected, according to McDowell, an anti-reductivist account of meaning and understanding remains open. The allegedly villainous assumption is what McDowell calls the 'master thesis' that we can separate a content from its 'vehicle' in the case of mental representations. He says:

> Kripke's reading of how the regress of interpretations threatens the very idea of understanding turns on this thesis: 'no matter what is in my mind at a given time, I am free in the future to interpret it in different ways'. This presupposes that whatever is in a person's mind at any time, it needs interpretation if it is to sort items outside the mind into those that are in accord with it and those that are not. There are always other possible interpretations, and a different interpretation, imposing a different sorting, may be adopted at a different time. Considered in themselves, that is, in abstraction from any interpretations, things in the mind just 'stand there'.[32]

On this assumption, we can distinguish between two elements of a mental representation: the vehicle and the content. The vehicle is thought to be a non-intentional state—such as a brain state, or a dispositional state. The vehicle is as devoid of intentional content as a stone, or an un-interpreted word. In order for the vehicle to have a content it needs to be interpreted. Without an interpretation, we cannot make sense of the idea of the vehicle applying correctly or incorrectly.

McDowell objects that the 'master thesis' is untenable. Although the sceptic tries to pass it off as a platitude, it is no such thing. On the contrary, McDowell maintains that it is *peculiar* to think of thoughts as non-intentional things plus interpretations. The master thesis mistakenly conceives of mental entities as though they were like linguistic tokens.

---

[31] For further discussion of these points, see Hattiangadi 2003.

[32] McDowell 1998, p. 226. McDowell quotes from Kripke 1982, p. 107.

Bits of written text might require interpretation to be meaningful, but mental entities, on the common sense account, are different; they are already meaningful and require no further interpretation:

> The master thesis is not just a piece of common sense, which we can sensibly leave unquestioned while we look in Wittgenstein for philosophical contrivances aimed at freeing it of paradoxical implications. … The master thesis implies that what a person has in mind, strictly speaking, is never, say, *that people are talking about her in the next room*, but at most something that *can* be interpreted as having that content, although it need not. Once we realize that that is what the master thesis implies, it should stand revealed as quite counter-intuitive, not something on which a supposed need for constructive philosophy could be convincingly based.[33]

According to McDowell, a regress follows only if we assume that a mental item 'stands there' in need of an interpretation if it is to acquire a correctness condition. But the master thesis is untenable; the relationship between the mental entity—thinking *that people are talking about me in the other room*—stands in no need of interpretation. According to McDowell, the right answer, in response to the looming regress, is to decline the temptation to assume that mental items are mere vehicles for interpretations: 'We should not suppose that the normative surroundings of the concept of understanding can be in place only thanks to there being a role for the concept of interpretation.'[34] Rather, if we give up the master thesis we can embrace,

> … the conception of meaning as reaching normatively into the objective world: for instance the idea that the meaning of, say, an instruction for extending a numerical series determines what is correct at any point in the series, in advance of anyone's working out the series to that point, so that meaning constitutes a standard of correctness for what any calculator or group of calculators does or might do.[35]

That is, McDowell suggests that if we give up the master thesis, we can assume full-blown representational states that have all of the hallmarks of the truth conditional picture. Thus, he says:

> If the threat posed by the regress is properly disarmed by discarding the master thesis … then ratification-independence … can fall into place as simply part of a way of thinking that we are now able to take in our stride. There seemed to be problems about the normative reach of meaning, but since they depended on a thesis that we have no reason to accept, they stand revealed as illusory.[36]

---

[33] McDowell 1998, p. 272.

[34] Ibid., p. 273.

[35] Ibid., pp. 273–4.

[36] Ibid., p. 274.

No doubt, McDowell is right to question the assumption that we can separate vehicles from contents of mental representations. However, McDowell is mistaken in resting the full force of the regress argument on the master thesis. On the contrary, the sceptic can grant McDowell's claim that mental representations are individuated by their contents, and that no sense can be made of the question whether a given representation-vehicle has a given content. For, having granted that assumption, the sceptic can still go on to ask what makes it the case that a given speaker has the beliefs that we say she does rather than some other beliefs or none at all. For example, suppose I believe *that people are talking about me in the other room*. Without separating vehicle and content, the sceptic can ask what makes it the case that I have the belief *that people are talking about me in the other room* rather than the belief *that Elephants are dancing in the basement*. If that is a reasonable question to ask, then the sceptical argument can be carried without at any point relying on the master thesis.

Here is how the argument goes: first, the sceptic asks what makes it the case that I believe *that p*. One thing that the sceptic does take to be crucial to my believing this is that I am following a rule—in this case one that tells me to believe *that p* if and only if p. Now, the sceptic asks, what makes it the case that I am following this rule? In answer, we can either appeal to my dispositions, in which case the objections to dispositionalism would come into play, or we can appeal to a further mental representation, such as my intention to believe *that p* if and only if p. But now the sceptic can ask what makes it the case that I intended to believe *that p* if and only if p. And so on.

Does this version of the regress argument secretly rely on the master thesis? It could be seen to rely on something very much like the master thesis—although one not so controversial. In order to get the regress started, the sceptic has to hold something fixed and vary something else. In the above regress argument it is not vehicles that remain fixed, while contents float free. Rather, what is held fixed are persons, and it is asked what makes it the case that a person has this belief rather than that. The beliefs themselves can be seen as individuated exclusively by their contents, assuming no possibility of differentiation between contents and vehicles. And while it may seem peculiar to think of beliefs as composed of vehicles and content, it is indeed a bit of commonsense that a person can have this belief or that. Hence, the above regress argument does not seem to be sensitive to the same sort of objection as the argument McDowell outlines.

On the improved version of the regress argument, if it is reasonable to ask what makes it the case that Jones believes *that p*, then an answer that appeals to a further mental representation will give rise to the same kind of question again, and the regress will get started. McDowell could respond to this by insisting that this is not a reasonable question and therefore one on which we should remain silent. However, this is not a viable option. Because this question no longer presupposes the master thesis, the dubiousness of the master thesis cannot be used to show that the question is unreasonable. To refuse to answer a reasonable question is simply evasive. Moreover, McDowell argues that Wittgenstein should not be seen as a quietist. Rather, he attempts to give a constructive answer to the question 'how is meaning possible?' with the use of the notion of a custom or a practice.[37] By appealing on the one hand to dispositions and on the other hand to customs, Wittgenstein hopes to steer a path between a 'bald' naturalism, in which there are no norms, no correctness or incorrectness, and the regress of interpretations. McDowell says:

> The point of *PI* § 198, and part of the point of § § 201–202, is that the key to finding the indispensable middle course is the idea of a custom or practice. How can a performance be nothing but a 'blind' reaction to a situation, not an attempt to act on an interpretation (avoiding Scylla); and be a case of going by a rule (avoiding Charybdis)? The answer is: by belonging to a custom (*PI* § 198), practice (*PI* § 202), or institution (*RFM* VI-31).[38]

Unfortunately, the appeal to custom here only works to provide a constructive response to the question 'how is meaning possible?' if it is already assumed that the regress argument can be avoided. McDowell does not wish to defend a form of communal dispositionalism, according to which the correct use of an expression is determined by the dispositions of the members of the community. This would be to treat a community as a mere aggregate of individuals. And McDowell complains that if an individual's dispositions to verbal behaviour, described in non-intentional terms, are insufficient to determine meaning, then adding more individuals will hardly be of any use. Rather, on the picture that McDowell envisages 'shared membership in a linguistic community is not just a matter of matching in aspects of an exterior that we present to anyone whatever, but equips us to make our minds available to one another, by confronting one another with a different exterior from that which we present to outsiders ... a linguistic community is conceived as bound together, not by a match in mere externals (facts accessible to just anyone), but by a capacity for a meeting of minds.'[39] However, this picture can only be a viable, constructive response to the sceptic if individuals have minds to meet. The constructive part of McDowell's response to Kripke's sceptic seems to fail because it depends crucially on the destructive part, which aimed to show that the sceptical regress argument could be avoided. Rejecting the 'master thesis' does not take the sting out of the sceptical argument, since the argument can be reformulated without assuming it. Since the regress has not been avoided, McDowell has not shown how we can appeal to full blown representations without inviting regress.

---

[37] McDowell 1998, pp. 275–6.

[38] Ibid., pp. 242.

INTERNAL RELATIONS

Like McDowell, Baker and Hacker maintain that the sceptical argument is spurious because it tries to keep thoughts separate from their contents, which are internally related. Baker and Hacker also point to additional mistakes that the sceptic allegedly makes, all of which ultimately have to do with the internal relations between concepts and conceptual knowledge. I will examine each in turn, and argue that none of these assumptions is essential to the sceptical argument.

First, Baker and Hacker claim that the sceptical argument rests on an assumption about endless justification. Kripke's sceptic argued that any rule you might give to justify your use of '+' in a given circumstance could itself be subjected to sceptical scrutiny. If you say you told yourself to count beads in a certain way, the sceptic can ask whether you really meant *quount* by 'count' when you told yourself what to do. If you formerly meant *quount* by 'count', you would now have to say that 68 + 57 = 5. Baker and Hacker argue that this regress cannot go on forever:

> If I am asked what an expression means, I can explain it. An explanation of meaning is a norm of correct use. If my explanation is not understood, I can clarify it, i.e. I can give a further explanation of my explanation (a rule for the application of the rule). Ultimately, perhaps, I will explain by giving a series of examples with an 'and so on' rider. *This too is an expression of the rule*. Now my explanations will terminate at the point of showing that *this* and *this* ... is what I call 'going on the same'. If I am *now* asked 'Why?', I can only say 'This is simply what I do'. I have no *further* justification. But I have given a justification for what I do, so I cannot be accused of having made a stab in the dark.[40]

---

[39] McDowell 1998, p. 253.

[40] Baker and Hacker 1984, p. 82.

This response belies a mistaken epistemological interpretation of the sort of justification Kripke's sceptic is after. The sceptic is made out to be asking for an explanation of how I *know* what I should say. Baker and Hacker then argue that but a few reasons are sufficient; there is no need for me to keep on giving reasons. But the sceptic is not asking for an epistemic justification for my action. The question, rather, is metaphysical: what *constitutes* my meaning *addition* rather than *quaddition* by 'plus'? If I tell you that my meaning *plus* is constituted by my grasp of the relevant rule, then the sceptic asks what my grasp of the rule consists in, not how I *know* which rule I have grasped. The sceptic asks a general question about what constitutes meaning—we cannot say that what constitutes my understanding of *p* is my understanding of *q* without raising the question of what my understanding of *q* consists in. It therefore seems that abandoning the requirement for infinite explanations or epistemological justifications simply misses the sceptical point.

Baker and Hacker additionally point to a feature of rules that accounts for the fact that justifications here are unnecessary. Since this might help answer the constitutive question, as well as the epistemological one, it is worth considering. They say that the rule is related to the acts that accord with it—its extension—*internally*. An internal relation is one that determines the identity of the *relata*. Internal relations, Baker and Hacker claim, cannot and need not be justified:

> Absence of grounds is a criticism if grounds are at least possible, and if doubt about justification is reasonable. But neither of these conditions obtains here, where justifications terminate. Precisely because a rule and its extension are internally related, because this nexus is grammatical, there can be no such thing as justifying it. For there is no such thing as justifying grammatical, conceptual connections by reference to reality. ... Writing '1002, 1004, ... ' after '1000' in the course of expanding the series of even integers is internally related to the rule of the series '+2'. Writing anything else would not be following *that* rule correctly.[41]

The claim here is certainly compelling. You cannot match up the concept *green* with the extension of *grue*—*green* is the concept with all and only green things in its extension. However, the fact that a rule and its extension are internally related does not help us to solve the sceptical problem, or even to avoid it. The sceptic can grant that the relation between the *plus*-rule and its extension, and the relation between the *quus*-rule and its extension, are internal. The sceptic can then ask what makes it true that *addition* is the concept I grasp, rather than *quaddition*. Even if rules are internally related to their extensions, what relation do these internally related pairs of rule and extension bear to *me*, such that it is true that I mean *addition* rather than *quaddition* by 'plus'? Of course, it cannot be that the speaker has the whole extension in mind when she thinks of the concept because in some cases, the extension will be infinite. So, presumably she has the concept in mind. But what is it that determines whether the speaker has concept C with extension E, or the slightly different concept C^* with the slightly different extension E^*? That C and E are internally related offers no help in answering this question.[42]

---

[41] Baker and Hacker 1984, p. 83.

There are three things involved here: the conceptual rule, its extension, and the speaker. The sceptic can grant that the conceptual rule and its extension are internally related, because this is an essential assumption of the truth conditional picture of content. In order for the speaker to mean *addition*, then, she must grasp the concept of *addition* with *its* extension. However, Baker and Hacker additionally claim that the sceptic makes unwarranted assumptions about the relation between the speaker's understanding of a rule and her understanding of what accords with it. Because a rule is internally related to its extension, Kripke's assumption that any case could be *novel* is mistaken. Thus, Baker and Hacker argue:

> The rule sceptic thinks that what counts as doing the same thing is determined here by precedent. And he cannot see how to extract from precedent what 'doing the same' to a new case would be. In fact, what 'doing the same' is is determined by the rule. … One cannot grasp a rule without knowing what it determines as doing the same.[43]

Indeed, a novel case is not novel from the point of view of the rule—because the rule is internally related to its extension. Moreover, it may be that grasp of a rule implies grasp of its extension. But the speaker herself cannot have the whole extension in mind; so, novel cases can be novel for the speaker. The sceptic simultaneously questions both whether I have grasped the concept of *addition*, and whether I know what answer I ought to give. So, the sceptic can accept that one cannot grasp a rule without knowing what it determines as doing the same, but then she casts doubt on both the grasp of the rule and the knowledge of what accords. What is it that determines that I have grasped the *addition* function, and that therefore, '125' is the answer that I should give?

