Errata & Addenda

4 3  
...tensions.' —> 'prima facie tensions.'

11 2  
'possible' —> 'possibly distinct'.

11 fn1  
'Meanings of expressions, and contents of concepts and propositions are italicised (e.g., the meaning of 'dog' is dog; the content of the proposition [there's a dog] is there's a dog), the meanings of unspecified words or unspecified concepts are given as 'F', meanings of unspecified declarative sentences or contents of unspecified propositions as 'p' —> 'Meanings of expressions italicised (e.g., the meaning of 'dog' is dog), the meanings of unspecified words are given as 'F', meanings of unspecified declarative sentences 'p'.'

14 10  
...networked' —> '...networked. (Note, however, that my final proposal bears no resemblance to Wittgenstein's own theory—if he had a theory).'

15 5  
'Though they...' —> 'In so far as they analyse the general meaning-relation 'α means p', they...'

15 fn9  
'Cf. Harman ibid.' —> 'Cf. Harman ibid. Notice that conceptual role theories are mainly concerned with how concepts get particular meanings assigned, thus they give an account of how α comes to mean dog, rather than and, etc. But surely (the beginnings of) an account of the general meaning-relation 'α means p' is at least implicit in such an account (e.g., for α to have some meaning or other α must have some conceptual role).'
Horwich’s approach shares something with conceptual role semantics. They both give accounts of particular relational meaning-properties, such as \( \alpha \) means dog, \( \alpha \) means and, and so on. But they also, Horwich most explicitly, give accounts of the meaning-relation \( \alpha \) means \( p \).

...about meaning. (At least, being disposed to mouth of words doesn’t demonstrate the asymmetry between first- and third-person authority about meaning: I have just as good access to your utterances as you have yourself).

...meanings are robust and error is possible.

seem

impart

allow us to have

plativudes

tensions

...of the concept of meaning.' —> ‘of the concept of meaning.

It makes sense to talk about meanings of sentence types because by ‘sentence type’ I mean sentences whose meanings are determined by the patterns of use at a world, under the assumption that that world is the actual world. Such sentence types can be evaluated counterfactually, i.e. evaluated at other worlds under the assumption that their meaning is determined in the actual world. This is a non-standard
use of the term ‘sentence type’ in so far as indiscernible grammatical types may count as different sentence types on this approach, if they are interpreted differently at different worlds. I settle for this notion of sentence type because I wish to side-line the kind of context sensitivity associated with counterfactual truth evaluation (see also Ch. II.2.a.iii, and p. 119).'

21 6 'There are two ways [...] constitute semantic relations.' —> ‘To argue this point I must show how the analysis of meaning gives us reason to believe that for $q$ to mean $p$ is it necessary that $q$ possesses U, and by showing, that possession of U is also, perhaps with other conditions, sufficient for $q$ to mean $p$. In Ch. III I argue that some use-dependent property is necessary for meaning. In Ch. IV-VII I begin to establish the claim that, together with some background assumptions, a property of use-dependence is sufficient to constitute facts about meaning.’

23 5 ‘...the platitudes are arrived at a priori, if they are arrived at relying merely on non-empirical sources’ —> ‘however the platitudes are actually arrived at, our recognition of them as fixing the concept of meaning is a priori, i.e. relies merely on non-empirical sources,’

23 10 ‘actual’ —> ‘true’

24 -6 ‘... the definiens.’ —> ‘...the definiens. What if there are no things which the platitudes are true of? Imagine, e.g., that the platitudes we arrived at are false. Is there then no concept of $F$? Here we need to distinguish two kinds of case, corresponding to two projects one may be interested in. There is a clear sense in which there will still be a possible concept of $F$, even if the platitudes are false (think again of Euclidean geometry). So there is a sense in which being committed to some set of platitudes, true or false, is sufficient for instigating a meaning (We might say, in Horwich’s words, that ‘regarding them as true’ is sufficient for instigating a meaning (see Ch. VIII)). But, in another sense, it ought to make sense to ask: is this actually our concept of $F$? We may, that is, be
interested in our actual concept of $F$, rather than some possible concept $F$, instigated by some platitudes. On this approach we hope to have arrived at our actual platitudes—the platitudes which are truly ours—, irrespective of whether they are true or false. If we have not arrived at the actual platitudes there is a sense in which there is no concept defined by them, even though we could of course instigate a new concept by regarding the ‘non-actual’ set of platitudes as true (imagine we get the axioms constitutive of Euclidean geometry wrong, in that case we have failed to analyse the concepts of Euclidean geometry, even though there may some other quasi-geometrical concept that we have arrived at).'

24 1 'Each of them...' —> ‘To say that they are core is to say that each of them...’

25 11 ‘...is aware of such errors and...' —> ‘is aware of such errors about meaning and...’

25 -3 'In general, if I believe or desire that $p$, and if $\varphi$ means $p$, then I can use $\varphi$ to indicate the belief or desire that $p$ to someone.' —> ‘Normally, if $\varphi$ means $p$ and $S$ uses $\varphi$, then $S$ believes or desires that $p$’.

25 -10 ‘...platitudes.’ —> ‘...platitudes. What are the grounds for supposing that these five exhaust the platitudes? Naturally, any philosophical treatment of the platitudes is going to involve some structuring of what is recognised a priori as the platitudes. I think the above five capture most of the platitudinous things we hold true for the concept of meaning. It is perhaps a matter a book-keeping how they should be written out and whether some of them, e.g., the error and robustness platitude, should really be spelled out as several platitudes. Here are three candidates for platitudes that people may think are left out: (a) the meaning of a sentence is dependent on the words in it. (b) The meaning of a word has normative import concerning what it ought to be applied to. (c) the meaning of a declarative sentence, together with how the world is, determines whether or
not the sentence is true. I think (a) is captured under the communication platitude for surely a speaker can convey different information by substituting individuals words in a sentence so as to produce a new sentence; the way we put words together allow us to convey different things about how we take the world to be. If this is not what is meant by the communication platitude, then it is hard to see what it means. (b) emerges easily from the explanatory meaning and error and robustness platitudes for if we know that there can be errors in our use of a word, and that we can explain use of words by citing facts about meaning, then we have a way of saying whether some use of a word was indeed in accord with its correct use. (c) is relevant in so far as we have limited ourselves to declarative sentences (though notice my reasons for doing this (p. 21, fn. 1)). But anyway, this seems to me to be captured by the platitudes as well (in particular the communication, explanatory meaning, and error platitudes). Assume error is possible, and that we can tell what it is for a word to be applied incorrectly. Then we can see how the world may corporate such that, if the words are applied correctly, then we may transmit true beliefs in communication. True, people say that the meaning of a declarative sentence is its truth-conditions, but this concerns one of the semantic relata, the meanings, and I have nothing to say about them, apart from what it is for one of those things to be related to an expression (as I have stressed above p. 11-13).

28 -6 ‘used’ —> ‘normally used’.

30 5 fn after ‘them.’ : ‘(I) may seem trivial (if we assume that ‘correct use of φ’ is defined as ‘used to indicate a belief in the proposition φ means’). I address this worry below (sec. 2.a.iv). For now, notice that (I) is not trivial if you think meaning is independent of use. It still holds of course, that a supervenience theorist should not explain what the subvenient properties are in terms of the supervenient properties (cf. II.3.a), but there is no reason why, in the case of (I), the notion of correct use must be defined in terms of the notion of meaning (cf. Chs. V and VI for a solution).’
fn after 'platitudes.' : 'Do we already have a solution, viz. the thought that use\textsubscript{1} = 'what concept \( \varphi \) is used to indicate' and use\textsubscript{2} = 'the external circumstances in which \( \varphi \) is used'? This may be a solution, but only if we can make good explanatory sense of the idea that '\( \varphi \) was used in the external circumstances C \because \( \varphi \) is used to indicate the concept F'. It is not immediately clear that this is what is implied by the explanatory meaning platitude, or indeed, how these two sense of use can come apart (for more detail see Ch. 5).

Insert 'but:' between line 11 and line 12.

Insert 'so, by MTT on the conditional in (I):' between line 13 and line 14

'...a simple perception:' —> ‘...a simple, indeed platitudinous, perception:'

'...those practices too.' —> ‘...those practices too. If the revision is so weak it amounts to a mere reformulation of the platitude we can say that the tension in question was just a prima facie tension. It may be a vague issue where reformulation ends and genuine revision begins.’

'...sentence type.' —> ‘...sentence type. We may legitimately ask how this account may then be extended to incorporate context sensitive sentences since there may not be very many context invariant sentences. To this I say: the sentence types I have in mind are not context invariant in the sense that context could not influence their meaning. When I say 'sentence type' I mean 'sentence type as interpreted as \( w \) under the assumption that \( w \) is the actual world'. This means that my sentence types are more like propositional characters ('diagonal propositions' in Stalnaker's two-dimensional modal logic). That is, their content are arrived at by evaluating them at a world under the assumption that that world is the actual world. UD helps determine these propositional characters for it is the dispositions for use at each world which determines them. This is
of course highly context sensitive for it depends on the context of use at the world under consideration. In what sense does it then make sense for me to insist that I deal with sentence types? It makes sense for there is another important way of evaluating sentences: propositional characters may be evaluated counterfactually, at other possible worlds, under the assumption that their propositional character is fixed by their interpretation in the actual world. This is a way of saying how the same ('diagonal') sentence type may be evaluated in different contexts. The contribution of context to this kind of evaluation lies beyond the scope of this essay for it concerns how the facts determine truth value, given we know the meaning, i.e. the propositional character, of the expression. I deal only with the latter, i.e. with how the 'diagonal' propositional characters are determined.
Meaning as Use

JAKOB HOHWY

A thesis submitted for
the degree of Doctor of Philosophy
of the Australian National University

Philosophy Program
Research School of Social Sciences
Institute of Advanced Studies
October 1998
DECLARATION

This dissertation is my own work,
extcept where otherwise acknowledged

Jakob Hohwy
ACKNOWLEDGMENTS

I wish to thank my supervisory panel: Frank Jackson, Philip Pettit and Daniel Stoljar. They all commented on versions of this thesis in its various stages. Their impact has been enormous, and I am very grateful for their support and patience.

A generous scholarship from the Danish Research Academy enabled me to study for the Ph.D. at the ANU. I wish to thank the Academy, especially Birgit Christensen and Tinna Meyer for their professional and friendly assistance.

Many people have read parts of earlier drafts or have discussed the core issues with me. I especially wish to thank Lars Binderup, Simon Blackburn, Grant Gillett, Ian Gold, Lars Gundersen, Richard Holton, Adrienne Lehrer, Daniel Nolan, Huw Price, Michael Smith, Natalie Stoljar, Thomas S. Wentzer, and all my fellow students at the Philosophy Program at RSSS. Thanks also to Judith Sellers for great administrative support.

At the AAP Annual Conference in Auckland 1997 I read a version of Chapter 7. I wish to thank the people in the audience for valuable comments. I read a version of Chapter 3 at the Truth and Knowledge Conference in Aarhus where I benefited from valuable feedback from the audience, especially Paul Boghossian and Crispin Wright. In 1998 I read part of Chapter 8 at the Wittgenstein, Semantics and the Deflationary Turn Conference in Sydney, where I also received valuable feedback. I am very grateful to Paul Horwich for letting me read and work on the manuscript of his forthcoming book.

I wish to thank Linda Barclay for discussions, proof-reading, and support.
In this thesis I argue for a dispositional account of meaning. The main premise in my argument is that meaning is a cluster concept whose constituents stand in various tensions. My conclusion is that the tensions can only be resolved if we adopt a dispositional account of meaning.

I begin by showing that five uncontroversial platitudes, or core beliefs, are constitutive of the concept of meaning. These are platitudes concerning communication, meaning and use, explanatory meaning, error, and first-person authority. However, there are three tensions among the platitudes: i) the meaning is use problem, i.e. that meaning both explains use, and is explained by use; ii) the problem of error; and iii) the problem of the phenomenology of first-person authority (Chs. 1-2).

Even though the platitude about meaning and use is involved in all three tensions, we should not reject it. Without it, we can no longer make sense of the remaining platitudes (Ch. 3). This shows that there must be an aspect of use in an account of meaning. The particular account I endorse is this dispositional notion of ‘use-dependence’ (UD); the sentence type φ means p in a speaker’s language iff the speaker is disposed to use φ to indicate mental states with the content p. UD belongs to the family of response-dependent theories. I therefore discuss and amend the general notion of response-dependence (Ch. 4).

Given UD, the three tensions can be resolved. The meaning is use problem is resolved by appealing to the distinction between possession and manifestation of dispositions (Ch. 5). The problem of error is resolved by showing that if we adopt a sufficiently sophisticated notion of dispositions, then all versions of the problem fail (Ch. 6). The problem about first-person authority is resolved by showing that UD doesn’t entail behaviourism or social externalism in any problematic sense (Ch. 7).

Lastly, I consider Paul Horwich’s recent theory of meaning. He argues that a use-based account of meaning must be combined with deflationism about truth. I argue that the issue of truth can and should be factored out of the use-based account (Ch. 8).

Many philosophers find the “meaning is use” slogan either obscure, obviously false, vague, or platitudinous. My argument shows that if we embrace the slogan in its platitudinous form, together with the four other platitudes surrounding the concept, then we can indeed conceive of meaning as use, because that will allow us to resolve the three tensions among the platitudes. I therefore conclude that the concept of meaning is the concept of a use-dependent property.
## CONTENTS:

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DECLARATION</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>ACKNOWLEDGMENTS</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>TABLE OF CONTENTS</td>
<td>5</td>
</tr>
<tr>
<td>Chapter 1</td>
<td>INTRODUCTION</td>
<td>11</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>FIVE PLATITUDES ABOUT THE CONCEPT OF MEANING, AND THE TENSIONS AMONG THEM</td>
<td>20</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>INCOMPATIBILIST APPROACHES TO THE TENSIONS AMONG THE PLATITUDES: PRIMITIVISM AND CAUSAL THEORIES</td>
<td>66</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>TOWARDS COMPATIBILISM: USE-DEPENDENCE</td>
<td>93</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>RESOLVING THE FIRST TENSION: THE &quot;MEANING IS USE&quot; PROBLEM</td>
<td>128</td>
</tr>
<tr>
<td>Chapter 6</td>
<td>RESOLVING THE SECOND TENSION: ERROR AND ROBUSTNESS</td>
<td>153</td>
</tr>
<tr>
<td>Chapter 7</td>
<td>RESOLVING THE THIRD TENSION: FIRST-PERSON AUTHORITY</td>
<td>207</td>
</tr>
<tr>
<td>Chapter 8</td>
<td>MINIMALISM ABOUT MEANING, USE-BASED MEANING, AND DEFLATIONIST TRUTH</td>
<td>228</td>
</tr>
<tr>
<td></td>
<td>CONCLUSION</td>
<td>261</td>
</tr>
<tr>
<td></td>
<td>BIBLIOGRAPHY</td>
<td>265</td>
</tr>
</tbody>
</table>
Chapter 1  INTRODUCTION

1. Three types of meaning-theory.  
2. Semantical theories and the "meaning is use" slogan.  
3. The analysis of the concept of meaning.  
4. Overview.

Chapter 2  FIVE PLATITUDES ABOUT THE CONCEPT OF MEANING, AND THE TENSIONS AMONG THEM

1. Analysing the concept of meaning.  
a) The analysandum.  
b) Platitudes and analyses.  
2. The five platitudes about the concept of meaning.  
a) Communication.  
b) Meaning is use.  
c) Explanatory meaning.  
d) Error and robustness.  
e) First-person authority.  
f) Remarks on the platitudes.  
3. Three tensions among the platitudes.  
a) The first tension: the "meaning is use" problem.  
b) The second tension: the problem from irregular and infallible use.  
c) The third tension: the problem of behaviourist and social use.  
d) Networked platitudes.  
4. Strategies for resolving the tensions.  
5. Summary remarks.

Chapter 3  INCOMPATIBILIST APPROACHES TO THE TENSIONS AMONG THE PLATITUDES: PRIMITIVISM AND CAUSAL THEORIES

1. Introduction.  
a) Against incompatibilism.  
b) Primitivism and causal theories.  
2. Contingent Primitivism.
a) A Moorean parallel. 69
b) A Lockean parallel. 72

3. Back to basic primitivism. 75
   a) Paring down the doctrine. 75
   b) The problem for basic primitivism. 75

4. Essential Primitivism. 77
   a) Problems for primitivism with essential relations. 77
   b) Twin-Earth cases and externalist content. 78
   c) Primitive informational content - the disjunctive view. 78
   d) Conceptual role semantics. 80
   e) Individualised content primitivism. 81
   f) Supervenience Primitivism. 82

5. Causal theories. 84
   a) Primitivism and causal theories. 84
   b) Pure causal theories. 84
   c) Problems for the causal theory. 86
   d) The *qua*-problem and the possibility of error. 89
   e) Further remarks. 91

6. Conclusion. 92

---

**Chapter 4  TOWARDS COMPATIBILISM: USE-DEPENDENCE** 93

1. Introduction. 93
2. Introducing the use-dependence of meaning. 94
3. Use-dependence as a response-dependent notion. 96
   a) What is response-dependence? 96
   b) Global or local response-dependence. 99
   c) Successful response-dependence: four constraints. 100
   d) Dispositions and response-dependence. 110
   e) What type of dispositional responses? 115
4. Refining and restating the theory and conception of use-dependence. 118
   a) Refining and restating the theory of use-dependence. 118
   b) Clarifying the UD-based conception of meaning. 121
   c) Remarks on the UD-based conception of meaning. 125
5. Concluding remarks. 126
Chapter 5  RESOLVING THE FIRST TENSION:
THE "MEANING IS USE" PROBLEM  128

1. Use-dependence, explanatory meaning
and the "meaning is use" problem.  128

2. Getting the phenomenology right: varieties of explanation.
   a) The type of explanation.  132
   b) What is getting explained by meaning?  135
   c) What is doing the explaining of use?  137

3. The Missing Explanation Argument.
   a) Response-dependence and MEA.  142
   b) Critical assessment of MEA.  144

4. Conclusion.  151

Chapter 6  RESOLVING THE SECOND TENSION:
ERROR AND ROBUSTNESS  153

1. Robust disagreement about error, and crass conventionalism.  154
   a) Crass conventionalism.  155
   b) Dispositions are not necessarily irregular.  157
   c) Communication and dispositionalism.  158
   d) Crass conventionalism and U-ness.  159

   a) Generating the objection.  164
   c) Assessing the problem about unity.  166
   d) An answer to the problem about unity: joking and daring.  167
   e) Substantiality and uniformity in the
      specification of C-conditions.  171
   f) Two further versions of
      the problem from belief-desire holism.  173