---

[42] Baker and Hacker frequently talk as if words are internally related to their extensions, when we naturally think that their *meanings* are internally related to extensions; or that *concepts* are internally related to their extensions. I suspect, however, that they mean 'word type' by 'word', where two words are of the same type if they have the same meaning. If, in contrast, they mean that token words are internally related to their extensions, they are mistaken. Word tokens are not internally related to extensions as they can be related to different extensions in different languages, or in different contexts within the same language.

[43] Baker and Hacker 1984, p. 87.

In answer, Baker and Hacker make the stronger claim that there is an internal relation between the knowledge of the rule and the speaker's behaviour:

> It is assumed that an individual's behaviour is merely inductive or quasi-inductive evidence for his understanding a rule-formulation (or for what he understands by it). . . . The notion that one must postulate one's own understanding on the basis of one's past behaviour rests on misconceptions about understanding and intentions that are peculiar to rule-scepticism. What these premises of rule-scepticism share is a failure to acknowledge that acting in certain ways (what is called 'acting in conformity with the rule') are *criteria* for understanding a rule, and that acting otherwise is a criterion for failing to understand it. . . . The rule-sceptic distorts this internal relation between acts and rules by treating acting in accord with a rule as making understanding the rule merely a probable hypothesis.[44]

The idea here seems to be that the internal relation between the rule and its extension maps on to the speaker. Understanding the rule necessarily implies acting in accordance with it; acting otherwise constitutes a failure of understanding. This claim, however, is too strong—I am free to act as I like in any given situation regardless of my understanding of the rule. That is, I can choose not to add if I want to be perverse, silly, or provocative. Moreover, I can make a mistake in the application of a concept that does not belie my misunderstanding of the concept. For example, I may understand the concept *red* perfectly well, and yet fail to apply it correctly on a given occasion because it is too dark, or because I do not look carefully enough. Thus, it is not invariably the case that my behaviour manifests my understanding.

Baker and Hacker might, however, be making the weaker claim that understanding of a rule implies knowledge of what constitutes acting in accordance with it. But then this internal relation is unhelpful in answer to the sceptical problem. If to understand is to grasp a rule, and to understand a rule is to know when it applies correctly, the same regress that threatened other anti-reductionists will threaten Baker and Hacker's. Obviously, I cannot have all of the correct applications before my mind when I understand a rule. So, in order to grasp a rule, I must represent it to myself. But then, we are once again appealing to representations, which is, after all, what the sceptic calls into question.

---

[44] Ibid., p. 103.

In all of these cases, the internal relations that Baker and Hacker accuse Kripke of ignoring are irrelevant to dissolving the sceptical problem; the sceptical challenge can be reformulated while paying full homage to all of the genuine internal relations between concepts, extensions, and speakers. The sceptic can accept that if I understand the rule then I know what constitutes acting in accordance with it, but then she goes on to call *both* into question. That is, the sceptic challenges us to account for what determines whether I understand concept C or C^*, and thus what accounts for whether I ought to answer a or a^*. It might be true that a accords with C and a^* accords with C^*, but which of these is the concept that I understand? And, more importantly, what is it about me that determines which concept I understand? The internal relations between my understanding of the concept and my knowledge of what constitutes accordance with it do not seem to help us to answer this question at all.

CONCLUSION

In Chapter 5, I argued that, as it stands, the sceptical challenge cannot be met by any reductive theory; in this chapter, I argued that it does not help to assume semantic, intentional, or normative facts as primitive. Though I do not purport to have exhausted all the anti-reductionist possibilities, the foregoing should suggest that the prospect of an anti-reductive solution is woefully bleak. Assuming non-representational mental states is of no use; we only encounter the very objections raised against dispositionalism. Unfortunately, we are no better off assuming representations. If we do so, we are caught in a vicious regress. None of the ingenious suggestions for avoiding this regress, such as those made by Wright, Brandom, and McDowell, have been able to stop it. Platonism, although initially promising, is unhelpful because ultimately some psychological state must be the arbiter of which concepts I understand. After all, it is a speaker's *grasp* of a given concept, as opposed to another, that concerns the sceptic.

Since naturalism and non-naturalism exhaust the possibilities for 'straight' solutions to the sceptical problem, this chapter and the previous one together strongly suggest that, as it stands, the sceptical problem resists solution. It looks as if the truth conditional picture of meaning and understanding is in serious trouble. Moreover, with the *a priori* argument against reductionism and the *a priori* presumption against anti-reductionism tucked into his belt, the sceptic can argue that, not only do extant theories fail, but that *no* theory will be capable of succeeding. If we assume the truth conditional picture, and if we allow the sceptic to assume both *Norm-Relativity* and *Normativity*, there can be no theory that accounts for what it is to understand the meaning of a word. Moreover, as I argued in Chapter 4, there seems to be no prospect of a *sceptical* solution to the sceptical problem either. So, we seem to be in a genuine bind; if there can neither be a 'straight' solution to the problem, nor a 'sceptical' one, there appears to be no such thing as meaning anything by any word, and no prospect of ever legitimating our ordinary ascriptions of meaning and content.

The question now is whether the resounding failure of the extant theories should lead us to give up hope of ever finding a solution to the sceptical argument. Indeed, with a paradox of this magnitude, we should be suspicious of the assumptions that were made in the course of the sceptical argument. In the next chapter, I will argue that it is the *a priori* part of the sceptical argument that we must reject. If we do so, then even though we may now have no adequate theory in hand, we would not be justified in concluding that none will be forthcoming.

This page intentionally left blank

7

Is Meaning Normative?

Up to now, I have been arguing that all of the extant responses to the sceptic are inadequate as they stand. The sceptical problem has not yet been solved, but we need not conclude that no solution is possible. Indeed, I think that we should not adopt the sceptical conclusion, because the *a priori* argument against all possible solutions, which was reconstructed in Chapter 3, rested on the thesis that meaning is *categorically normative*. This assumption, which is intuitive to many of the participants of the discussion, is worthy of further investigation.[1] The purpose of this chapter is to show that the thesis that meaning is normative is mistaken, and therefore, that the sceptical argument is unsound.

In the next section, I will consider, more closely, the principle called *Normativity* (introduced in Chapter 3), which is meant to capture the thesis that meaning is normative. In subsequent sections, I will consider a number of reasons why one might think that *Normativity* (or a weaker version) is true. I will argue that none of these reasons is compelling.

THE NORMATIVITY PRINCIPLE

According to the semantic realists, the meaning of a word can be given by its correctness conditions. We can specify what 'blue' means by saying that it applies correctly to all and only blue things. But this tells us nothing about understanding—for that, we need the assumption that understanding a word is akin to following a rule. This rule can either be a prescriptive rule, or not. If it is not, then meaning is merely norm-relative, but not also normative. That is, *Norm-relativity* follows. Recall that this was the thesis:

---

[1] Proponents of the thesis that meaning is normative include: Blackburn 1984; Bloor 1997; Boghossian 1989; Brandom 1994*a*; Glock 1996; Kripke 1982; McDowell 1994; Miller 1998; Pettit 1990*a*; Wright 1984; Zalabardo 1989. For further critical discussion of the normativity of meaning, see Davidson 1984; Dretske 2000; Glüer 1999*b*, 2001; Glüer and Pagin 1999; Papineau 1999; Wikforss 2001; Wilson 1994.

> *Norm-Relativity*: S means F by $t \rightarrow (a)(S$ applies x 'correctly' to $a) \leftrightarrow a$ is f

If meaning-determining rules are thought to be *prescriptive* rules then, plausibly, *Normativity* follows. Recall the *Normativity* principle:

> *Normativity*: S means F by $x \rightarrow (a)(S$ ought to (apply x to $a)) \leftrightarrow a$ is f

The speaker's semantic obligation has this biconditional content, because *Normativity* was intended to capture the semantic realist's view that the meaning of a word can be given by its correctness conditions. The result is that if I mean something by an expression, I thereby incur an obligation to use it if and only if it can be described as 'correct,' given what I mean by it. *Normativity* thus says that if someone means something by an expression she incurs an obligation under certain conditions. However, *Normativity* seems to be too strong. To see why, first break down *Normativity* into its two conditionals:

(1) S ought to (apply x to $a) \rightarrow a$ is f

(2) a is $f \rightarrow S$ ought to (apply x to a)

Whereas (1) might be doable, (2) clearly is not. In many cases, it is not up to me whether a is f; so, in order to do (2), I would have to apply x to every a that is f. That is, if I meant *green* by 'green', in order to carry out my semantic obligation, I would have to apply 'green' to all the green things. If we take into consideration all the other words I know, the demands of semantics would simply be too high. Moreover, because 'ought' implies 'can', (2) is not just too demanding, it is false.

This difficulty should already arouse our suspicions. Unlike *Normativity*, *Norm-Relativity* can be formulated in biconditional form: that is, according to *Norm-Relativity*, if I mean *horse* by 'horse', 'horse' will apply correctly (i.e. refer) to all and only horses. There is no analogous difficulty in saying that the correct application of 'horse' outstrips what I can do—'correct' or 'refers' does not imply 'can'. Thus, it seems as though *Norm-Relativity* might be a natural extension of the truth conditional picture of determinate meaning to a thesis about a speaker's understanding, whereas *Normativity* can only require that a speaker adhere to one side of the biconditional correctness condition in order to grasp a determinate meaning.

Since *Normativity* is clearly too strong, perhaps 'semantic normativitists' ought to assume a weaker principle, such as the following:

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

According to *Normativity**, a speaker who means something by an expression ought to ensure that she uses it only when it is correct to do so. This avoids the above problem, because it is something the agent can achieve. And if *Normativity** is defensible as a commitment of any reasonable, semantic realist account of understanding, it may still allow the sceptic to advance the claim that there can be no adequate theory of meaning and content—so long as it can be shown that semantic obligations are of just the right kind.

ESSENTIALLY SEMANTIC CATEGORICAL OUGHTS

*Normativity** is a thesis about what it is for a speaker to mean (or understand) something by a word—what it is for a speaker to mean *addition* by 'plus' or *green* by 'green'. Thus, Kripke says that 'a candidate for what constitutes the state of my meaning one function rather than another, by a given function sign, ought to be such that, whatever in fact I (am disposed to) do, there is a unique thing that I *should* do.'[2] Similarly, Brandom says that 'our ordinary understanding of states and acts of meaning, understanding, intending, or believing something is an understanding of them as states and acts that *commit* or *oblige* us to act and think in various ways.'[3] These passages suggest that, for the sceptic—indeed, for any 'normativist'—it is constitutive of my meaning something by an expression that I ought to use the expression in certain determinate ways. The thought is that meaning imposes specifically semantic obligations, and a speaker *must* have the relevant obligations if she is to mean something by her words.

Thus, *Normativity** should not be confused with the idea that sometimes we are obliged to use language in certain ways for *moral*, *prudential*, *rational*, *legal* or other reasons. Although we may indeed have such non-semantic obligations, there is no reason to think that a theory of meaning has to make sense of them. For example, imagine that someone were to ask you whether it is safe to skate on an apparently frozen lake, when you happen to know that the ice is thin. Of course, you ought to say 'the ice is thin', or something to that effect, but this is a *moral* obligation, not a semantic one; you must look to a moral theory to tell you what you ought to do in this situation. The fact that you mean *ice* by 'ice' and can expect your interlocutor to do the same is merely a fact that is relevant to the question of what you ought to do. Similarly, imagine that you are in Spain, you speak Spanish fluently, and you want to communicate. Well, obviously, you ought to speak Spanish. But this is clearly a *prudential* obligation. In contrast, semantic normativists hold that it is essential to my meaning *blue* by 'blue' that I ought to say that something is blue only if it is. This clearly has nothing to do with what I morally, prudentially, legally or rationally ought to do. Moreover, if *Normativity** implied only that meanings were *relevant* to our obligations, it would be trivial, since all sorts of things may be relevant to what I ought to do without themselves being, in any interesting sense 'normative facts'. For example, whether or not I ought to take a leisurely stroll depends on the weather, but that does not make facts about the weather irreducibly normative. A theory of rain would not have to accommodate the fact that I ought to carry an umbrella in the rain. Similarly, there is no reason why a theory of meaning should account for the prescriptions to which meaning is merely *relevant*. Thus, we should not confuse *Normativity** with the true but trivial claim that semantic facts, just like facts about the weather, can sometimes be relevant to what someone ought to do.[4]

---

[2] Kripke 1982, p. 24.

[3] Brandom 1997, p. 13.