3. Dispositionalism, error, and the problem of infallible use.  176
   a) Error and the importance of error.  177
   b) Dispositions and errors.  178
   c) First alternative version of the problem of error:
      an objection about infallible use.  180
   d) Second alternative version of the problem of error:
      guaranteed correctness.  181
   e) Third alternative version of the problem of error:
      being disposed to make errors.  185
4. On C-conditions.
   a) Interdefinitions of C-conditions for use-dependence. 192
   b) How meaning-constituting dispositions are set apart from other dispositions. 197
5. Restating UD - concluding remarks. 205

Chapter 7  RESOLVING THE THIRD TENSION: FIRST-PERSON AUTHORITY 207

1. Introduction. 207
2. Use-dependence is not behaviourist in any problematic sense. 209
   a) Which types of behaviourism are problematic for first-person authority? 210
   b) UD doesn’t entail the problematic kind of behaviourism that threatens the first-person perspective. 210
   c) UD doesn’t entail the kind of behaviourism which is a threat to the a priori privilege of first-person authority. 211
   d) Wherein does first-person authority consist? 212
3. UD is consistent with a plausible social externalism. 214
   a) UD and social externalism. 215
   b) UD-based social externalism and first-person authority. 217
4. First-person authority and linguistic insiders. 219
   a) A worry about diachronic first-person authority. 220
   b) The threat to diachronic first-person authority is a threat to first-person authority as such. 222
   c) Elusive authority about language. 223
   d) Comments on the idea of elusive a priori evidence for first-person authority. 226
5. Conclusion. 226

Chapter 8  MINIMALISM ABOUT MEANING, USE-BASED MEANING, AND DEFLATIONIST TRUTH 228

1. Introduction - minimalism and metaphysically infected concepts. 228
   a) An example of minimalism from the debate about truth. 229
   b) Minimalism as a general strategy. 230
2. ‘Minimal’ minimalism. 231
3. How minimalism relies on use. 233
   a) Horwich's project. 235
   b) A central problem for the traditional approach. 236
   c) Deflationism and meaning. 238
   d) What are the meaning-constituting properties and how do we identify them? 239

5. Basic regularities, functionalism, and permutation.
   a) Diverse *explananda*. 241
   b) Functionalism? 244
   c) Are the *explanantia* too similar? 246
   d) Permutation and relationality: some general remarks. 249

6. Explanations of meaning and explanations of content.
   a) An account of meaning *justified* by deflationism about truth. 252
   b) Explaining aboutness. 253
   c) Meaning and content. 255
   d) Two objections and a final worry. 257

7. Summary remarks. 259

CONCLUSION 261

BIBLIOGRAPHY 265
INTRODUCTION:
TYPES OF MEANING-THEORY,
AND OVERVIEW

1. Three types of meaning-theory.

In this essay I focus on the concept of meaning. In general we can discern three possible philosophical approaches to the concept of meaning: an ontological, an epistemological and a semantical approach. I shall give a brief characterisation of each and explain why I prefer the semantical approach.

An ontological theory of meaning is focused on the nature of meanings. Thus, the ontological meaning-theorist offers a theory about the nature of the relatum on the right hand side of relational expressions of the form \( \alpha \text{ means } p \) where \( \alpha \) is some expression type and \( p \) its meaning. According to the ontological theorist, meanings may be, e.g., extensions, concepts, or propositions.\(^1\) For example, if we assume that

\(^1\)I shall adopt the following conventions: single quotes are used for mentioning expressions (words and sentences); double quotes for citations; corner quotes for mentioning contexts of unspecified expressions, Greek letters will be names of unspecified expressions (in particular, \( \varphi \), \( \psi \), and \( \xi \) for unspecified declarative sentences). Concepts and propositions are referred to in square brackets (e.g. the concept \([\text{meaning}]\); the proposition \([\text{there's a dog}]\), except where it is clear from the context and I write, e.g., 'the concept of meaning' or 'the proposition that dogs are ugly'; unspecified concepts are \( F \) and unspecified propositions \( p \). Meanings of expressions, and contents of concepts and propositions are italicised (e.g., the meaning of 'dog' is \textit{dog}; the content of the proposition \([\text{there's a dog}]\) is \textit{there's a dog}), the meanings of unspecified words or unspecified concepts are given as \( F \), meanings of unspecified declarative sentences or contents of unspecified propositions as \( p \).
the word 'platypus' means *platypus*, then the ontological theorist offers a theory of what the meaning *platypus* is by saying, e.g., "it's the set of all possible platypuses", or "it's the abstract object *platypus*" etc.

Epistemological meaning-theorists are primarily concerned with knowledge of meaning, that is, they ask "given 'dog' means *dog*, how can speakers of a language L know this, on what basis can they have this knowledge, and what is it for them to have it?" The epistemological meaning-theorist might be concerned with how we can know meanings given the sparsity of linguistic input in our childhood learning, or with the implicit or explicit nature of our knowledge of meaning.

Finally, a meaning-theorist may adopt a semantical approach and ask "what is the nature of the semantic relation in expressions of the type 'α means *p*'?" A meaning-theorist of this ilk can take the semantic relata for granted and concentrate on what has to be the case for semantic relations between the relata to occur. The semantical theorist can choose between a general and a more particularist approach: either the focus is on semantic relations as such, or on the question "what makes it the case that the expression 'F' has the specific meaning *F* rather than any other meaning?"

These approaches are potentially distinct: one may consciously adopt one or the other. Of course, this is not to say they cannot overlap and cross-fertilise: views about the ontology and epistemology of meaning are apt to crop up in one's approach to the semantical question, and, likewise, views about the nature of semantic relations are apt to play a role in one's approach to the ontological and epistemological questions.

My approach in this essay will be semantical. I am interested in the semantic relations between expression types and meanings. Through an analysis of the concept of meaning I shall set out what is necessary for expression types to have meanings assigned, and I shall show that a property of dispositional use is sufficient to assign
particular meanings to particular expression types. The expression types I focus on are declarative sentences $\varphi$ whose meanings are propositions, $p$. That is, I am interested in the concept of meaning as given by the relation expressed by "$\varphi$ means $p$". Notice that though I define the relata as declarative sentence types and propositions respectively, I am not committed to any ontology or epistemology of meaning. A semantical meaning-theorist can, at least at the outset, afford to be agnostic about those questions.

2. Semantical meaning-theories and the "meaning is use" slogan.

The reason I focus on the semantical theory is that I think it provides the least prejudiced and potentially most fruitful way to begin to understand the slogan that "meaning is use". Making sense of the slogan by adopting an ontological theory seems counter-intuitive: we do not normally say that meanings are uses. Making the epistemology of meaning central likewise seems wrong: there is nothing inherently epistemological about the slogan as it stands. This is not to say that ontological and epistemological matters play no role in an elucidation of the "meaning is use" slogan. Rather, my claim is that the slogan first and foremost is concerned with the semantic relations between expressions and meanings, that is, it is concerned with the question "what is it about the use of some expression that can make it the case that it is related to some meaning?" Of course I will have to be committed to a conceptual analysis of the concept of meaning to begin to answer that question, but I do not have to be committed to any specific ontological or epistemological theories.

The "meaning is use" slogan originates in Wittgenstein's works, in particular in the *Philosophical Investigations* (1953). But, in some form or other, it appears in much of post-war philosophy of language. Unfortunately, the slogan is often alluded to

---

2The *locus classicus* is Wittgenstein 1953, §43.
3Cf. e.g. the wide range of papers in Margalit (ed.) 1979.
with some apprehension—indeed many people find it obscure, vague or trivial.\textsuperscript{4} A pretty representative view could be Gregory McCulloch:

Wittgenstein said that ‘if we had to name anything which is the life of the sign, we should have to say that it was its use’ (Wittgenstein 1958, p. 4). This can be made to mean all sorts of things, but, interpreting it as innocently as possible, we end up with something of a platitude.\textsuperscript{5}

I think this is exactly right. The slogan is indeed platitudinous. Importantly however, it is not the only core platitude that is constitutive of the concept of meaning, and, moreover, we get interesting results when we look at how the set of platitudes are networked.

I shall focus on the slogan, not as it appears in Wittgenstein, but more as it has entered mainstream philosophy of language in this more or less platitudinous form. My argument is that the slogan can be defended in a specific and substantial way—when set in the context of a semantical theory, together with other platitudes about meaning, and when read dispositionally.

Versions of the “meaning is use” slogan appear in positions as diverse as W. v. O. Quine’s translationism, Michael Dummett’s justificationism, conceptual and inferential role semantics, Paul Grice’s intentionalism, David Lewis’ conventionalism, and more recently Paul Horwich’s deflationism. Their differences notwithstanding, it is possible to think of each of these positions as primarily concerned with the semantical question, which then potentially leads to lessons about ontological and epistemological issues.