Second, we need to keep in mind that, as I have formulated it, the sceptical argument relies on the thesis that meaning is normative in order to rule out all possible theories of meaning *a priori*. This argument depends on the assumption that semantic 'ought's are *categorical*. This means that semantic oughts cannot simply be *means/end* prescriptions, that is, concerning what one ought to do in order to achieve the object of a wish or a desire or to achieve an end. An example of an ordinary means/end prescription is this: 'If you want to get from Oxford to Cambridge by noon, you ought to take a morning train.' In this case, what you ought to do is really what you ought to do *in order to satisfy your desire*; if you cease to want to get from Oxford to Cambridge by noon, you no longer ought to take a morning train. A *categorical* prescription, in contrast, says what one ought to do not merely as a means to satisfying a desire or achieving an end. This distinction is not

[4] Gibbard 2003; Wikforss 2001.

about form: categorical prescriptions can have a conditional structure so long as the consequent of the conditional does not constitute a means to achieving the end expressed in the antecedent. Hence, 'If you are a moral agent, you ought to keep your promises' is a categorical prescription.

The crucial point is that semantic prescriptions must be *categorical* as opposed to merely *instrumental* if the sceptic is to use the normativity of meaning in his service. Remember, for example, that Hume reasoned that you cannot derive an 'ought' statement from a consistent set of 'is' statements because 'ought' statements have a necessary connection with motivation and the will. If I judge that I ought to give to charity, then necessarily, I am motivated to give to charity. The Humean ban on deriving an 'ought' from an 'is' does not apply to instrumental oughts. Suppose that I judge that *if* I want to get to Edinburgh in the shortest possible time, I ought to take a plane. This is a purely descriptive judgement. I can make it without having the slightest motivation to fly to Edinburgh. Even if I do have the belief, I will not be motivated to fly to Edinburgh unless I have a *desire* to get to Edinburgh in the shortest possible time. It is the desire that does the motivating, not the belief that flying is the quickest way to get to Edinburgh. In the absence of the desire, the judgement that flying is the quickest way to get to Edinburgh has no motivational force. Since hypothetical 'ought' statements do not have a necessary connection with the will, they do not introduce a 'new relation' that is not contained in descriptive statements, and so are not normative, or at least, not normative in the sense that is required for Hume's argument that you cannot derive an 'ought' from an 'is'.

In contrast, the fact that it is morally good to give to charity gives me a categorical reason to give to charity—I have a reason to give to charity even if that is not what I want to do. Moreover, assuming motivational internalism, if I am a rational agent, and if I judge that it is good to give to charity, then I am *ipso facto* motivated to give to charity. And it is this link between what a rational agent judges and her motivation that gives rise to Moore's and Mackie's arguments against reductive and anti-reductive accounts of normativity. This sort of problem does not arise, however, for facts that are normatively inert. Though I have a reason to fly to Edinburgh, if I want to get there in the shortest possible time, we wouldn't say that if someone judges that flying is the fastest way to get to Edinburgh she is *ipso facto* motivated to fly to Edinburgh. Since the judgement that flying is the fastest way to get to Edinburgh does not motivate one to fly to Edinburgh, the fact that flying is the fastest way to get to Edinburgh is not a categorically normative fact. Thus, in order for the sceptic to be able to make use of Moore's and Mackie's arguments, semantic correctness must supply categorical reasons—it must be that a speaker's judgement that some use of an expression is correct necessarily motivates her to make that use of the expression.

This suggests that several intuitive reasons for thinking that there might be categorical semantic obligations fail. For instance, it makes sense to say that if you want to speak the truth, you ought to say that the sky is blue; but this is not a categorical semantic obligation. This obligation is not categorical because if I no longer want to speak the truth, then I no longer have reason to say that the sky is blue. Similarly, obligations that are instrumental to my desire to communicate cannot be categorically prescriptive because their motivating force disappears with the desire to communicate. An exception to this general rule might seem to be obligations that are instrumental to my desire to speak meaningfully. One might argue that obligations that are apparently instrumental to the desire to speak meaningfully are in fact categorical semantic obligations—just as we might say that obligations that are instrumental to the desire to be rational are really categorical obligations. However, this analogy falls apart, because even though we might think that an agent who lacks the desire to be rational still ought to act rationally, we do not think that a speaker who lacks the desire to speak meaningfully nevertheless ought to do so. If that were true, kindergarten teachers who frequently sing nonsense rhymes would violate categorical semantic prescriptions.

It might be tempting to point out, right away, that the sceptical challenge is only to account for instrumental prescriptions. For Kripke, the problem is that 'if I intend to accord with my past meaning of "+", I *should* answer "125".'[5] Is this 'ought' not instrumental to satisfying my intention to accord with my past meaning? Indeed, it might seem so. However, Kripke's idea is that the intention to conform to a certain pattern of usage is *constitutive* of meaning something by an expression. Thus 'if I intend to accord with my past meaning … ' is a paraphrase of 'if I mean what I did previously by "+"', which, given the assumption that I meant *plus* previously, is just to say, 'if I mean *plus* by "+"'. Hence, this *categorical* prescription can be expressed as conditional only on what I mean. That is, although the prescription has a conditional form, it is not an instrumental 'ought'. The assumption is not that using '+' as one ought is a means to meaning *plus* by '+', but that, necessarily, if one means *addition* by 'plus' one ought only to use 'plus' in certain ways. Even if I no longer want to speak correctly, if I mean *addition* by 'plus', my mere judgement that some use accords with the addition function must motivate me to make that use of 'plus'.

---

[5] Kripke 1982, p. 37.

The upshot, then, is that the problem of accounting for semantic prescriptions will be damaging *if* it is possible to show that there are meaning-constituting semantic oughts that are not merely instrumental to satisfying the desires or aims of the speaker. In the next sections, I examine some of the more compelling reasons to think that there are such semantic prescriptions.

MISTAKES AND LIES

Kripke presents the normativity thesis in the course of his argument against the dispositionalist, whose failure to naturalise meaning is diagnosed as a failure to capture semantic oughts. Kripke's dispositionalist maintains that the correctness conditions of a speaker's terms, and thus what she means by them, can be read off from her dispositions. That is, the dispositionalist says that a speaker who means *horse* by 'horse' is one who is disposed to apply 'horse' to some creature if and only if it is a horse. Kripke points out that the dispositionalist will have difficulties dealing with people who are disposed to make mistakes. Imagine, for example, someone who systematically forgets to carry when doing complex sums. We would normally want to say that such a person systematically makes mistakes. The dispositionalist cannot say this on pain of circularity. If someone is disposed to give the 'right' answer under some conditions and the 'wrong' answer under other conditions, then which disposition determines what the speaker means? The dispositionalist must specify, in non-semantic, non-intentional terms, which of the speaker's dispositions are meaning constituting and which are error producing. But this, Kripke argues, the dispositionalist cannot do.[6] Whether or not Kripke is right about the prospects for dispositionalism, he diagnoses the dispositionalist's failure as the failure to capture semantic prescriptions. He says that a full specification of my dispositions will tell you what I *will* do, never what I *ought* to do. The dispositionalist fails, that is, because he fails to capture semantic 'ought's. Indeed, the 'problem of error' is a variant of the familiar problem, made prominent by Dretske, that any theory of representation must allow for the possibility of misrepresentation.[7] It is tempting to gloss this, as Kripke does, in prescriptive terms. But in the end, the gloss is just gloss and nothing prescriptive lies beneath.

---

[6] Kripke 1982, pp. 42–5.

Kripke is right to point out that dispositionalism, at least as he formulates it, faces a difficulty. The dispositionalist seems to say that the correctness conditions of a term can be read off of a speaker's dispositions to use that term; that we can find out what 'horse' refers to by looking at the conditions under which a speaker would use the term 'horse'. The problem, as Kripke casts it, is that under some conditions, the speaker would apply the term 'horse' to non-horses, and will thereby have made a mistake. But this is just a special case of a more general difficulty for dispositionalism which has nothing to do with mistakes, let alone semantic 'ought's. The problem for the dispositionalist results from his trying to identify the correctness conditions of a term with the conditions under which I would utter the term. Yet, not only are there occasions under which I would apply 'horse' to a non-horse, but there are other occasions under which I would fail to apply 'horse' to a horse. The dispositionalist seems constrained to suppose that there are some horses that do not fall within the extension of 'horse', just as he seems constrained to suppose that some non-horses fall within the extension of 'horse'. In both cases in which I apply 'horse' to a non-horse and in which I fail to apply 'horse' to a horse, I do not necessarily make a mistake. Sometimes, when I apply 'horse' to a non-horse, I am lying, joking, speaking ironically, or speaking sarcastically, in which case we would not say that I have made a mistake or fail to speak as I should. Similarly, if I fail to apply 'horse' to a non-horse, it might merely be because I have better things to do with my time, not because I have made a mistake. It is true that on some occasions when I either fail to apply 'horse' to a horse or when I apply 'horse' to a non-horse, I have made a mistake, but it is not the fact that these are mistakes that poses a problem for the dispositionalist. The point is that if the dispositionalist solution is a failure, it is a failure because our dispositions do not generate determinate correctness conditions, not because they fail to account for the wrongness of mistakes.

These considerations bring to light an important point: that the correct use of a term is not always the use that we ought to make. Sometimes, given our intentions, we ought to use our words incorrectly—if we want to lie, amuse, or mislead.[8] Yet, according to *Normativity**, a speaker who means something by a term ought only to apply it correctly, where 'applies correctly' stands in for 'refers to', 'is true of', 'denotes', and so forth. Thus *Normativity** seems to make it a condition of meaning something by a term that a speaker ought only to speak the truth. But this requirement is too strong. Of course, there may be all sorts of extra-linguistic consequences of what we say that determine what we ought to say. It may be that I morally ought to tell the truth. It may be that in some special cases I morally ought to tell a lie. But in all such cases what I mean will only be relevant to what I ought to say. The fact that some use of a term is correct—that is, refers to, is true of, or denotes its object—does not seem to carry with it any obligation. What grounds could we have for thinking that when someone lies she contravenes a specifically *semantic* obligation to tell the truth?

---

[7] Dretske 1986. Although Dretske (2000) argues that misrepresentation has nothing to do with normativity.

One might argue that what we should learn from this case is that lying is parasitic on truth-telling, and that what someone means by an expression is constituted by obligations she incurs on occasions of *sincere* use.[9] So, what Matilda (who told such dreadful lies) means by a term is constituted by the obligations she incurs when she aims to tell the truth. These obligations are contingent on her aim to tell the truth, but they are nevertheless essential to what she means—in order for her words to have meaning, she must aim to use them correctly at least most of the time. On those rare occasions when Matilda does not strive to tell the truth, the meaning of the term—as constituted on occasions of sincere use—simply carries over. Thus, Matilda can meaningfully lie, but at the same time, meaning is constituted by semantic obligations.

There are several difficulties with this suggestion. The first is that the obligations incurred on occasions of sincere use will not yield the right pattern of semantic obligations. Suppose, for example, that Matilda wants to sincerely tell us that her house is on fire. In that case, given what she wants, and given that her house is on fire, she ought to say 'my house is on fire'—if she is to satisfy her desire to tell the truth. Matilda does not acquire, on that occasion of sincere use, the obligation to use 'is on fire' of something only if it is on fire. Rather, she acquires the obligation to utter the sentence 'my house is on fire' on that occasion, because the sentence is true, on that occasion. The obligation itself only concerns a single occasion of use. Now, suppose that Matilda has incurred obligations, on separate occasions, to say 'this house is on fire', 'this log is on fire', 'this oven mitt is on fire', and so forth. All these taken together do not add up to the obligation to say that something is on fire only if it is. Matilda just has a list of obligations to say one thing in one circumstance, another thing in another. Since it is permissible for Matilda to lie, on the suggestion under consideration, there will be some occasions on which it is not the case that she ought to apply her terms correctly. There is thus no way that the conjunction of Matilda's singular obligations will add up to the obligation to apply 'fire' only to fire. If the suggestion does not even give us *Normativity**, however, it is difficult to see why these semantic obligations should be thought to determine what Matilda means, given the intuitive, truth conditional, or correctness conditional picture of meaning.

---

[8] Wikforss 2001; Glüer 1999*b*.

[9] This has been suggested to me in conversation by Martin Kusch and Simon Blackburn.

The second problem is that these obligations are only instrumental to satisfying Matilda's desires. Even if these oughts are genuinely or essentially semantic, they are nevertheless merely instrumental—Matilda ought to say 'my house is on fire' only as a *means* to satisfying her desire to tell the truth, in those circumstances. What Matilda means here is *relevant* to what she ought to say, certainly, but this does not ensure that meaning is normative in the sense that is required for the sceptic. For Matilda could judge that *if* she wants to tell the truth, she ought to say 'my house is on fire', and nevertheless feel no motivation whatsoever to tell the truth.

PRIMA FACIE *OBLIGATIONS*

Perhaps Matilda does have a semantic obligation to tell the truth, which is not contravened when she wants to lie, but overridden. Obligations that can be overridden are called *prima facie* obligations.[10] For example, if I promise to meet Cathy tomorrow for tea, some would say that I undertake a *prima facie* obligation to meet her tomorrow for tea. This is only a *prima facie* obligation because I might be justified in breaking my promise under extenuating circumstances—for instance, if at the time of our tryst, Krister requires me more urgently. My duty to Krister overrides my obligation to Cathy, but this does not mean that I had no obligation to Cathy in the first place, nor does it mean that the obligation went away; it was just trumped, as it were, by my obligation to Krister. The fact that I have a *prima facie* obligation to Cathy is the fact that, *all things being equal*, I ought to drink tea with Cathy. Hence, one might argue that my semantic obligations are *prima facie* obligations. If I mean something by an expression, I have an obligation to tell the truth, all things being equal; my obligation will be overridden if I wish to tell a lie.

---

[10] Cf. Ross 1987.

I suspect that the surface plausibility of this suggestion trades on our smuggling in desires on which the obligation to tell the truth is contingent. In order to test whether I have a *prima facie* obligation to use an expression correctly, we need to consider a case in which all is genuinely equal. So, not only should we assume that I have no desire to tell a lie, but that I have no desire to tell the truth—otherwise, my obligation to apply a term correctly might just be contingent on that desire. Moreover, we have to rule out the possibility that the obligation is really moral or prudential. So, we have to assume that I have no desire to communicate; that if I have an audience, my audience is utterly indifferent to whether or not I tell the truth; and that nothing whatsoever hangs on what I say—it will not lead me to act imprudently or irrationally. The question is, given that my audience and I are indifferent to whether I tell the truth, and given that nothing hangs on what I say, am I still obligated to tell the truth? I see no reason why. If what I say affects nothing and no one, it hardly seems to matter whether I apply the term 'horse' only to horses. One might suppose that it is necessary for an account of a speaker's meaning that she has a *prima facie* obligation to use her terms correctly.[11] However, this does not seem to be true. Given that I mean *horse* by 'horse', on the intuitive view, this means that 'horse' refers to all and only horses. Given *Norm-Relativity*, however I use the term 'horse', it will be correct in application to something if and only if it is a horse. This seems to be sufficient to account for the meaning of 'horse'; nothing is added by saying that, *ceteris paribus* I ought to use the word 'horse' only if it is correct to do so.

Furthermore, there is a difficulty in supposing that a speaker has a *prima facie* obligation to use her words correctly—because the obligation would be too easily overridden. The thought is that I have an obligation to use the term 'horse' only if it is correct to do so, unless I have a desire to lie or mislead. In other words, my obligation to use 'horse' correctly is overridden merely by my desire not to do so. In contrast, if I have promised to meet Cathy for tea, my obligation to keep my promise cannot be overridden by my mere desire not to do as I promised. For a *prima facie* obligation to be overridden, it must be overridden by another obligation, such as an obligation to help Krister, which bears more weight. If we assume that an obligation can merely be overridden by a desire not to carry it out, this seems to undermine the idea that an obligation gives me a categorical reason to do something. Thus, it seems more plausible to explain the intuition that a speaker ought to use her words correctly by appeal to a hypothetical obligation contingent on her desire to speak truthfully.

---

[11] Cf. Wikforss 2001.

THE NORM OF TRUTH

It is frequently suggested that semantic obligations follow from 'the norm of truth' or from the intimate relationship between meaning and truth. If this just means that we have obligations contingent on our desire to tell the truth, it should be obvious, for the reasons given above, that this will not do. Obligations that follow from the aim of truth-telling cannot constitute meaning, because such obligations are inessential to meaning. My obligation to say something because it is true is not a *semantic* obligation, though it might be moral, prudential, or merely contingent. Moreover, contingent obligations can be analysed reductively, and so are not what the sceptic is after.[12]

The argument from truth to semantic obligations can be more sophisticated, however. It could be argued that truth-telling is in some sense *necessary* for meaning. If that is so, then obligations contingent on a speaker's desire to tell the truth are essential to meaning. In striving to tell the truth, the speaker incurs a contingent obligation to tell the truth. This obligation, although contingent on a desire, is nevertheless essentially semantic because the desire on which it is contingent is essential to meaning. Gary Ebbs suggests roughly this characterization of Kripke. He says:

> Kripke thinks that meaning is normative because he believes that *our grasp of rules is inextricably linked to our aim of asserting and judging in accordance with the truth*. The links go in both directions: we can't strive to make true assertions unless we can grasp rules that determine the truth conditions of our assertions,

---

[12] Cf. Wikforss 2001; Glüer 1999*b*; Glüer 2001.

and we can't grasp rules that determine the truth conditions of our assertions unless we strive to make true assertions.[13]

Certainly, we cannot make true assertions unless we grasp the meanings of our expressions. But does it make sense to say that we cannot grasp the relevant rules or meanings without striving to make true assertions? I think not. First, if we consider what I mean *now*, it does not seem as if I have to strive to tell the truth to preserve my meaning. Obviously, I can *now* strive to tell a falsehood, to tell a *lie* without meaning something non-standard by my words. It could be argued, however, that when I *learned* the meaning of 'blue' in the past, at *that* point I had to strive to tell the truth. However, I can think of no reason why this should be true *a priori*. Perhaps adults need to strive to tell the truth, at least sometimes, when teaching a language, but it seems unlikely that a child needs to strive to tell the truth in order to learn the language. What would striving to tell the truth even amount to in the early stages of language learning? Indeed, it is far more likely that children must strive to achieve pragmatic ends when learning a language—such as getting biscuits—rather than to tell the truth. Moreover, language is learned in a wide variety of different ways, not all of which require the child or adult to strive to speak the truth. Jokes and misrepresentations play an important role in learning to speak and nursery rhymes aid language learning, but can nevertheless be nonsensical.

The expression 'norm of truth' seems to suggest a rule, and once again, the rule can either be prescriptive or not. If it is not prescriptive, nothing is implied about what a speaker ought to do. For example, suppose that for a speaker to understand any meaning or have any belief, it is necessary that she grasp, however tacitly, that the belief *that p* is true if and only if p. Even if this is necessary for understanding, nothing follows about what a speaker ought to do. The semantic obligations only follow from there being a 'norm of truth' if the norm is prescriptive, such as, 'you ought to have only true beliefs' or 'you ought to aim to make sure that most of your beliefs are true'. If the sceptic can persuade us that it is necessary for grasping any meaning, or having any belief, that one follows a prescriptive norm of truth, then it seems to follow that meaning is normative.

To defend this line of reasoning, one might employ (albeit out of context) Michael Dummett's argument for the value of truth.[14] Dummett makes an analogy between the concept of truth and that of winning at a game such as chess. He observes that you could specify all the rules of chess, specify how all the pieces move, what constitutes winning or losing the game, and still something would be left out: that the whole *point* of the game is to win. Similarly, Dummett argues, specifying the conditions under which a belief is true leaves out the important fact that truth is valuable, that truth is the aim of belief. If truth is the aim of belief then we ought to believe only what is true. And, if we ought to believe only what is true, then we ought to apply concepts only when it is correct to do so.

---

[13] Ebbs 1997, p. 19.

[14] Dummett 1978.

Dummett's argument is compelling because we do value truth—we clearly want to believe only what is true. The trouble, however, is that the obligations in question are instrumental, as opposed to categorical. The obligation to believe only what is true can be seen to be instrumental to satisfying my desire for all my desires to be satisfied. It is a truism that a true belief is more likely to lead to a successful action than a false belief. So, if I have true beliefs, I will be more likely to be successful in getting what I want. This is obviously true of particular beliefs. Consider Pam, who wants some ice cream, and who believes that there is ice cream in the freezer. Clearly, Pam wants her belief to be true because if it is true, then she can successfully act on it—she can go to the freezer and find ice cream. This is not just to say that Pam wants there to be ice cream in the freezer; rather, she wants to believe that there is ice cream in the freezer only if there is. If there is no ice cream in the freezer, she would rather believe that there is no ice cream in the freezer, so she will be prevented from making an unnecessary trip. Now, it is easy to generalise from these particular cases. Pam has desires. Having true beliefs is more likely to produce actions that satisfy Pam's desires than having false beliefs.[15] That is, if Pam has any desires, she has an instrumental reason to acquire only true beliefs, since those are more likely to lead to the satisfaction of her desires.

Horwich points out that the foregoing explanation of why we want to have true beliefs captures another important intuition.[16] We do not take the 'norm of truth' to imply that we ought to have no beliefs at all, for that is surely the best way to ensure that we only have true beliefs. Indeed, assuming simply that there is a categorical obligation to believe only what is true does not explain why we value truth, which is why it would seem to be just as well to have no beliefs at all. The instrumental explanation fares better because it explains *why* we value truth, and explains why we want to have some true beliefs, not just avoid having false ones.

---

[15] See Whyte 1990; Mellor 1990; Horwich 2000.

[16] Horwich 1998, 2000.

Given that the value of truth can be explained prudentially, it is difficult to see what motive one would have for insisting, on the grounds that truth is valuable, that its value must be irreducible. If we can explain the obligation to believe only what is true by appeal to our desires, then what more work would a categorical semantic obligation do? It would have to be the case that *Norm-Relativity* is insufficient for meaning; that we need to adopt the more demanding *Normativity**. But the observation that truth is valuable does not make that additional commitment necessary.[17]

COMMUNICATION AND COMMUNITY

Another reason why people readily believe that there are semantic prescriptions is that we expect people who are learning a language to abide by grammatical and semantic rules, to speak as they ought. When we teach a language, we tell people what they *ought* to say in a given situation. Wittgenstein himself spent considerable energy describing the way in which students learn to follow rules, particularly focusing on what the teacher knows and says. Teachers, in this picture, frequently say things like: 'no, that is wrong', 'you must say *this*', 'you should go on in the same way' and so forth. It might look, from these descriptions, that semantics abounds with obligations.

It is not difficult to see, however, that the overt prescriptive language of the classroom has nothing to do with semantic obligations. According to *Normativity**, the structure of my semantic obligations is constitutive of *what I mean*; it is because I mean *blue* by 'blue' that I ought to apply 'blue' only to blue things. In the classroom, by contrast, the students' obligations are a function of the social situation that they are in, not a function of what they mean. In the classroom, *ceteris paribus*, a student ought to do what the teacher tells her to do. However, this is not a *semantic* obligation. The students have an obligation simply to do as they are told. If the teacher tells them to use 'green' for all and only green things, then *that* is what they ought to do. If the teacher tells them to use 'green' for green things until tomorrow and then for blue things, *that* is what they ought to do. What the teacher means by 'green' and tries to convey to her students, is thus merely relevant to what they ought to do.

It might be objected that there is an overriding obligation in the classroom for the teacher to teach the appropriate meanings to students, so that they ought to learn to apply *blue* only to blue things, *square* only to square things and so forth. However, even if there are such obligations in the classroom, they are clearly contingent on the aim of communication. We want little Otto to mean *metaphysics* by 'metaphysics' so that he will be able to communicate with the rest of us in the future. That is, in general, we want children to learn to use words as we do. But this says nothing about obligations *constituting* meaning—it does not tell us that if Otto really means *metaphysics* by 'metaphysics' then he ought to apply 'metaphysics' only to metaphysics. Rather, so long as Otto wants to communicate with us, or so long as we want to teach him to communicate with us, then he ought to use 'metaphysics' as we see fit. In this case, once again, meanings might be *relevant* to how Otto ought to speak, but Otto's meaning *metaphysics* is not constituted by the pattern of his obligations.

---

[17] For further argument on semantic normativity and the value of truth, see Wikforss 2001.

Mark Lance and John Hawthorne proposed a related argument for the normativity of meaning. Lance and Hawthorne argue that all of our meaning discourse is normative in the sense that it has prescriptive force. So, if I am a radical interpreter attempting to compile a translation manual for jungle-dwellers who I am studying, if I write that 'gavagai' means 'rabbit' what I do is *prescribe* a consistent usage for jungle-dwellers and English-speakers. That is, '"gavagai" means "rabbit"' means something like 'for all speakers of both languages, if they use "gavagai" in some circumstance, they should accept that "rabbit" is correctly used in those circumstances as well.' 'Translational claims', they say, 'are not descriptions in the first place. They are, rather, speech acts whose point it is to influence the structure of social practices, to impose a (possibly new) socially recognised constraint upon behaviour'.[18]

Even if what Lance and Hawthorne say about translation were true, however, *Normativity** would not follow. If, when compiling a translation manual, I prescribe a future use for expressions of both languages, these prescriptions cannot, in retrospect, constitute what the two communities meant by 'gavagai' and 'rabbit' respectively. More importantly, however, Lance and Hawthorne's argument cannot show that meaning is, in general, constituted by 'ought's as a consequence of some prior act of prescription. For, if meaning is prescriptive, then sentences of the translation manual themselves must have meaning. According to *Normativity** for the words of those sentences to have meaning, it must be the case that they ought to be used under some circumstances rather than others. If those obligations are constituted by some prior prescription, then the words of those prior prescriptive sentences must have meaning and so presuppose some further prior prescriptions. If the view were intended to show that meaning is essentially prescriptive because prescriptive sentences determine meanings, it winds up in a regress.

---

[18] Lance and Hawthorne 1998, p. 61.

A similar line of argument for *Normativity** might go as follows: the origin of semantic obligations lies in the community, in the social conventions governing the correct use of terms in natural languages. The operative idea would be that membership in a given linguistic community implies the adoption of the conventions of that community. Unlike the scenario where I adopt a semantic standard on my own, in the context of communal conventions, I actually make something of a *promise* to the others in my community. These commitments to the communal conventions then yield obligations to the community to carry out the actions to which one had committed.