\textsuperscript{4}For some random but not-so-hostile examples see Edwards 1995, p. 90: ‘[verificationism] would give substance to the otherwise vague slogan ‘meaning is use’”; Bilgrami 1986, p. 101: ‘The ideas and ideologies summarized in the dictum “meaning is use” are notoriously various and vague”; Stroud 1990, p. 28: ‘The answer [that meaning is use] does not tell us what the meaning of any particular word is, or how the meaning of one word differs from the meaning of another’, and Dummett 1978, p. 94: ‘The weakness of Wittgenstein’s dictum lies in its extreme generality’. Fogelin 1987, p. 122 suggests that Wittgenstein has no theoretically apt view of how use and meaning are related.

\textsuperscript{5}McCulloch 1989, p. 165f.
Though Quine has an ontological theory of stimulus meanings, his main concern is with how expressions get meanings assigned in virtue of use (and for Quine 'use' is "dispositions to gross overt behaviour").

Conceptual role semantics and inferential role semantics "represent one thing that might be meant by the slogan 'meaning is use'". Though they sometimes can appear to be ontological theories, they are best interpreted as concerned with the semantical relations between concepts and contents.

Though Dummett has a very strong epistemological bias he is not lead by an interest in epistemology, but by the conviction that sentences get their meanings (i.e. the truth-conditions) assigned in virtue of the manifestable use of their subsentential components, and he is interested in what this tells us about the concept of truth.

Grice and Lewis, as well as Horwich, have adopted positions which are more obviously semantical. According to Grice, language is used to express one's thoughts in communication, such that the meaning of an expression is given by the thought it is used to express. Grice gives an intentional account of how such use constitutes semantic relations for speaker's meaning, and, further downstream, linguistic meaning. Lewis explicitly presupposes possible languages, i.e. possible pairings of expression-types with meanings, and then asks: "in virtue of what is a population speakers of one rather than another of these possible languages?" His answer is:

---

6Quine 1975, p. 87; cf. also Quine 1979, p. 1: ‘What then, is the semanticist to study and analyse, if not the meanings of words? the use of words’, as well as Quine 1960.

7Harman 1982, p. 242. See also the statements of CRS in Block 1986 and Field 1987. There is a nice early statement in Sellars 1950, p. 301-2: ‘the linguistic meaning of a word is entirely constituted by the rules for its use...[S]ymbols [are] governed by a system of rules which implicitly define these symbols by giving them a specific task to perform in the linguistic economy’.

8 Cf. e.g. Brandom 1994, p. 618. I believe the reason for their ontological appearance is that conceptual/inferential role semantics claim that the relations between meaning-bearers and meanings are essential: once the meanings change the meaning-bearer will change.

9 Cf. Harman ibid.

10 Cf. e.g. Dummett 1993, p. 115-6, in a remark on Quine's 1975 quoted above: ‘This is in full consonance with what I have myself repeatedly insisted on, that a meaning-theory, being a theoretical representation of a practical ability, must not only say what a speaker must know in order to know the language, but in what his knowledge consists, that is, what constitutes a manifestation of it’. Cf. also Dummett 1987, p. 117; 1991a, p. 93ff; 1976, p. 216ff; 1993, p. 187; 1978, p. 221. It is also interesting to note Dummett's close affiliation with verificationism evidenced in the complaint that the verificationists' only real error was that they treated 'all synthetic statements as standing on the same level, that is, [they ignored] the articulated structure of language', 1991b, p. 85 (for a general discussion of verificationism and the "meaning is use" slogan cf. Skorupski 1997).

viro of conventional use. Recently, Horwich has proposed a use-based semantical theory of meaning which is similar to, and partly justified by, deflationism about truth.

These semantical use-based theories of meaning agree on the idea that a property of use of expressions is what determines which meaning that expression gets assigned, and they then go on to exploit this in different ways, depending largely on which property of use they find fit for the concept of meaning. From this perspective the property of use one settles on may have quite wide philosophical ramifications. I am sympathetic to this methodology, I wish however to approach the problematic through an analysis of the concept of meaning which is more inclusive with respect to the phenomenology associated with the concept of meaning than is usual among the use-based accounts.

3. The analysis of the concept of meaning.

An analysis of the concept of meaning must respect the phenomenology of meaning, as represented in the core beliefs—the platitudes—we have about the concept of meaning, otherwise we end up with an *analysans* which cannot explain how the concept fulfils the function we care about it fulfilling. Here, I think, previous use-based accounts of meaning have not been sufficiently inclusive.

Quine's scientistic anti-mentalism does not leave space for the platitude that we have some kind of privileged first-person authority about meaning. Conceptual and

---

12 A *language* L is a function that assigns truth conditions to certain verbal expressions, called the sentences of L. [...] We [...] use L iff, by convention, we are truthful and trusting in L. That [...] is how use determines meaning'; Lewis 1992, p. 106. Cf. also Lewis 1972a, 1972b. The core of Grice and Lewis' account, that our words express our thoughts, can be found as far back as Aristotle and Locke (cf. Aristotle 1963, 16a3-16a26, 19a3; and Locke 1690, Bk. III, 2.i; Bk. III, 2iv-v. Cf. also Guyer 1994. For an overview of the tradition, cf. Avramides 1997.

13 Horwich, Ms.
inferential role semantics have serious problems accounting for the platitude that without possibility of error, meanings are not sufficiently robust. Dummett’s justificationism focuses on the normativity of meaning at the expense of the platitude that meaning in some natural sense is use; it also focuses on the idea that communication is essentially public, at the cost of the idea that our words express our thoughts. Grice and Lewis focus on the idea that language is used for communication but then both underplay the platitude that meanings are normative, that is, that error is possible, and they also neglect the idea that sometimes facts about meaning can explain facts about use in a robust manner.¹⁴ I am more enthusiastic about Horwich’s theory and devote Ch. 8 to a discussion of it.

These alleged shortcomings point us to the features of a properly inclusive analysis of the concept of meaning. It must make sense of the platitudes that we have first-person authority about meaning, that we can have disagreements about meaning and make occasional errors about meaning, that meaning somehow is (or supervenes on) use, and that facts about meanings can explain facts about use.

Unfortunately such an inclusive analysis is difficult to hold together. There seems to be counterexamples to the idea that meaning supervenes on use, and moreover there are three internal tensions among the platitudes i) how can meaning supervene on use and explain use? ii) how can our own use of words incur fallibility and robustness on facts about the meanings of those words? and iii) how can the apparent social externalism and behaviourism associated with use provide us with first-person authority about meaning?

¹⁴The Grice-Lewis approach has also been attacked in other ways, notably by Schiffer 1987, Ch. 9. In this thesis, which is sympathetic to the overall Lewisian framework, I do not address these arguments directly. I agree with some of Schiffer’s points (e.g. that understanding isn’t a matter of intellectualised inference, p. 261), but disagree with his general suspicion about propositional attitudes and compositional semantics.
The counterexamples and the three internal tensions among the platitudes must be addressed in order to get an inclusive analysis of meaning as use off the ground. The overarching aim of this exercise is not, then, to provide a scientifically respectable account of meaning, or an account of meaning which throws light on the concept of truth, or which explains how the purpose of language is communication, etc. Rather it is to provide an analysis of the concept of meaning which explains how and in what sense the platitudes constitutive of the concept can all be true.

4. Overview.

The essay has the following plan. In Ch. 2 I analyse the concept of the relation expressed by ‘φ means p’ by motivating the five platitudes that are constitutive of the concept. Here I also discern the three tensions among the platitudes. In response to the platitudes one can either adopt incompatibilism (i.e. claim that the platitudes are incompatible and that one or more of them therefore must be rejected), or adopt compatibilism (i.e. claim that, when viewed in the right light, the platitudes—or close versions of them—are compatible after all). I critically assess the most economical incompatibilism in Ch. 3. This is the doctrine that rejects the platitude that meaning is use (the ensuing position I call ‘primitivism about meaning’). Rejecting this platitude however makes it difficult to make sense of the remaining platitudes. I conclude that this type of incompatibilism doesn’t work. I therefore turn, in Ch. 4, to the prospects of a compatibilist approach to the platitudes. I suggest that we should think of the concept of meaning as the concept of a use-dependent property. This, I argue, is a version of the broader class of concepts of so-called response-dependent properties. I discuss this notion in some detail, and on this basis I formulate my idea for use-dependence. By doing this I invite the reader to begin to think of the concept of meaning in dispositional terms.
In Chs. 5-7 I demonstrate how the three tensions can all be resolved, given the doctrine of use-dependence. The core claim of these chapters is simply this: the arguments that purport to lead to the tensions are arguments that either overlook the dispositional nature of use-dependence, or they are arguments that work not only against meaning-dispositionalism, but against the very notion of a disposition, and which would therefore deprive us of the right to talk of such paradigmatic dispositional properties as fragility and water-solubility. In Ch. 5 I argue that, given use-dependence, we can see both how use explains meaning, and how meanings can explain uses. In Ch. 6 I show how meanings can be considered robust, so that we can disagree purposefully about meaning, and I demonstrate how a dispositionalist doctrine like use-dependence can explain the possibility of error while keeping faith with the other aims of this kind of analysis. In Ch. 7 I argue that use-dependence gives us a plausible idea of wherein the defeasible, yet a priori, nature of first-person authority about meaning consists.

The last chapter, Ch. 8, is concerned with an alternative suggestion for compatibilism, viz. minimalism, or deflationism about meaning. This is the kind of view put forward by, for example, Mark Johnston. I argue that such views must be use-dependent, and that therefore it is hard to see in what sense they are minimal. I give special attention to Horwich’s theory which is a use-based theory similar to, and justified by, deflationism about truth. I argue that the core of such a theory in principle can survive a divorce from deflationism about truth.