Margaret Gilbert has defended the idea that joint commitments, or promises to members of a community, create obligations on the part of those who commit or promise. Gilbert argues that when a joint commitment is made, two people are committed to doing something as a body, so they are each individually obligated to perform the actions that are necessary for the proper function of the group.[19] On this basis, it could be argued that the need for communication and co-ordination in a community leads to the joint commitment on the part of the community as a body to adhere to the local linguistic conventions. In the course of learning the meanings of words, a child is 'brought into the community', and in the process, the child incurs obligations to adhere to the linguistic conventions of the community.

Once again, if this view is taken to supply a defence of *Normativity**, it will fail. For, such communal conventions are constituted by intentions and beliefs—we all intend to use our expressions in such and such ways, and we all believe that others will do the same. If you and I commit to going for a walk together, we must both form intentions—that is we must both form the intention to go for a walk together. If those intentions constitute our obligations, they must have content. And if it is true that for the intentions to have content, we must be obligated to apply the words and concepts involved in those intentions in certain determinate ways, then we must have formed some prior intention to use the words or apply the concepts involved in the intentions in certain determinate ways. In general, intentional acts, such as forming commitments, might constitute social conventions, but then those social conventions cannot be taken to constitute the obligations that then constitute the contents of the intentions in the first place. Content cannot be constituted by social conventions if intentional states are necessary to constitute social conventions. The result will either be a vicious circle or a vicious regress.

---

[19] Gilbert 1996.

A similar criticism applies to Brandom's defence of normativity in terms of social conventions. Brandom says:

> The particular norms of concern ... are *discursive* normative statuses, the sort of commitment and entitlement that the use of concepts involves. These norms, it will be claimed, are instituted by *social* practices. ... Elaborating an account along these lines is pursuing three of Wittgenstein's grand themes: the insistence on the *normative* character of language and intentionality, the *pragmatist* commitment to understanding these norms in terms of practices ... and the recognition of the essentially *social* character of such norms.[20]

To refresh the memory, Brandom's picture is one in which communities 'institute' semantic or conceptual conventions by taking attitudes of approval or disapproval towards bits of behaviour. Any behaviour that tends to meet with disapproval is thereby wrong, by the community's lights, whereas behaviour that tends to meet with approval is thereby right. What a given speaker ought to say, then, is what tends to meet with approval rather than disapproval. Thus, the social practice of treating types of behaviour as correct or incorrect creates semantic obligations.

Whatever the merits of this as a picture of meaning, it cannot be marshalled in *defence* of the thesis that meaning is normative. Some argument must be given for saying that meaning is *essentially* social, and in Brandom's case, the argument that is given depends on the assumption that meaning is normative. Like Kripke, Brandom assumes that the normativity of meaning is basic to our intuitive view. He then proceeds to argue that no naturalistic theory of meaning would do—and concludes that normativity must be assumed as primitive.[21] It is not clear why this would make meaning essentially social; since whatever normativity is brought in at the social level might equally be brought in at the individual level.[22] Nevertheless, by introducing communal practices, Brandom purports to explain how we create obligations—obligations he takes to be there, standing in need of explanation. This just fails to

[20] Brandom 1994*a*, p. 55.

[21] Ibid., pp. 42–5.

[22] See Blackburn 1984.

speak to my concerns; my worry is that there are no semantic obligations to be explained in the first place.

Moreover, there is once again a distressing circularity in the claim that semantic obligations are constituted by communal attitudes. Brandom seems to think that we can catapult ourselves into language via our attitudes. But this is like pulling ourselves up by our own socks, or turning ourselves into promise-keepers by promising to keep our promises. If semantic obligations are supposed to be essential to content, then there can be no content without the relevant obligations. But if the obligations are created by our attitudes, then they are themselves the products of contentful states. If I approve of your saying 'red' in the presence of this apple, my endorsement has content—I endorse *that you say 'red' in the presence of this apple*. If I can have an attitude with this content, then for one thing, I must grasp the concept *apple*. Assuming *Normativity**, this means that I ought to apply the concept *apple* only if it is correct to do so. But if obligations such as these are uniformly created by attitudes, then how is my obligation to apply the concept *apple* created? If by another attitude, we embark on a regress. If by my dispositions, the view will ultimately bottom out into a dispositional theory, which Brandom rejects on the grounds that it fails to accommodate normativity.[23] So, Brandom faces a dilemma. Either way, we have no reason to think that we must assume that there are semantic obligations, standing in need of an explanation.

Finally, it should be noted that the very idea that there must be *norms* or *conventions* that determine linguistic meaning has come under sustained attack by Davidson and others.[24] If Davidson is right that convention is not necessary for meaning, then the arguments from convention to normativity would fail. I have argued, however, that even if we grant that conventions are necessary for meaning, it still would not follow that meaning is normative in the relevant sense.

CONTRACTUAL COMMITMENTS

Kripke, Wright, and McDowell (among others)[25] suggest that we are committed to *Normativity** because we intuitively subscribe to what

---

[23] For a more detailed criticism of Brandom see Hattiangadi 2002.

[24] Davidson 1984; Glüer 2001; Bilgrami 1993.

[25] Cf. Brandom 1994*a*; Kripke 1982; McDowell 1993; Pettit 1990*a*; Wright 1980. Kripke (1982) claims that *intentions* yield obligations, whereas Wright (1980) and

Wright calls the 'contractual theory' of understanding. According to the contractual theory, to understand an expression is to be committed to a particular pattern of application of that expression. This commitment then confers an obligation on the speaker to carry out the pattern of application in question. McDowell presents this assumption as follows:

> We find it natural to think of meaning and understanding in, as it were, contractual terms. Our idea is that to learn the meaning of a term is to acquire an understanding that obliges us subsequently—if we have occasion to deploy the concept in question—to judge and speak in certain determinate ways, on pain of failure to obey the dictates of the meaning we have grasped; that we are 'committed to certain patterns of linguistic usage by the meanings we attach to expressions'.[26]

The reasoning behind this goes roughly as follows: to accept the intuitive picture of content is to accept that a term has content in virtue of its correctness conditions. In order for me to mean something by an expression, I must have adopted some standard of correctness for that expression. In adopting a standard of correctness, I must be committed to acting in accordance with that standard. Otherwise, what is it for me to have adopted the standard in the first place?

The trouble is that the contractualist needs to show that when I form a commitment or an intention to accord with a standard, I incur the right set of obligations. That is, the contractualist needs to show that if I grasp the concept *horse*, I ought to apply it only to horses. This is because semantic obligations are only problematic for the naturalist if those obligations are meaning-determining. The contractual theorist thus holds that in order for me to mean something by an expression, I must be committed not just to using the standard in any old way, but to meeting it; I am committed to *carrying out* a particular pattern of use for the expression I understand. Wright suggests this when he says that we intuitively suppose that '[w]hen I assent to the rule: F is to be applied only to individuals which are ϕ, I commit myself to a quite determinate way of using F.'[27]

If the contractual theory is indeed our intuitive view—which I doubt—we ought to revise our intuitive view. To begin with, if an agent were to form the intention to use an expression correctly, then

---

McDowell (1993) both suggest that *commitments* yield obligations. I discuss both versions.

[26] McDowell 1993, p. 257. McDowell quotes from Wright 1980, p. 21.

[27] Wright 1980, p. 36.

given a referential standard of correctness, this would be tantamount to her forming the intention never to lie. Think again of Matilda, who claims that her house is on fire when she knows full well that it is not. In order for her lie to be genuine—and not just a case of mistaken belief—she has to form the intention to say 'the house is on fire' despite her awareness that the statement is false. Assuming the contractual theory, however, if Matilda means *fire* by 'fire', she must have the intention to apply the expression only if it is true. The trouble is that when she lies, of course, she forms the intention to utter a statement that she believes to be false. This obviously results in a difficulty: Matilda cannot simultaneously intend to apply x to a only if a is f, and intend to apply x to a when it is not the case that a is f. Intentions (as opposed to desires or beliefs) cannot stand in conflict, since intentions are the result of firm decisions to carry out actions. Hence, at best, the contractual theory makes it impossible for anyone to consistently form the intention to lie. At worst, the contractual theory implies that someone who forms the intention to lie thereby means something non-standard by her words and thus ends up unwittingly telling the truth.

One might think that this objection can be met if the contractualist sticks to the idea of someone being *committed* rather than merely *intending* to act in accordance with a rule.[28] However, if Matilda is committed to apply x to a only if a is f, then she still cannot intend to apply x to a while sincerely believing that it is not the case that a is f without affecting her commitment. A one-off intention to lie might not destroy her commitment, but if Matilda intended to lie too many times, presumably we would no longer wish to say that she was committed to speaking the truth. But it is odd to think that what Matilda means is hostage to her intentions in this manner. Compulsive liars, like Matilda, do not gradually begin to attach non-standard meanings to their expressions. When Matilda lies, jokes, or misleads, she *presupposes* the ordinary or literal meanings of her terms, so she cannot change them just by lying.

Perhaps the contractualist could meet this objection by restricting himself to conceptual, as opposed to linguistic, content. In Matilda's case, to say that she lies is to say that she believes that the house is not on fire, when she says that it is. In her mind she carries out her semantic obligations, while in her speech she uses the meanings of her words to convey false information. Given that she grasps the concept *fire*, she is committed at least to think *fire!* only when there are flames about.

---

[28] I am grateful to Bob Hale for making this suggestion.

If this is the contractualist's argument, however, it will succumb to an objection that was made previously: we have here a prudential obligation masquerading as a semantic one, and the prudential obligation can be given a reductive explanation. Although there might be value in only believing what is true, this seems to follow from a general prudential obligation contingent on a desire to get what we want. The obligation in question is therefore not semantic.

Moreover, even if Matilda were committed to telling or believing the truth, there is no good reason to think that undertaking a commitment implies that one thereby incurs an obligation to carry it out.[29] Someone might be committed to torturing everyone who disagrees with her, but this does not imply that she is obligated to torture everyone who disagrees with her. Of course, she might have an obligation to torture everyone who disagrees with her which is *contingent* on her commitment to doing so, and similarly, Matilda might have an obligation to believe or tell the truth that is *contingent* on her commitment to doing so (if she has such a commitment). But these contingent obligations are not categorical and so would not give rise to Moore's or Mackie's arguments against attempts to account for them.

It does not seem as if meaning or understanding some content implies an intention to adhere to the correct pattern of use. Moreover, even if we *did* form intentions towards the use of expressions when we grasped their meanings, a categorical obligation still would not result. However, Kripke seems to think otherwise, and he suggests that the key to understanding the relationship between meaning, intention, and obligation is the notion of an *internal* relation. For instance, he says that 'Wittgenstein's view that the relation between the desire (expectation, etc.) and its object must be "internal", not "external" parallels corresponding morals drawn about meaning in my text below (the relation of meaning and intention to future action is "normative, not descriptive").'[30] It might be possible to argue that the relation between intention and future action is *internal* and that this implies that the relation between intention and future action is *prescriptive*.

There are internal relations both between a desire and the conditions that fulfil it, and between an intention and the action that constitutes carrying it out. If I intend to eat an apple, then I *must* eat an apple in order to fulfil my intention. Nothing else will do. Intentions, as a result of their internal relations, seem to specify cases in which they are fulfilled and cases in which they are not fulfilled. This distinction between success and failure in carrying out one's intentions is supposed to support the distinction between correct and incorrect uses of words. It is my intention towards the use of words that supplies the distinction between what I ought to say and what I ought not to say.

---

[29] Vogel Cary 1975.

[30] Kripke 1982, pp. 25–6.

However, intentions will only get us contingent obligations, they will not yield irreducible and essentially semantic ones. Unless we interpret 'I intend to *x*, therefore I ought to *x*' as a truncated hypothetical imperative, there is no reason to think that my intention results in an obligation. Eating a banana instead of an apple does not comply with my intention to eat an apple but that does not mean that I have failed to do what I *ought* to do in anything other than the trivial sense that I have violated the hypothetical imperative conditional on my intention.[31] Similarly, if I intend to use '+' to mean *addition*, then in order to carry out my intention, the only thing that will qualify is my saying that 57 + 68 = 125, when queried. This is not what I ought to say, but the only thing that constitutes carrying out my intention to add.[32]

On balance, it seems that although it might be intuitive to suppose that adopting a standard for the correct application of an expression might be necessary for meaning, it does not follow that speakers have specifically semantic obligations. All told, semantic realists need not be committed to the contractual theory, and even if they were, this would not commit them to *Normativity*.

RULES

In the background of this discussion lurks the analogy between meaning and rule-following. Indeed, the analogy between meaning and rule-following is largely responsible for the popularity of the normativity thesis. If we grant that meaning is a matter of rule-following, we must take correctness conditions, or truth conditions, to be rules that a speaker follows. But rules, as the intuitive line of reasoning might run, tell us what we ought to do. The Highway Code tells us how we ought to drive. The Ten Commandments tell us how we ought to act. So, if there are semantic rules, surely they will tell us what we ought to say.

---

[31] Wikforss 2001. See also Bilgrami 1993.