I conclude that, given the platitude-based analysis of the concept of meaning, we can coherently view meaning as use, that is, view the concept of the relation expressed by ‘φ means p’ as the concept of a use-dependent property.
1. Analysing the concept of meaning.

a) The analysandum.

In what follows I shall analyse of the concept of meaning. By 'analysis' I mean an explication, on *a priori* grounds, of the content of our ordinary concept of meaning. The aim is to obtain a conception of which type of property we talk about when we talk about meaning. That is, I wish to obtain a conception of which kind of relational property we talk about when we say, for example, "the sentence 'there is a wombat' means *there is a wombat*". In general, where 'φ' is a declarative sentence type and 'p' is a proposition, I am interested in the relation expressed by:

"φ means p",

in so far as that relation is semantical, that is, in so far as it satisfies the meaning-platitudes I shall set out below. Notice that I shall have little to say on the nature of the relata φ and p. This indicates that the analysis is not of meanings themselves, but
rather of the concept of the relation expressed by \( \forall \phi \text{ means } p \), and it is the platitudes surrounding this concept which take centre stage below.\(^1\)

After setting out the platitude-based analysis of the concept of the relation expressed by \( \forall \phi \text{ means } p \) I shall argue that talk of that concept is talk of the concept of a use-property, that is, that a property of use, U, is what constitutes the semantical relations between declarative sentence types and propositions. There are two ways this argument can proceed: either by showing that for \( \phi \) to mean \( p \) it is necessary that \( \phi \) possesses U; or by showing that possessing U is sufficient for \( \phi \) to mean \( p \). After the analysis in this chapter I try to establish the first claim. In chapter 3 I argue that no use-independent property can constitute semantic relations, but then, in Chapters 4-7, I go on to argue that a particular use-property (viz. a subject's correct dispositional use) is indeed sufficient to constitute semantic relations.

b) Platitudes and analyses.

How should we begin to talk about the concept of meaning? It seems natural, and itself in keeping with the idea that meaning is use, to focus on the things we implicitly or explicitly hold true and believe about the concept of meaning. Among these beliefs we should focus on the beliefs that are central to how we conceive of meaning, that is, the core beliefs—the platitudes—which are such that, were they false, then we would no longer recognise the concept as our concept of meaning. In general, talk about a concept F is, in Mark Johnston's words:

\(^1\)I settle for declarative sentences for the following reason. Block 1998, §2, employs a distinction between linguistic and metaphysical semantics. I am interested in metaphysical semantics, i.e. in 'investigating the nature of meaning, especially what it is about a person that gives his words or thought whatever meaning they have in the first place' (notice that Block explains it in relational terms: it is about how speakers' words or thoughts are related to meanings). The contrast is with linguistic semantics which 'deals with the meanings of particular expressions in particular languages and how they fit together to make up meanings of larger expressions, for example, how the meaning of 'John hopes angels exist' is related to the meaning of [the subsentential expression] 'angels exist'’. Block goes on to say that 'linguistic and metaphysical theories are largely independent' (ibid.). This implies that one can afford to deal with the metaphysics of meaning without dealing with the question of compositionality. Therefore it shouldn't, prima facie, matter whether one settles for sentences or subsentential expressions and I have settled on declarative sentences for reasons of simplicity, and I believe the analysis quite easily can be extended to non-declarative sentences.
...just a way of talking about the core of a conception or cluster of beliefs *de dicto* about F.\(^2\)

In our case, talk of the concept of meaning is talk of the platitudes about the relation expressed by "φ means \(p\). The idea is that setting out the concept in platitudinous terms will allow us to see that talk of the concept of the relation expressed by "φ means \(p\) is talk of the concept of a particular property, in this case the concept of a use-dependent property.\(^3\)

This style of analysis, then, comes in two stages: a conceptual and a constitutive stage. The conceptual stage consists in presenting the *a priori* platitudes about meaning. The constitutive stage consists in pointing out that the concept of meaning is the concept of a use-dependent property, in such a manner that it can be seen how this property could make the platitudes about meaning come out true.

The purpose of the analysis is not to provide an interesting analytic reduction of the concept of meaning in completely non-semantic terms (indeed I shall draw heavily on a notion of the contents of mental states). I do not know whether such reductive analysis is possible, or even desirable. Rather, the purpose of the present analysis is to show how, and to what extent, human use must be involved in determining facts about semantic relations, given what we think is platitudinous about the concept of meaning. That is, it is an attempt at setting out, in a substantial and non-trivial manner, what is involved in adhering to a version of the "meaning is use" slogan. Whereas this does not result in a reduction to non-semantic terms, it is a non-circular and non-regressive analysis of the concept of the relation expressed by "φ means \(p\)."

---


\(^3\)I take my cue from Johnston 1991 and 1993, but this general style of analysis has been developed by Frank Jackson, 1998, in particular, building on work by Ramsey, Carnap and Lewis (see e.g. Lewis 1970, 1972c). For examples of applications of the analysis see, for colour discourse, Johnston 1991, and, for moral discourse, Jackson & Pettit 1995.
In this chapter I am primarily concerned with the first step: presenting the platitudes surrounding our concept of meaning and discussing their interrelations. This must be done in an *a priori* manner, in order to merit the label of philosophical conceptual analysis, rather than the label of empirical investigation. The notion of aprioricity I have in mind is this: the platitudes are arrived at *a priori*, if they are arrived at relying merely on non-empirical sources, such as, *e.g.*, knowledge of the correct use of relational expressions of the form ‘φ means p’. This does not mean, of course, that the suggested platitudes necessarily will be the true platitudes: the fact that the justification of a claim is *a priori* does not show that the claim is necessarily true.\(^4\) It is possible, in other words, to arrive *a priori* at a set of platitudes which are not the actual platitudes for the concept in question.

At a fairly intuitive level the idea that talking about a concept is talking about the core platitudes associated with the concept, can be associated with a familiar notion of implicit definition. For example, the concepts of point and line are implicitly defined by the axioms of Euclidean geometry. Were those axioms different the two concepts would have different contents. An implicitly defined concept itself is used in the implicitly defining axioms which participants in the discourse regard as true. When doing geometry we do not question the axioms, but we may do so when we discuss the content of the terms of geometry. (The fact that we were fallible about Euclidean geometry goes to show that *a priori* justification may be wrong—it doesn’t show that the concepts of point and line had no content).\(^5\)

Whereas this is fairly intuitive in the case of highly theoretical concepts like point and line, it is perhaps less clear how it could work in general, and for less theoretical concepts like the one expressed by ‘φ means p’. From the point of view of a general

\(^4\)Kitcher 1990, p. 15-16, has a good account of distinctions among the *a priori*. The above distinction would correspond to her *a prioriQ* and *a prioriK*.

\(^5\)Cf. Horwich 1997b for a discussion of meaning and implicit definition.
common-sense or analytical functionalism, geometry is but a special, albeit clear-cut, case. In general we can talk about the contents of concepts by talking about the core platitudes of ordinary competent participants in the relevant discourse, because this is a way of giving that concept's functional role. This might be so even if it is platitudinous that there are paradigm cases (e.g. for the case of colour, if it is \textit{a priori} that some judgements about, say, redness are paradigm cases, though it is not \textit{a priori} which particular cases are the paradigm cases). So, the idea is that contents of concepts can be specified by identifying the possibly interconnected core platitudes that form the functional role for a given concept.\textsuperscript{6}

The idea that this style of analysis is potentially non-circular is motivated by the fact that implicit definitions (as in the geometry case) can be turned into explicit definitions by ramsifying. This is done by setting out the core platitudes involving the concept of F, and then stripping away all occurrences of the concept from these platitudes. What remains of the platitudes—in effect a description of the concept’s functional role—is then true of F’s, if there are any. It is then possible to define the concept of F explicitly by saying that it is the thing, whichever it is, such that what remains of the platitudes is true of it. By ramsifying we have ensured that there are no occurrences of the \textit{definiendum} in the \textit{definiens}.\textsuperscript{7}

2. The five platitudes about the concept of meaning.

There are five core platitudes which are constitutive of the concept of meaning. Each of them is indispensable to our concept of meaning: take one of them away and what is left will not be our concept of meaning. I believe the list is exhaustive but I may of course be wrong: claiming that one’s justification is \textit{a priori} is not to claim that it is

\textsuperscript{7}Cf. Braddon-Mitchell and Jackson, \textit{op.cit.}; as well as Smith 1994 and Miller 1997. For more detail on this style of analysis \textit{cf.} Ch. 4§3.c.iii. below.
infallible. My suspicion is that other potential core platitudes will find their place under the heading of one of the five platitudes I put forward.

The five platitudes are as follows:

*Communication*: sentences are used to indicate how things are believed or desired by the speaker to be.

*Meaning is use*: if two sentences have the same use, then they have the same meaning.

*Explanatory meaning*: sometimes the use of a sentence can be explained by its meaning.

*Error and Robustness*: a speaker can use a sentence incorrectly, and sometimes the speaker is aware of errors and reacts accordingly; speakers can have robust disagreements about meaning.

*First-person authority*: speakers have privileged, non-inferential knowledge of the meanings of their sentences, they do not have to investigate their environment to know what they mean.

Such are the platitudes. I shall now motivate them separately, and then in §3 discuss the extent to which they are consistently networked. In §4 I discuss what to do about various internal tensions among the platitudes.

*a) Communication.*

This is the platitude that sentences are used to indicate how things are believed or desired by the speaker to be. It is an incontrovertible fact about language and meaning that we can use the meaningful expressions of our language to transmit and obtain information about how things are taken to be, or how we desire things to be. In general, if I believe or desire that \( p \), and, if \( \varphi \) means \( p \), then I can use \( \varphi \) to indicate the belief or desire that \( p \) to someone. There is an endless variety of examples: you ask me
how the election went and I answer that it went thus-and-so; I want you to be on time so I tell you when the train leaves; I have a vision of a better world and I tell everybody about it; and so on, and on.

Putting the platitude about communication like this does not imply that communication is always a matter of letting you know what is on my mind, in contrast to letting you know how things are in the world. The platitude is neutral between these two communicative tasks: i) letting you know that I believe that the train leaves in an hour; and ii) letting you know that the train leaves in an hour. That is, the platitude does not imply that I can only ever inform others of my subjective ideas, but likewise it does not imply that I can never inform others of my subjective ideas.8

Saying that it is a priori that the meaningful expressions of a language can be used for communication is not equivalent to claiming that it is the prime function of language, in a teleological sense, to enable communication. This would turn the platitude into a non-platitudinous controversial claim, as only people possessing teleological concepts (e.g. Darwinians and creationists) could believe that this is a platitude. This would be misguided and, accordingly, the platitude only says that the analysis of the concept of meaning must make it clear that communication is indeed possible. Whatever else language and meaning may be about, it surely holds that meaningful sentences can be used for communication.

Moreover, the platitude is neutral between those, like Wittgenstein and Dummett, who argue that communication is essentially public, that is, who argue that there must be successful communication for there to be meanings;9 and those, like Chomsky, who stress the importance of our linguistic creativity and argue that language is essentially

8Thus the platitude is neutral vis a vis, e.g., Frege’s and Dummett’s discussion of the objective or public nature of communication, cf. e.g. Frege 1893/1964, preface; Dummett 1991a, ch. 4.
9Cf. e.g. Wittgenstein 1953 §§198-217, 253-8, Dummett, op.cit.
about enabling the free expression of thoughts, whether communicated or not.\textsuperscript{10} I have formulated the platitude so that it can capture both these views. Sentences can be used to indicate the contents of one’s mental states whether this is for one’s own sake, or for other’s sake; whether this is done in intra-personal communication (as in reminiscing) or in normal inter-personal communication; whether those contents are individuated by reference to the world or to internal psychology; or, finally, whether those contents are essentially private, or essentially public.\textsuperscript{11}

The platitude has one important implication, however, viz. that a philosophical account of meaning and communication at some level must involve the mental states of the speakers. This is what communication is about: using meaningful expressions to indicate the contents of mental states, no matter what theory we have about those contents themselves. But this is really uncontroversial: it is consistent with taking either side in the disputes I have mentioned. Those disputes mainly concern what the content of mental states must essentially be like, in order for communication to be possible, and thus they concern, not the semantic relations, but one of the semantic relata.

Notice finally that the notion of ‘use’ is employed in the formulation of this platitude: sentences are used to indicate how things are believed or desired to be. This is important, with respect to our understanding of the “meaning is use” slogan, because it begins to tell us which property of use may be important for the concept of meaning. It is not the use of words to, say, provide background noise (someone muttering the same word over and over), and it is not the use of written words as so-called webpage ‘wallpaper’, rather it is the use of words to indicate the contents of mental states.

\textsuperscript{10}For a survey of this idea and its history, see Chomsky 1966.

\textsuperscript{11}Hence, the platitude is also neutral in the debate between Dummett and Evans about the correctness of the Fregean model of communication; cf. Evans 1982, ch. 9, Dummett 1987, Ch. 15. It is also neutral, I think, in the debate about the so-called ‘private language argument’ associated with the §§ around Wittgenstein 1953, §258.
b) Meaning is use.

This is the platitude that if two sentences have the same use, then they have the same meaning. I believe this is what most naturally captures the main intuition behind the slogan "meaning is use". It captures the principle behind the intuition that if two speakers use the sentences 'it's a dog' and 'det er en hund' in the same way, then they have the same meanings, though they may belong to different languages (in this case English and Danish). The same goes for tokens of the same type of sentence, for example, if you and I use the sentence 'kookaburras look furry' in the same way, then we mean the same thing by that sentence.

i) elaborating the platitude.

The examples not only tell us that meaning is determined by use, they also tell us something about what is meant by the notion of 'use'. Judgements about meanings, as in the 'dog'/'hund' case, are based on what we take the correct use of the sentences to be. Obviously we do not make judgements about the sameness of meaning on the basis of the incorrect use of sentences: it is not a platitude that sameness of use implies sameness of meaning, irrespective of what counts as correct use.

It is possible, from the perspective of an analysis of the concept of meaning, to say more about what counts as the correct use of sentences. In particular the platitude reflects the following intuition: the correct use of a sentence has to do with the contents of the mental states which it is used by competent speakers to indicate. It would be an odd language where a sentence, say, 'there is a possum' is correctly used to indicate mental states with the propositional content there is a possum, and yet where 'there is a possum' means, say, motorcycles are awesome. It is facts about the correct use of sentences to indicate contents of mental states which we exploit when pronouncing that two sentences have the same meaning.
It might be thought that the platitude supports the further claim that, when we pronounce that two sentences have the same meaning, we exploit facts about what things the sentences are used in front of. For some sentences this is of course true (e.g. sentences containing ‘dog’ and ‘hund’), but for many other sentences it is not the case: we do not say that sentences containing the words ‘electron’ and ‘elektron’ have the same meaning because they are used in front of the same things, but rather because they are used to indicate the same types of contents of mental states. So the only conclusion licensed by our commitment to the platitude is that the correct use of a sentence is, at least in part, a function of which contents of mental states it is correctly used to indicate.

ii) the “meaning is use” platitude as a supervenience claim.

It is clear, I believe, from the examples I used to illustrate the “meaning is use” platitude that commitment to it is a commitment to the supervenience of meaning on use: facts about meaning are held to co-vary with facts about correct use such that, if two speakers use a sentence in the same way, then the sentence has the same meaning in their languages. Of course this is only a way of presenting the platitude, it does not imply that speakers who are committed to the platitude themselves possess the concept of supervenience.

Once we know the modal strength of this supervenience claim we can state the platitude more precisely. In the examples above the platitude is presented as a weak supervenience thesis:

---

12 In fact, since everything contains electrons, the two words cannot but be used in front of electrons, but that fact doesn’t impress us at all. The ‘dog’ vs. ‘electron’ example is Horwich’s. I use it for slightly other purposes and do not draw the same consequences as Horwich. *viz.* that the concept of meaning is not necessarily a relational concept. See my discussion in Ch. 8.

13 Notice that, for all I have claimed on behalf of the concept of meaning, we are not committed to the idea that if two sentences have the same meaning, then they must have the same use. That is, perhaps different uses can determine the same meanings. For example, different sets of beliefs and desires may be said to have the same content if the difference is located in some of the ever-shifting background beliefs and desires, *cf.* also Ch. 6§2.
(I) for all speakers S and S*, and all declarative sentences $\varphi$ and $\varphi^*$: if S’s correct use of $\varphi$ is the same as S*’s correct use of $\varphi^*$, then $\varphi$ and $\varphi^*$ have the same meaning.

That is, take any possible world, if two speakers in that world use a sentence in the same way, then that sentence has the same meaning for them. However, for reasons that will appear shortly, I think the platitude is best presented as a strong supervenience thesis:

(II) for all speakers S and S*, all declarative sentences $\varphi$ and $\varphi^*$, and all possible worlds $w$ and $w^*$: if S’s correct use of $\varphi$ in $w$ is the same as S*’s correct use of $\varphi^*$ in $w^*$, then $\varphi$ and $\varphi^*$ have the same meaning.\(^{14}\)

That is, for any two speakers in any two possible worlds, if they use a sentence in the same way in each their possible world, then that sentence means the same in their languages.

iii) apparent counterexamples to the supervenience of meaning on use.

Naturally, if the supervenience of meaning on use is a platitude about the concept of meaning, then it cannot be a controversial claim. It would be a controversial claim if there were plausible counterexamples where a sentence had the same use, but different meanings. In fact, there could appear to be precisely such counterexamples. However, I shall argue that they are in fact not counterexamples to the supervenience thesis as I have presented it (though they may be counterexamples to other types of supervenience theses).

Consider first the case of gold and fool’s gold. The apparent counterexample goes like this: two speakers, S and S*, have used the sentence ‘that is gold’ in the same way, that is, in front of gold as well as fool’s gold, until it was found out that fool’s gold is

\(^{14}\)The terminology of ‘weak’ and strong’ supervenience is Kim’s, cf. e.g. Kim 1994a, 1994b. Strong supervenience entails weak supervenience, but not vice versa since $w$ and $w^*$ may be the same world.
in fact not gold. After that scientific discovery, S stopped using the sentence in front of fool’s gold, whereas S* persists in calling fool’s gold ‘gold’. What this shows, according to the counterexample, is that before they found out about fool’s gold they meant different things by the sentence ‘that is gold’ in spite of using the sentence in the same way: in particular S meant gold, and S* meant, perhaps, golden heavy metal. Therefore, the counterexample goes, the supervenience thesis is false: similar use does not entail similar meaning.