[32] Someone might argue that the reason I am reluctant to countenance these prescriptions is simply that I have mistaken them for strong semantic obligations when they are *prima facie* ones. One could argue that forming an intention constitutively involves incurring a *prima facie* obligation to carry it out. But the example in which I intend to eat an apple does not seem to involve any overriding moral obligations—nor any other obligations, for that matter. This is simply a case in which it looks as if, all things being equal, the intention to eat an apple does not essentially confer an obligation on me to eat an apple.

For this argument for *Normativity** to go forward, semantic rules clearly need to be both meaning-constituting and prescriptive. Semantic rules will have to be meaning-constituting because otherwise they will not be essential to meaning; if they are not essential to meaning, they will pose no problem for theories of meaning. Of course, for the argument to go forward, semantic rules must be prescriptive, because if they are not, they will supply no semantic 'ought's. The trouble is that these two constraints pull in opposite directions.

Recall that a prescriptive rule is one that implies a prescription or an imperative, one that tells an agent to perform some action, *A*. Thus, a prescriptive rule is one that says 'Do *A*!' or 'You ought to do *A*', or 'If in circumstance *C*, do *A*!'. Of course, a rule can be prescriptive if it implies 'Do *A*!' given some of the relevant facts. For example, the Utility Principle, 'Do *A* if and only if *A* maximises overall utility' is a paradigmatic prescriptive rule. If some action *A* maximises utility, then the Utility Principle demands that an agent do *A*.

What, then, is a meaning-constituting rule? Recall that according to Searle, constitutive rules are those that define a practice or action; they are necessary for the possibility of performing that practice or action.[33] In contrast a regulative rule is one that regulates an antecedently existing practice or behaviour. Thus, a meaning-constituting rule would be one that supplies us with a way of describing linguistic behaviour that would not otherwise be available. Obviously, what a meaning-constituting rule will afford is the description of linguistic behaviour as distinctly *linguistic*, as *meaningful*. Moreover, a rule that is constitutive of the meaning of a term should tell us, presumably, exactly what that term means. Assuming semantic realism, a meaning-constituting rule for 'horse' should tell us that 'horse' applies correctly to all and only horses.

The question is whether a rule could be both meaning-constituting, in the above sense, and prescriptive.[34] Hans-Johann Glock has argued that a 'grammatical proposition' can satisfy both constraints. Consider, to begin with, one of Glock's examples of a grammatical proposition, 'All bachelors are unmarried'. Although Glock sometimes talks about rules, and although his claim is that these are norms or standards that we use, he does not formulate them as linguistic rules. Moreover, because we are concerned with whether these rules are prescriptive, it is important that they imply prescriptions or imperatives. Thus, consider the following prescriptive rule based on Glock's example of 'bachelor' and 'unmarried':

---

[33] Searle 1970.

[34] Searle (1964) argues that you can derive an 'ought' from an 'is' on the basis that, for instance, the constitutive rule for making a promise states that one places oneself under an obligation to carry out that promise. For the classic criticisms of this

> R3: If you say '*S* is a bachelor', don't say '*S* is married'; if you say that '*S* is married', don't say that '*S* is a bachelor'.

R3 is phrased negatively, because it is clearly not the case that if I say that someone is a bachelor, I ought also to say that he is unmarried. However, it does not seem as though it is necessary for understanding the concepts of *bachelor* and *married* that someone follow R3. For instance, someone might say that he is married on one occasion, to fend off unwanted attention, but say that he is a bachelor on another occasion to encourage the attention of someone else. That he might say this of himself without any change in his marital status or in the meanings he attaches to his words, implies neither that he fails to understand the concepts of *bachelor* or *unmarried*, nor that he speaks nonsense. On the contrary, he uses the expressions in the way he does *precisely because* he understands what they mean. Moreover, as it is formulated, R3 mentions only *one* of the correctness conditions of the concept *bachelor*; and this condition is satisfied by *spinster* as well. So, I could act in accordance with R3 and still mean *spinster* by 'bachelor'. Thus, R3 does not specify the meaning of 'bachelor' sufficiently accurately to disambiguate it from other alternatives, and R3 cannot be a meaning-determining rule.

Responses can be made to both of these objections. In the case of the first objection, Glock could argue that a rule like R3 needs to be restricted to concept application. I do not think *Glock* would argue this, because he stresses what we would *say*, for instance, when he writes that 'it is logically impossible for bachelors to be unmarried [*sic*] simply because we would never *call* anybody both "married" and "bachelor".'[35] Nevertheless, assuming that we could reformulate necessity at the level

---

argument, see Genova 1970; Hare 1969; Hudson 1969. Since Searle never argues that *meaning-determining* rules are prescriptive, I do not consider his 'is'–'ought' argument here. However, for a thorough discussion of why, on Searle's view, meaning-determining rules could not be prescriptive rules, see Glüer and Pagin 1999.

[35] Glock 1996, p. 202.

of concept application as opposed to linguistic usage, it could be reformulated as follows:

> R4: If you believe that *S is a bachelor*, don't believe that *S is married*; if you believe that *S is married*, don't believe that *S is a bachelor*.

The trouble with R4 is that it does not seem to be meaning-determining, and so still succumbs to the second objection mentioned above. Rather than determining the content of the concepts of *bachelor* and *unmarried*, R4 presupposes the contents of these concepts. It is only because we possess those concepts that we can understand a rule like R4, since the merely formal connection between the concept of a *bachelor* and the concept of *married* is insufficient to determine their content. To see this, consider a more formal characterisation of R4:

> R5: If you believe that *x is F*, don't believe that *x is G*; if you believe that *x is G*, don't believe that *x is F*.

Clearly, R5 applies to numerous different concepts. For instance, where *F* is the concept of *tomato* and *G* is the concept of *hippopotamus*, R5 says that I should not believe that something is both *F* and *G*. It makes no sense to think that a tomato is a hippo. However, if R5 characterises the conceptual link between *tomato* and *hippo* as well as between *bachelor* and *unmarried*, then we seem to have a problem. How can a grammatical rule linking *F* and *G* determine the contents of *F* and *G* if the rule holds for numerous different values of *F* and *G*? If the conceptual link between *tomato* and *hippo* is the same as that between *bachelor* and *married*, then a rule like R5 cannot qualify as a meaning-determining rule in the sense required by the sceptic.

The most obvious response to these objections is to say that meaning-determining rules come in *systems*, so that a single isolated rule is necessary but not sufficient for determining the meaning of an expression or the content of a concept, whereas the system of interconnected rules is both necessary and sufficient for the determination of meaning. This is a plausible suggestion, because with enough rules like R5, perhaps we could at least distinguish the concept of *bachelor* from the concept of *tomato*. But I doubt that even a system of rules such as R5 would be enough to determine meaning on the truth conditional account, for to do so, we need meaning-determining rules that will distinguish a concept from *all quus*-like alternatives. The rule or rule-system that the sceptic takes proponents of semantic realism to be committed to must be capable of funding a potentially infinite list of semantic obligations. But systematicity will offer no help in this case.

The trouble is that the meaning-determining obligations need to match up exactly with correctness conditions. But the sorts of rules that can be built out of Glock's necessary propositions are inferential rules. And inferential rules, even when systematised, will not give the full complement of correctness conditions that we intuitively assume. Imagine that we systematically rearranged the correctness conditions of all our concepts, so that the conceptual links would be retained, but the concepts altered. Suppose, that is, that we reinterpret all of our concepts in a non-standard way, so that they switched correctness conditions at some time, *t*. Before *t*, *green* applies correctly to green things, and after *t*, *green* applies correctly to blue things. Moreover, before *t*, *emerald* applies correctly to emeralds and after *t*, *emerald* applies correctly to sapphires. If we reinterpret the relevant concepts systematically so that their correctness conditions change at *t*, we have a divergent interpretation of a whole system of conceptual links that is not ruled out by the conceptual links themselves. That is, a system of rules like R5 would not adequately disambiguate alternative conceptual contents, and so even a system of rules such as R5 would not qualify as meaning-determining in the requisite sense.

Of course, the inferentialist says that there is no difference between the two interpretations—if the inferential links remain, then the content must be the same. But we will have no reason to accept this on faith. So long as we can imagine alternative assignments of correctness conditions that preserve the inferential structure of the system, the inferentialist will not be able to fund the determinate obligations that are supposed to constitute meaning.

Some inferentialists argue that truth conditions and correctness conditions *can* be recovered by a system of rules.[36] However, even if this were possible, the inferentialist system would then encounter the very difficulties faced by rules that specify correctness conditions. It is implausible to think that a speaker ought to use expressions correctly all the time. Thus it seems that in general the two desiderata with which we began—that a rule be meaning-determining and prescriptive—cannot simultaneously be met. For a rule to be meaning-determining, it needs to be capable of specifying *all* the infinitude of correct applications of an expression; but it is unreasonable to suppose that anyone ought to apply an expression invariably correctly. Since an argument from rule-following to *Normativity* requires that some rules meet both desiderata, it seems as though such an argument cannot be made. There is thus no reason to think that the intuitive idea that meaning can be understood on analogy with rule-following should commit us to the idea that meaning must be constituted by a pattern of semantic obligations.

---

[36] Cf. Brandom 1994*a*.

The reasons why a rule cannot be both prescriptive and meaning-constituting are quite general, and will be familiar from the foregoing arguments. For a rule to be meaning-constituting, it must tell us what a particular term means. Assuming the intuitive, truth conditional picture of meaning, this is to say that for a rule to be meaning-constituting it must tell us the truth conditions of a sentence or reference conditions of the term. If a rule falls short of supplying the truth conditions of a term, as far as proponents of the intuitive view are concerned, it does not constitute the meaning of the term. Since *Normativity** was supposed to follow from the intuitive view, this constraint can hardly be ignored. Thus, a meaning-constituting rule must yield the full list of correct uses of a term and none, whatsoever, of the incorrect ones.

The trouble is that no prescriptive rule can supply the relevant truth conditions or correctness conditions. For a rule to be prescriptive, it must tell me what I ought to do. According to the intuitive view, the meaning-constituting rule for 'horse' must imply that 'horse' applies correctly to all and only horses. However, it is not the case that I ought to apply 'horse' to all and only horses—I am not obligated to apply 'horse' to all horses because I cannot do so, and 'ought' implies 'can'. The weaker rule, stating that I should apply 'horse' *only* to horses is no good, because it cannot constitute the meaning of 'horse'. The rule that tells me to apply 'horse' *only* to horses does not tell me that 'horse' applies to all horses, which, for the semantic realist, falls short of telling me what it means. In general, if a rule is to have even a remote chance of constituting meaning, it will have to imply the full list of correct applications of a term. But since 'ought' implies 'can', no such rule can also be prescriptive.[37]

CONCLUSION

I have looked at a number of arguments in favour of *Normativity*. In each case, I have claimed that the arguments fail to support the thesis that meaning is normative. Although there may yet be further arguments to consider, the argument with which I ended the chapter, although simple, shows clearly that there can be no rule that is both meaning-constituting, on a truth conditional picture of meaning, and categorically prescriptive. As a result, *Normativity* is indefensible as a condition of adequacy on theories of meaning which purport to reduce or otherwise make sense of what it is for a speaker to understand contents on the truth conditional picture of content. Thus, I conclude that we have very good reason to reject *Normativity*.

---

[37] For a more detailed discussion of this issue, see Glüer and Pagin 1999.

I have argued previously that *Normativity* is crucial to the sceptical argument; it allows the sceptic to generalize from the failure of a few theories to the failure of them all. If *Normativity* were tenable, the sceptic would have an argument against theories of meaning *a priori*; meaning would be on the wrong side of a yawning gap between *is* and *ought*, with no prospect of rescue. Given that *Normativity* is untenable as a constraint on theories of meaning, however, the sceptical argument turns out to be unsound. Thus, by abandoning *Normativity*, we can confute the sceptic, and thereby deprive the sceptical argument of its force. Ultimately, the sceptic does not give us any reason to hold that there is no fact of the matter what we mean.

This page intentionally left blank

8

Conclusion

The sceptic faces the daunting task of proving a negative: that if we assume semantic realism, the intuitive picture of meaning and content, there is no fact of the matter what anybody means by any word or thinks with any thought. Astonishingly, on one plausible interpretation of the sceptical argument, the sceptic succeeds. Given the assumption that meaning is normative, the sceptic is able to marshal *a priori* considerations against both reductive and anti-reductive accounts of the facts that constitute what we mean. This is because, if meaning is normative, then the concept of something's being semantically correct must have an internal connection with motivation—my belief that some use of an expression is correct must motivate me to make that use of the expression. If semantic correctness is understood to be normative, then, the concept of semantic correctness cannot be analytically defined in terms of any non-normative concepts—for, in virtue of being non-normative, those concepts will have no internal conceptual connection with motivation. Similarly, we will never find that semantic correctness is identical with some non-normative property, since the putative property of semantic correctness would have to be inherently action-guiding, whereas no non-normative properties would have any such effect. Furthermore, if meaning is normative, then semantic facts would have to be both objective and inherently action-guiding, so that the assumption that there are such facts commits us to a queer ontology. Granted that meaning is normative, then, the sceptical argument would be tremendously powerful; if it were cogent, we would have *a priori* grounds for the rejection of all possible facts that might constitute what I mean.

Devastating though this argument appears, it has a crucial weakness: the assumption that meaning is normative, which plays such a crucial role in the *a priori* argument against all possible theories of meaning, is untenable. *Normativity* is no part of the intuitive view—it follows from the assumption that understanding the meaning of a word is to follow a rule, and it relies on a controversial assumption of the nature of the rule that a speaker must be said to follow. What is more, *Normativity* seems to be false—we have no reason to suppose that a speaker who means *snow* by 'snow' is *ipso facto* motivated or obligated to apply 'snow' to all and only snow. Thus, it is more reasonable to assume *Norm-Relativity*, according to which a speaker who means *snow* by 'snow' is someone for whom 'snow' applies correctly to all and only snow. However, since *Norm-Relativity* gives rise to no *a priori* argument against all meaning theories, the sceptic cannot use this assumption to rule out theories that have yet to be formulated, or to rule out facts of which we currently have no knowledge.

Without an *a priori* argument against all possible facts that might constitute what I mean, the sceptic is left with the weaker conclusion that we do not now know of the facts that constitute what I mean. However, although we do not now know of a fact that will decide between the hypothesis that I mean *snow* by 'snow' and the hypothesis that I mean *schmow* by 'snow,' it does not follow that there can be no fact of the matter what I mean. The sceptic places no restrictions on the sorts of fact we are entitled to consider—he allows even that the fact that I mean something by a word is one accessible only to an omniscient God. Hence, the fact that constitutes what I mean might well be one of which we are not now aware. The sceptic is not justified in concluding that there is no fact of the matter simply because we do not know of one. Similarly, if all the facts of which we are aware point equally to suspect A as to suspect B, we are not licensed to conclude that there is no fact of the matter who committed the crime.

What is more, as I argued in Chapter 3, the 'no fact thesis' leads inevitably to incoherence. To assume that there is no fact of the matter which ascriptions of content are correct, the sceptic is unable to so much as legitimate an ascription of content to his own thesis. For, in order to say that it is legitimate to ascribe the content *that there is no fact of the matter what anybody means* to the sentence 'there is no fact of the matter what anybody means', the sceptic must presuppose that there is a fact of the matter what somebody means. Moreover, the sceptical solution fails to supply a suitable surrogate conception of meaning, since the notion of communal agreement presupposes representations with determinate content. Without assuming semantic realism at the outset, the semantic non-factualist cannot legitimate meaning ascriptions to sentences that comprise the conclusion of his argument, and the sentences leading up to it; he cannot claim that his sentences are true even in the weakest, deflationary sense. It turns out that the non-factualist conclusion of the sceptical argument is irremediably self-defeating. This gives us positive reason to endorse semantic non-factualism: if it is incoherent to maintain that there is no fact of the matter what we mean, then we must presume that there is a fact of the matter what we mean. Moreover, semantic realism is indispensable, since it must be presupposed in the sceptical solution.

Given that the sceptic can no longer reasonably argue that there is no fact of the matter what we mean, what is left of the sceptical argument, then, is the problem of under-determination—of finding the fact that constitutes someone's meaning *addition* rather than *quaddition* by 'plus'. However, since we have reason to think that there *is* a fact of the matter what we mean, we can continue to hope that some adequate account will emerge. Moreover, by abandoning *Normativity*, the sceptical problem can be reduced to one that is, if not more manageable, then at least more familiar. Fundamentally, the so-called 'sceptical' problem is really just the hoary old problem of *intentionality*—of developing a theory that tells us what some linguistic or conceptual token is *about*, what it *represents*. Of course, it will be difficult to develop a fully adequate theory of representation; the problem of intentionality has resisted the persistent efforts of numerous philosophers of tremendous acumen and ability. The arguments against some of the most compelling of those efforts given in Chapters 5 and 6 give some indication of the nature of the difficulties. What I have shown, however, is that we do not need to develop a fully adequate theory of representation in order to confute the sceptic. Though we may not yet have an adequate theory of representation, we have no reason to believe that none will be forthcoming. Indeed, the incoherence of the sceptical conclusion gives us some reason to believe that a fully adequate theory of intentionality *will* be forthcoming. The important thing, then, is to keep on trying.

This page intentionally left blank

Bibliography

Ayer, A. J. 1936. *Language, Truth and Logic*. Oxford: Basil Blackwell.

Baker, G. P. and Hacker, P. M. S. 1984. *Scepticism, Rules and Language*. Oxford: Basil Blackwell.

Belnap Jr., N. D., Camp Jr., J. L. and Grover, D. L. 1975. 'A Prosentential Theory of Truth', *Philosophical Studies*, 27: 73–125.

Bilgrami, A. 1993. 'Norms and Meaning', in R. Stoecker (ed.), *Reflecting Davidson*. Berlin: W. de Gruyter.

Bird, A. 1998. 'Dispositions and Antidotes', *Philosophical Quarterly*, 48: 227–34.

Blackburn, S. 1984. 'The Individual Strikes Back', *Synthese* 58: 281–301.

Blackburn, S. 2005. 'Success Semantics', in H. Lillehammer and D. H. Mellor (eds), *Ramsey's Legacy*. Oxford: Oxford University Press.

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

Boghossian, P. A. 1989. 'The Rule-Following Considerations', *Mind* 98: 507–49.

Boghossian P. A. 1990. 'The Status of Content', *Philosophical Review* 99: 157–84.

Boghossian, P. A. 1994. 'Sense, Reference and Rule-Following', *Philosophy and Phenomenological Research* 54: 139–44.

Boghossian, P. A. 2003. 'The Normativity of Content', *Philosophical Issues*, 13: 31–45.

Brandom, R. 1994*a*. *Making it Explicit*. Cambridge, Mass.: Harvard University Press.

Brandom, R. 1994*b*. 'Unsuccessful Semantics', *Analysis* 54: 175–8.

Brandom, R. 1997. 'Replies', *Philosophy and Phenomenological Research* 57: 191–207.

Burge, T. 1998. 'Individualism and the Mental', reprinted in P. Ludlow and N. Martin (eds), *Externalism and Self-Knowledge*, pp. 21–85. Stanford: CSLI Publications. Originally published 1979. In *Midwest Studies in Philosophy* 4: 73–122.

Chomsky, N. 1980. *Rules and Representations*. New York: Columbia University Press and Oxford: Basil Blackwell.

Coates, P. 1997. 'Meaning, Mistake and Miscalculation', *Minds and Machines* 7: 171–97.

Crane, T. and Mellor, D. H. 1990. 'There is No Question of Physicalism', *Mind* 99(394): 185–206.

Darwall, S., Gibbard, A. and Railton, P. 1992. 'Toward *Fin de siècle* Ethics: Some Trends', *The Philosophical Review*, 101(1): 115–89.

Davidson, D. 1980. *Essays on Actions and Events*. Oxford: Oxford University Press.

Davidson, D. 1984. *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

Davidson, D. 2001. *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press.

Devitt, M. and Sterelny, K. 1993. *Language and Reality: An Introduction to the Philosophy of Language*. Cambridge, Mass.: The MIT Press.

Dokic, J. and Engle, P. 2002. *Frank Ramsey: Truth and Success*. London: Routledge.

Dretske, F. 1986. 'Misrepresentation', in R. J. Bogdan (ed.), *Belief: Form, Content, and Function*, pp. 17–36. Oxford: Clarendon Press.

Dretske, F. 2000. *Perception, Knowledge and Belief: Selected Essays*. Cambridge: Cambridge University Press.

Dummett, M. 1978. *Truth and Other Enigmas*. Cambridge, Mass.: Harvard University Press.

Ebbs, G. 1997. *Rule-Following and Realism*. Cambridge, Mass.: Harvard University Press.

Feldman, F. 1978. *Introductory Ethics*. Englewood Cliffs, NJ: Prentice-Hall.

Field, H. 1994. 'Deflationist Views of Meaning and Content', *Mind* 103(411): 249–85.

Fodor, J. 1990. *A Theory of Content and Other Essays*. Cambridge, Mass.: The MIT Press.

Foot, P. 2002. *Virtues and Vices*. Oxford: Oxford University Press.

Frankena, W. 1939. 'The Naturalistic Fallacy', *Mind* 48(192): 464–77.

Genova, A. C. 1970. 'Institutional Facts and Brute Values', *Ethics*. 81(1): 36–54.

Gibbard, A. 2003. 'Thoughts and Norms', *Philosophical Issues*, 13: 83–98.

Gilbert, M. 1996. *Living Together: Rationality, Sociality and Obligation*. London: Rowman and Littlefield.

Glock, H.-J. 1996. 'Necessity and Normativity.' in H. Sluga, and D. G. Stern (eds), *The Cambridge Companion to Wittgenstein*, pp. 198–225. Cambridge: Cambridge University Press.

Glüer, K. 1999*a*. *Sprache und Regeln. Zur Normativität von Bedeutung*. Berlin: Akademie.

Glüer, K. 1999*b*. 'Sense and Prescriptivity.' *Acta Analytica* 14: 111–28.

Glüer, K. 2001. 'Dreams and Nightmares: Conventions, norms and meaning in Davidson's Philosophy of Language', in P. Kotatko, P. Pagin and G. Segal (eds), *Interpreting Davidson*. Stanford: CSLI Publications.

Glüer, K. 2002. 'Explizites und implizites Regelfolgen', in U. Baltzer and G. Schönrich (ed.), *Institutionen und Regelfolgen*, Paderborn: Mentis.

Glüer, K. and Pagin, P. 1999. 'Rules of Meaning and Practical Reasoning', *Synthese* 117: 207–27.

Godfrey-Smith, P. 1994. 'A Continuum of Semantic Optimism', in S. Stich and T. Warfield (eds), *Mental Representation*. Oxford: Blackwell.

Grice, P. 1989. *Studies in the Way of Words*. Cambridge, Mass.: Harvard University Press.

Grover, D. L. 1992. *A Prosentential Theory of Truth*. Princeton, NJ: Princeton University Press.

Hale, B. 1997. 'Rule-Following, Objectivity and Meaning', in B. Hale and C. Wright (eds), *A Companion to the Philosophy of Language*, pp. 369–96. Oxford: Blackwell.

Hare, R. M. 1952. *The Language of Morals*. Oxford: Oxford University Press.

Hare, R. M. 1969. 'The Promising Game', in W. D. Hudson (ed.), *The Is–Ought Question*. London: Macmillan.

Hattiangadi, A. 2002. 'Is Meaning Normative?' *Mind and Language*, 21(2): 220–40.

Hattiangadi, A. 2003. 'Making it Implicit: Brandom on Rule Following', *Philosophy and Phenomenological Research* 66: 419–31.

Haugeland, J. 1982. 'Heidegger on Being a Person', *Nous* 16: 15–26.

Haugeland, J. 1998. *Having Thought*. Cambridge, Mass.: Harvard University Press.

Heil, J. 1983. 'Believing what one Ought.' *The Journal of Philosophy* 80: 752–65.

Horwich, P. 1990*a*. *Truth*. Oxford: Clarendon Press.

Horwich, P. 1990*b*. 'Wittgenstein and Kripke on the Nature of Meaning', *Mind and Language* 5: 105–21.

Horwich, P. 1998. *Meaning*. Oxford: Clarendon Press.

Horwich, P. 2000. 'Norms of Truth and Meaning', in A. O'Hear (ed.), *Philosophy: The Good, the True and the Beautiful* (Supplementary Volume): 19–34.

Hudson, W. D. 1969. *The Is–Ought Question*. London: Macmillan.

Hume, D. [1888] 1968. *A Treatise of Human Nature*. L. A. Selby-Bigge (ed.), Oxford: Clarendon Press.

Jackson, F., Oppy, G. and Smith, M. 1994. 'Minimalism and Truth Aptness', *Mind* 103(411): 287–302.

Katz, J. 1990. *The Metaphysics of Meaning*. Cambridge, Mass.: The MIT Press.

Katz, J. 1994. 'Replies to Commentators', *Philosophy and Phenomenological Research* 54, 157–83.

Kitcher, P. and Stanford, K. 2000. 'Refining the Causal Theory of Reference', *Philosophical Studies* 97: 99–129.

Kripke, S. 1972. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.

Kripke, S. 1982. *Wittgenstein, Rules and Private Language*. Cambridge, Mass.: Harvard University Press.

Kusch, M. 2005. 'Fodor v Kripke: Semantic Dispositionalism, Idealization and *Ceteris Paribus* Clauses', *Analysis*, 65, 2 (286): 156–63.

Lance, M. N. and O'Leary Hawthorne, J. 1998. *The Grammar of Meaning: Normativity and Semantic Discourse*. Cambridge: Cambridge University Press.

Lewis, D. 1970. 'How to Define Theoretical Terms', *Journal of Philosophy* 67: 427–46.

Lewis, D. 1999. 'Psychophysical and Theoretical Identifications', *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

McDowell, J. 1993. 'Wittgenstein on Following a Rule', in A. W. Moore (ed.), *Meaning and Reference*. Oxford: Oxford University Press. Originally published 1984, *Synthese* 58: 325–63.

McDowell, J. 1994. *Mind and World*. Cambridge, Mass.: Harvard University Press.

McDowell, J. 1998. *Mind, Value and Reality*. Cambridge, Mass.: Harvard University Press.

Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin.

McGinn, C. 1984. *Wittgenstein on Meaning*. Oxford: Basil Blackwell.