This is not a counterexample to the supervenience thesis as I have presented it. I was careful to point out that the platitude concerns the way sentences are used to indicate contentful mental states, not how sentences are used in front of things. The counterexample trades on a conflation of these two notions of use. Far from being a counterexample to the supervenience of meaning on use, it lends support to it when ‘use’ is understood unequivocally in the more careful sense of ‘used to indicate mental states with the content $F$’. What the story about gold and fool’s gold should make us say is this: true, our intuitions about the case tells us that S and S* meant different things by ‘that is gold’ even before the discovery about fool’s gold, but this is because at that time S and S* also used that sentence to indicate mental states with different contents, just as supervenience would have us believe. For example, S used the sentence to indicate mental states with the content that is the actual golden, rare, heavy, valuable stuff of our acquaintance, and S* used it to indicate mental states with the content that is whatever is the golden, rare, heavy, valuable stuff of our acquaintance. If that is how the notion of use is conceived, then the purported counterexample is not a counterexample to the supervenience of meaning on use, rather it demonstrates that S had a lot of false beliefs about gold, whereas S* had a lot of true beliefs about gold-and-fool’s-gold.

The story might work as a counterexample to the idea that meaning supervenes on use, when the term ‘use’ is understood as ‘use in front of things’. But that latter idea is not
platitudinous. Once ‘use’ is taken in the more guarded sense of ‘use to indicate mental states with the content $F$', then we gain access to a whole set of intuitions which help show that the counterexample doesn’t apply. These are intuitions which endorse quite fine-grained descriptions of contents, and which thereby help bring the counterexamples into line with the supervenience theses (I) and (II).

Although the endorsed contents of mental states are fine-grained, the deliverances of the intuitions are not arbitrary and ad hoc. What shows the failure of the counterexample is just the fact that the contents of our mental states may come in two different types: contents may either be world-bound or world-neutral. In the example above, S uses ‘this is gold’ to indicate a world-bound content, i.e. a content such that its truth-evaluation is sensitive to which world is in fact the actual world. For S it is not only the superficial characteristics which gold shares with fool’s gold which matters, but also facts about which substance gold actually is. The individuation of the content is therefore bound to the actual world. In contrast, for S* the actual substance of what has the superficial characteristics of gold and fool’s gold doesn’t matter—it doesn’t matter for the individuation of content which world takes the place of the actual world.\(^{15}\) It is this difference which is important in dismantling the counterexample: S and S* had different uses all along because S uses ‘this is gold’ to indicate a world-bound content and S* uses ‘that is gold’ to indicate a world-neutral content.

Construing use and the content of mental states like this should be uncontroversial. It is a strategy which proceeds by noting that we have intuitions which support the idea that we have world-bound as well as world-neutral contents, but it avoids taking a stance on the issue of which type of content is conceptually prior, and moreover does not rely on construing world-bound contents as, e.g., object-dependent, and world-

\(^{15}\)For the terminology of world-boundedness see Menzies, Nolan and Pettit, forthcoming.
neutral contents as, e.g., psychologically narrow states.\textsuperscript{16} The reason I can be agnostic about those substantial debates is that they concern a semantic relatum, viz. meanings themselves, as given by the contents of mental states. The debates do not concern the focus of my interest, viz. the semantic relations between sentences and meanings.

Keeping in mind the distinction between semantic relata and semantic relations, we are now in a position to see why another class of apparent counterexamples miss the mark. This is a class of counterexamples which takes the offending notion of the supervenience thesis to be the idea of use as ‘use to indicate contents of mental states’. The counterexample goes like this: there are cases where what is in the mind in two speakers is indiscernible, but where they mean different things by the same sentence. Thus, take two psychologically indiscernible speakers S and S* in each their possible world, \(w\) and \(w^*\) who both utter the words ‘that is water’. Assume that the only difference between \(w\) and \(w^*\) is that what people call ‘water’ on \(w\) is the substance \(\text{H}_2\text{O}\), whereas what people on \(w^*\) call ‘water’ is the substance \(\text{XYZ}\), and that no-one on either world knows about the actual chemical composition of the stuff they call ‘water’. If the content of ‘water’ is world-bound for both of them, then S means \(\text{H}_2\text{O}\) by ‘water’ and S* means \(\text{XYZ}\) by ‘water’.\textsuperscript{17}

This could seem to undermine my answer to the previous counterexample because now, by hypothesis, I cannot help myself to psychological differences between S and S*. But this misses the point: my answer to the previous counterexample did not rely on a notion of narrow psychological states, it relied on different ways of individuating

\textsuperscript{16}That our intuitions support both types of content is an uncontroversial result of Putnam’s “The meaning of ‘meaning’”, 1979; the idea that world-bound contents must be object-dependent appears in Boghossian 1997b, p. 174 (notice that, though Evans 1982 proposes object-dependence he still allows a class of world-neutral ‘non-Russellian descriptive names’); the relative priority between world-bound and non-world-bound contents is discussed in Block and Stalnaker (forthcoming), §9. The best account I know of fine-grained contents that allow both world-bound, and world-neutral intuitions is Jackson 1998b, and Chalmers, forthcoming.

\textsuperscript{17}This is a (modal) version of Putnam’s 1979 Twin-Earth cases.
the contents of mental states indicated by the use of sentences. So, again, what we should say about this new counterexample is this: it is not a counterexample because the story presupposes that S and S* use 'water' to indicate different contents of their mental states. They mean different things by their respective utterances of 'that is water', but that is because their respective uses are also different, and therefore the counterexample does not work.

In sum, the two counterexamples fail to work because they are directed against the wrong types of supervenience thesis. The first counterexample is at most directed against a thesis which has meaning supervene on use of words in front of things, and the second is directed against the thesis that meaning supervenes on psychologically narrow use. The meaning is use platitude supports another supervenience thesis, viz. that meaning supervenes on use of sentences to indicate the contents of mental states, no matter what we may say about the status of content of mental states, and no matter what things the sentences are used in front of. The counterexamples work against supervenience theses which are very controversial, in contrast to the platitudinous thesis, and of course, its platitudinous nature is what ensures immunity to the counterexamples.

Notice that I told the story about S and S* and 'water' in strong modal terms. The story considers two speakers in each their possible world, and yet the supervenience platitude was strong enough to refute the purported counterexamples. This demonstrates that commitment to the meaning is use platitude is commitment to a strong supervenience thesis, in the sense captured by (II) above: if two speakers in each their different possible worlds use two sentences similarly, then those sentences have the same meaning.

The story could have been told in a non-modal version (viz. as concerning two speakers on near-indiscernible planets in the same possible world, say, Earth and
Twin-Earth on opposite sides of the sun) and then only the weak supervenience thesis (I) would have been necessary to refute the second counterexample. However, it is desirable to exclude the possibility that modal versions of the counterexample are significant for the analysis of meaning. Luckily refuting the modal version of the counterexample with the help of (II) does not introduce any controversial philosophical machinery with respect to the concept of meaning, and should therefore not offend against our platitudinous grasp of the concept of meaning.

iv) is the platitude interesting?

Should we be suspicious of a platitude which handles these counterexamples in such an off-hand manner? Would there be anything which could refute it, or is it really masking a tautology, i.e. the thesis that if two sentences have the same meaning, then they have the same meaning? There is a scenario which could refute the supervenience of meaning on use, viz. where the meaning of a sentence is independent of the mental states which the sentence is used to indicate. Imagine a Platonist semantics for a language L which pairs sentences and propositions independently of how speakers of L use those sentences. For any such sentence, \( \psi \), it will be possible that it means *that is a waratah*, but that the speakers of L never use \( \psi \) to indicate mental states with the content *that is a waratah*. Hence, on such a Platonist semantics it is not the case that, if two sentences have the same use, then they have the same meaning. So there is a putative class of counterexamples to the supervenience platitude and therefore it is not insignificant. This tells us that one must defend a Platonist or kindred semantics in order to contest the supervenience of meaning on use as a platitude about the concept of meaning. In Chapter 3 I discuss the feasibility of such moves. In addition, throughout the essay I shall address the important question about which kind of property could be such that it determines the correct use of a sentence as the use to indicate all and only mental states with the particular content \( p \). Though the platitude about meaning and use seems to require the possibility of such a property it is by no
means obvious that there could be such a property, nor that it will be easy to identify it.

c) *Explanatory meaning.*

This is the platitude that sometimes the use of a sentence can be explained by its meaning. It captures ordinary locutions such as “she said ‘there is the billabong’ because that sentence means *there is the waterhole* in her language”; and “the kid said ‘wobbaly’ when that wallaby jumped past because the kid means *wallaby* by that homespun word”, and so on. Often we use meaning to explain use when we are interested in factoring out the contribution of facts about meaning from the contribution of the relevant facts of the matter. For example, if you say “there is the billabong” and I say “there is the waterhole”, pointing in the same direction, then our communicative dealings are best served by explaining the discrepancy between us in terms of meaning, rather than explaining it in terms of the relevant facts of the matter, in this case, the locations of waterholes. Since these types of explanations are straightforward parts of our dealings and competence with the concept of meaning, an adequate analysis of the concept must reflect our entitlement to them.