Maddy, P. 1984. 'How the Causal Theorist Follows a Rule', *Midwest Studies in Philosophy* 9: 457–77.

Martin, C. B. 1994. 'Dispositions and Conditionals', *Philosophical Quarterly* 44: 1–8.

Martin, C. B. and Heil, J. 1998. 'Rules and Powers', *Philosophical Perspectives*, Vol. 12: *Language, Mind and Ontology*, 283–312.

Mellor, D. H. 1990. 'Telling the Truth', in D. H. Mellor (ed.), *Ways of Communicating*. Cambridge: Cambridge University Press, pp. 81–95.

Mellor, D. 2000. 'The Semantics and Ontology of Dispositions', *Mind* 109: 757–80.

Mill, J. S. 1904. *A System of Logic*, 8th edn, London: Longmans, Green & Co.

Millar, A. 2004. *Understanding People*. Oxford: Oxford University Press

Miller, A. 1998. *Philosophy of Language*. Montreal and Kingston: McGill-Queens University Press.

Miller, A. 2003. *An Introduction to Contemporary Metaethics*. Cambridge: Polity Press.

Millikan, R. A. 1984. *Language, Thought and Other Biological Categories*. Cambridge, Mass.: Bradford Books/The MIT Press.

Millikan, R. A. 1990. 'Truth, Rules, Hoverflies, and the Kripke–Wittgenstein Paradox', *The Philosophical Review* 90: 323–53.

Millikan, R. 1995. *White Queen Psychology and Other Essays for Alice*. Cambridge, Mass.: The MIT Press.

Moore, G. E. 1993. *Principia Ethica*. T. Baldwin, ed. Cambridge: Cambridge University Press.

Neander, K. 1995. 'Misrepresenting and Malfunctioning', *Philosophical Studies* 79: 109–41.

Papineau, D. 1987. *Reality and Representation*. Oxford: Basil Blackwell.

Papineau, D. 1999. 'Normativity and Judgement', *Aristotelian Society Supplementary Volume* 73: 17–43.

Pettit, P. 1990*a*. 'The Reality of Rule-Following', *Mind* 99: 1–21.

Pettit, P. 1990*b*. 'Affirming the Reality of Rule-Following', *Mind* 99: 433–9.

Putnam, H. 1975. *Mind, Language and Reality*. Cambridge: Cambridge University Press.

Putnam, H. 1983. 'Models and Reality', in *Realism and Reason: Philosophical Papers*. Vol. 3. Cambridge: Cambridge University Press.

Quine, W. V. O. 1953. 'Two Dogmas of Empiricism', *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

Quine, W. V. O. 1960. *Word & Object*. Cambridge, Mass.: MIT Press.

Quine, W. V. O. 1969. 'Ontological Relativity', in *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Quine, W. V. O. 1970. 'On the Reasons for Indeterminacy of Translation', *Journal of Philosophy* 67(6): 178–83.

Quine, W. V. O. 1987. 'Indeterminacy of Translation Again', *Journal of Philosophy*, 84(1): 5–10.

Quine, W. V. O. 1993. 'Mind and Verbal Dispositions', Reprinted in A. W. Moore, *Meaning and Reference*. Oxford: Oxford University Press.

Quine, W. V. O. 2001. 'Truth', reprinted in Michael Lynch (ed.), *The Nature of Truth*. Cambridge, Mass.: The MIT Press.

Ramsey, F. P. 1965. 'Theories', in *The Foundations of Mathematics*. Totowa, NJ: Littlefield, Adams and Company.

Read, R. 1995. 'The Unstatability of Kripkean Scepticisms', *Philosophical Papers* 24: 67–74.

Rosen, G. 1997. 'Who Makes the Rules around Here', *Philosophy and Phenomenological Research* 57: 163–71.

Rosenberg, J. 1997. 'Brandom's Making it Explicit: A First Encounter', *Philosophy and Phenomenological Research* 57: 181–9.

Ross, D. 1987. 'Prima Facie Duties', reprinted in C. Gowans (ed.), *Moral Dilemmas*. Oxford: Oxford University Press.

Searle, J. 1964. 'How to derive "ought" from "is" ', *Philosophical Review*, 73(1): 43–58.

Searle, J. 1970. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Soames, S. 1998*a*. 'Facts, Truth Conditions, and the Sceptical Solution to the Rule-Following Paradox', *Philosophical Perspectives*, Vol. 12: *Language, Mind and Ontology*, 313–48.

Soames, S. 1998*b*. 'Scepticism about Meaning: Indeterminacy, Normativity, and the Rule-Following Paradox', Ali A. Kazmi (ed.), *Canadian Journal of Philosophy: Meaning and Reference*. Supplementary Volume 23: 211–49. Calgary: University of Calgary Press.

Smith, M. 1994. *The Moral Problem*. Oxford: Blackwell.

Stampe, D. 1979. 'Towards a Causal Theory of Linguistic Representation', *Midwest Studies in Philosophy* 2: 81–102.

Teichmann, R. 1992. 'Whyte on the Individuation of Desires', *Analysis* 52: 103–7.

Tennant, N. 1997. *The Taming of the True*. Oxford: Oxford University Press.

Vogel Carey, T. 1975. 'How to Confuse Commitment with Obligation', *The Journal of Philosophy* 72(10): 276–84.

Whyte, J. T. 1990. 'Success Semantics', *Analysis* 50: 149–57.

Whyte, J. T. 1997. 'Success Again: Replies to Brandom and Godfrey-Smith', *Analysis* 57: 84–8.

Wikforss, Å. 2001. 'Semantic Normativity.' *Philosophical Studies*, 2: 203–26.

Wilson, G. 1994. 'Kripke on Wittgenstein and Normativity', *Midwest Studies in Philosophy*, Vol. 19: *Philosophical Naturalism*, 366–90.

Wilson, G. 1998. 'Semantic Realism and Kripke's Wittgenstein', *Philosophy and Phenomenological Research* 58: 99–123.

Wittgenstein, L. 1922. *Tractatus Logico-Philosophicus*. C. K. Ogden, trans. London: Routledge & Kegan Paul.

Wittgenstein, L. 1953. *Philosophical Investigations*. G. E. M. Anscombe, trans. New York: The Macmillan Company.

Wittgenstein, L. 1956. *Remarks on the Foundations of Mathematics*. Oxford: Basil Blackwell.

Wright, C. 1980. *Wittgenstein on the Foundations of Mathematics*. London: Duckworth.

Wright, C. 1984. 'Kripke's Account of the Argument Against Private Language', *The Journal of Philosophy* 81: 759–77.

Zalabardo, J. L. 1989. 'Rules, Communities and Judgements', *Crítica, Revista Hispanoamericana de Filosofia* 21: 33–58.

Zemach, E. 1994. 'Katz and Wittgenstein', *Philosophy and Phenomenological Research* 66(2): 419–31.

Index

Ayer, A. J. 39, 47–8

Baker, Gordon 90, 172–6
Biosemantics, *see* Teleosemantics
Blackburn, Simon 22–3, 51, 111, 116–120, 146, 180, 187,196
Brandom 84–5, 98, 122, 123, 161–8
Boghossian 3, 12, 15, 22, 25–7, 29, 35, 51, 53, 58, 60, 70–4, 111, 114, 135, 146, 152, 158, 161, 180
Burge, T. 144–5

Capacities 151–3
Correctness conditions 2, 3, 5, 7, 12, 13, 15, 17, 22, 34, 47, 52–4, 56–7, 61, 66–7, 75, 82–4, 88, 90–9, 118, 146–7, 155, 157, 162–5, 169, 171, 180, 184–6, 188, 198–9, 202–206, 209
Categorical and hypothetical obligations 190–202

Davidson, Donald 17, 53, 100–102, 139, 166, 179, 197
  Triangulation 101, 102
Deflationism 70–4
Dispositionalism:
  and problem of error 24, 25
  infinitude, objection from 22, 23
  and Ramsey sentences 26, 27
  and normativity 28
  Communal dispositionalism 29
Dispositions:
  Multi-track 23
  Dispositional realism 105–109
  Conditional analysis of 112–16
  Finkish 111–12
Dretske, F 128–132
Dummett, Michael 2, 191–2

Error, problem of 22, 24–6, 28–30, 106–9, 111, 114, 116–17, 119, 122, 132–4, 152–3, 141, 145–6, 185

Externalism, semantic:
  Epistemic 18, 141, 144

Field, Hartry 2, 81–84
Fodor, J. 131, 132–140
Functionalism 29, 120

Gilbert, Margaret 195
Glock, Hans-Johann 51, 202–3, 205, 179
Grice, H. P. 154

Hacker, P. M. S. 172–6
Hare, R. M. 46, 47, 203
Hawthorne, John 51, 194
Horwich, Paul 77, 192
Hume: 40, 41
  Hume's Law 40, 41, 43
Hypothetical means/end obligations, *see* instrumental obligations

Infinitude, problem of 14, 22, 24–5, 27, 106–7, 110, 116
Instrumental obligations 190–202
Internalism, semantic 18
Internal relations 172–6

Judgement dependence of intentions 66, 159–161

Katz, J. 153–158

Lance, Mark 51

Mackie, J. 38, 48, 49, 51, 183–4, 200
McDowell, J.: 168–172
  Contractual theory of understanding 197–201
McGinn, C. 141–3, 151–3
Mellor, Hugh 109–116
Mental Representations 14, 15
Millikan, R. G. 126–130
Motivational internalism 41–3
Moore, G. E. 43–5

Naturalism, semantic 5
Non-cognitivism, ethical 89
Non-factualism, semantic 66
Normativity:
  and dispositionalism 22, 35
  and Prescriptivity 6, 39, 48–50, 54–7, 59, 60, 62–3, 106, 202–3, 205–7
  of meaning 3, 5, 6, 8, 22, 35
  of morals 8, 28, 37–41, 43, 45–50, 57, 181–3, 187, 189–90, 200–202
Norm-relativity 7, 8, 53–8, 61
Norm of Truth 190–3

Open Question Argument 43–6
Kripke, Saul 3–8, 11–23, 25, 28–35, 37–40, 43, 45–8, 50–1, 53, 57–8, 61, 65–70, 73–6, 80–1, 85–9, 91–2, 95, 98–100, 102, 105–7, 109–112, 115–16, 126–8, 141–2, 145–6, 151–2, 155, 158–9, 161, 168, 172–4, 176–9, 181, 184–6, 190, 196–7

Platonism 33, 34, 153–158
Prima Facie obligations 188–90
Primitivism 32
Private Language Argument: 100
  Kripke's account of 3–4
Prosentential theory of truth 84
'Qua' problem 142–143

Quale 29
Quietism 171
Quine, W. V. O.: 4, 16–18, 34, 77–80, 97–8
  in comparison to Kripke 16–18, 34
  empiricism 16, 17
  indeterminacy of translation 4, 16, 17
  Truth 77–9
  Stimulus meaning 78, 97

Redundancy theory of truth 70, 75, 76
Response-dependence 132–40
Representational content 15, 21, 30, 66–7, 126, 130, 133, 136, 140
Robust conception of truth predicate 71–4
Rules: 31, 53–4, 56–7, 59, 62, 91, 127, 139, 162, 164–5, 173–5, 201–206, 180, 190–3
  Constitutive vs. prescriptive 54
Rule Following: 31, 53–4
  in response to skeptical challenge 31
  in skeptical argument 53, 54

Sceptical argument: 11–35, 48, 51, 61–4, 99, 104, 66–7, 71–3, 86, 90, 152, 159, 168, 170, 172, 177, 179, 182, 207, 209, 211
  as reductio 3
Sceptical challenge 2, 11
Sceptical solution 3, 6, 8
Semantic Realism 2, 5, 6–8, 12, 13, 15, 52, 53
Searle, John 54, 202–3
Smith, Michael 42, 46, 68, 70, 75, 88–91, 93–7, 99–102, 138, 149, 123
Success Semantics 120–125
Oughts: 6, 181–5, 188
  Categorical 6

Teleosemantics 125–132
Truth: 1–3, 5, 7, 190–3, 12, 13, 58, 63, 41–2, 45, 102–3, 66–7, 69–81, 83–6, 90, 92, 102–3, 121–4, 128, 169, 174, 177, 199, 200, 180, 184, 187–193, 202, 204–7
  and Meaning 63, 102–3
  deflationary theory of 72–3, 75–7, 83–6
  redundancy theory of 70, 75
  prosentential theory of 84
  minimalist theory of 70
  correspondence theory of 70
Truth conditions: 1, 2, 5, 12, 13, 66, 69, 75, 83, 85, 90, 92, 122–4, 202, 205, 190–1
  and semantic realism 1, 2, 13, 75, 76
  and meaning 12, 13, 69, 72, 81, 83–4, 90, 206

Under-determination 16, 17, 19

Wright, Crispin 20, 146, 147, 158–161, 3, 32–5, 51, 69, 71, 74, 92, 176, 179, 197, 198
Whyte, Jamie 120–125, 192

Wittgenstein: 2–5, 7, 13, 16–18, 53, 93, 98, 69, 86, 91, 124, 193, 196, 200
  Kripke's interpretation of 4–5, 66–8, 70, 73, 75–6, 80–1, 85–6
  Tractatus 75
  Philosophical Investigations 13, 91, 69, 171