It must be undisputed that we sometimes avail ourselves of facts about meaning to explain uses. But is it a core platitude about meaning? That is, could one imagine a recognisable concept of meaning which does not allow such explanations? Davidson argues that sometimes communal linguistic meaning is not necessary to explain use. For example, if I say “let’s hit the frog and toad” my use of that sentence is not easily explained by the ordinary communal linguistic meaning of ‘frog’ and ‘toad’. In those cases context plays the important role in the explanation of use.18 Davidson suggests that this demonstrates that linguistic meaning is inessential to communication. Does this show that the platitude about explanatory meaning is not a core platitude about meaning after all? I do not think so. Davidson’s comments are directed against the

---

ubiquity of explanation of use in terms of communal linguistic meaning, but my
version of the platitude is not committed to explanations of use solely in terms of such
communal meanings. As my examples above show, sometimes we explain use in
terms of idiolectic, or dialectic, meaning. In fact, this is often the most natural use of
such explanations since the situations where we call upon them are most often when
we are baffled by differences in idiolects or dialects. So Davidson’s examples of non-
standard communication do not work directly against the platitude.

Perhaps, however, the examples of successful non-standard communication (which
also include malapropisms, spoonerisms, slips of the tongue etc.) are intended to
demonstrate that when communal linguistic meaning fail to explain use, then other
meaning-independent facts step in to explain use. I think this is a *non sequitur*: the
dialectic could just as well work the other way. Such non-standard cases of successful
communication are only enabled precisely because we know what it would be for
‘let’s hit the frog and toad’ to have a certain meaning that could explain its use. Thus,
one way to deal with such cases is to say the following: it is easy to see how, in some
nearby possible world, ‘let’s hit the frog and toad’ could have the same meaning as
‘let’s hit the road’, and in that world its meaning would explain its use. My suspicion
is that if it didn’t work somewhat like this, then we would be hard put to distinguish
between semantic and non-semantic phenomena at all. Of course, fleshing this out
would depend on how meaning is constituted; but that account would probably cite the
fact that, in this world, the use of the two idioms overlap to some extent. Viewed like
this, the platitude retains its core status: whenever there is a semantic phenomenon we
can imagine situations where the meaning of a sentence can explain its use. I do not
think this is controversial: it only draws on the intuition that if something counts as a
semantic phenomenon, then we should expect there to be semantic relations between
sentences and meanings, and it is these relations—however fleeting and fragile they
may be—which we can draw upon in explaining why a sentence was used as it was.
Notice that the platitude is neutral in the debate about the relative conceptual priority of communal languages and idiolects. Since it is a platitude, both views must operate with a notion of explanation of use in terms of meaning. If communal languages are prior, then we should want explanations of an individual's use which exploit facts about communal meanings (e.g. “she said 'that is a kookaburra' because ‘kookaburra’ has the communal meaning *kookaburra*”). If idiolects are prior, then we should want explanations of an individual's use which exploit facts about idiolectic meanings (e.g. “he said ‘there is a kookaburra’ because in his idiolect ‘kookaburra’ means *kookaburra*”). Presumably there will be differences in the significance and substance of those two types of explanation (maybe because there are arguments showing that communal explanations are parasitic on idiolectic explanations, or vice versa). But such differences do not derive from the platitude about explanatory meaning itself, they will derive from the arguments one gives in support of one or the other priority thesis.

In all this I have not said much about the content of the notions of ‘explanation’ and ‘use’. What counts as an ‘explanation’ and what is the ‘use’ which is being explained by meaning? Of course these questions have to be sorted out in order to understand fully what the platitude means. However, I shall postpone this task to Chapter 5 because explication of this platitude depends very much on the explication of the meaning is use platitude and our conception of the property which the concept of meaning is a concept of. I shall mention one crucial aspect however, viz. that the explanations of use in terms of meaning should exploit facts like this: S used φ because S knows/grasps/believes something about/understands the meaning of φ. That is, the explanations must attribute intentional or motivational capacities to the speaker whose use is being explained by meaning. The explanations must conform to this because it is platitudinous that we do not explain use in terms of meanings which somehow lie beyond the cognitive purview of the speaker. For example, it would be

---

19 Cf. e.g. Davidson 1994 and Dummett's response in the same volume.
very odd to explain my use of an expression by adverting to what someone else means by that word (e.g., I use ‘kant’ like I do because it means edge in my language Danish, not because it means Immanuel Kant in German). This is not yet to say anything substantial about what would qualify as a motivational explanation for the purposes of the concept of meaning. Rather, my remarks register a feature of our ordinary explanations of meaning in terms of use. Someone offering non-motivational explanations (as the one about ‘kant’ above) simply demonstrates a lack of core competence with the concept of meaning (here I have defined ‘motivational’ negatively as: does not exploit meanings which the speaker couldn’t be motivated by).

This last point emphasises the motivational nature of the explanations required by the platitude about explanatory meaning. The explanations have to have a special motivational character to count as explanations of use in terms of meaning. This gives an immediate line of response to someone who baulks, firstly, at the idea that meanings—i.e. abstract objects—can occur in empirical explanations at all, and, secondly, at the idea that mere meaning-facts can explain uses of sentences which, supposedly, are intentional actions. This kind of objector would then object to the platitude like this: “the notion that meaning explains use is not a platitude about the concept of meaning. Rather the following is a platitude: use is explained by speakers’ beliefs about meaning.” The objector would therefore say that if there is a causal chain it goes via the beliefs of the speaker. But this is fine according to the platitude as I have now set it out. The explanations must be motivational, so they could go like this: the fact that Ψ means p causes S to have the belief that Ψ means p, and that belief (together with some appropriate desires) causes S to use Ψ in a certain way. The result is that the platitude is neutral in the debate about the causal efficacy of abstract objects, as well as in the debate about the right characterisation of intentional action.
d) Error and Robustness.

This is the platitude that a speaker can use a sentence incorrectly, sometimes the speaker is aware of errors and reacts accordingly; speakers can have robust disagreements about meaning. It is a commonplace that sometimes speakers make errors in the way they use their language. It happens often for speakers who are trying to learn a new language, but a speaker can make errors within his or her own native language too: one can have slips of the tongue, fail to be attuned to changes in meaning, fail to pick up new meanings correctly, and so on. Also, though children are amazingly fast at picking up language, they make numerous errors. In the case of children these errors are picked up by parents and teachers, and they are then corrected. The same goes for errors made by adults: the errors of a speaker trying to break into a new language are detected, explained and corrected; as are errors made within one's native language. So there is a rich phenomenology of error which conforms with the platitude: we do make errors, we can be made aware of them, we do in fact often react accordingly (though sometimes repeated correction is called for), and this happens through bona fide disagreements about meaning. From the phenomenological standpoint it seems a platitude about meaning that we can make errors and have robust disagreements.

But is it an a priori core platitude? I think it is, and there is a line of reasoning which fits the phenomenology nicely. Imagine a situation where there are no facts about what is to count as the correct application of the sentence 'there's an echidna', where there is no correct answer to the question "was that utterance correct?", hence a situation where the ideas of appraisal of error and corresponding reaction, and the idea of robust disagreement, can get no grip. I think it is clear we would have to say that in this situation the sentence has no meaning. It is a situation where we are not able to draw upon resources by which we normally identify linguistic phenomena. Think for instance of how hard it would be to teach someone the meaning of such a 'sentence'. This is not to say that there must actually be errors in the use of every sentence. It is to
say that we must have a clear conception of what *would* count as an error in the use of a sentence. This shows that the platitude about error and robustness is really a core platitude: without it we will fail to recognise linguistic phenomena.\(^{20}\)

What type of error are we talking about? The question, "was that utterance correctly used?" needs to be qualified: correctness in *what* with respect to *what*? It seems clear that an error about meaning issues in falsity in what was said. But such falsity may in fact be produced in two different ways: i) it might be produced by falsity in what was believed, or ii) it might be produced by getting the meanings wrong. For example, if S says "there's an echidna" when in fact it is a common hedge-hog in front of her, then there is falsity in what S said. Naturally, this has to be possible, but such errors need not be produced by S getting the meanings wrong. The reason is simply that, in the example, there need be no error about meaning: S may *believe* that it is an echidna, and on that basis utter the sentence 'there's an echidna'. That application of the sentence would then be correct with respect to meaning, in spite of the fact that there was falsity in what was said. In contrast, where S may truly believe that there is an echidna in front of her, but *says* 'there is a marsupial mole', she has produced a falsehood in what was said as a result of an error about meaning because in her language, let's assume, it is false that 'there is a marsupial mole' means *there is an echidna*.

This leads to the following question: which type of error is relevant for the analysis of the concept of the relation expressed by (*φ* means *p*)? As this analysis solely concerns the semantic relation, not the semantic relata, the answer must be that it concerns the type of error which is produced by getting the meanings wrong, irrespective of whether or not there is falsity in what is believed. Naturally this restricts the scope of my analysis because, for all I am going to say, it may be possible that we should be

\(^{20}\)Of course, this kind of reasoning has been heavily exploited in the discussion of the rule-following considerations, especially the debate ensuing Kripke's 1982.
sceptics about the possibility of error in what is believed. In this essay I shall simply assume that such error is possible. However, the possibility of error in what is said does not follow from the assumption that there is possibility of error in what is believed. Therefore it is important to discuss how such error may be possible.

The above considerations show that the possibility of error in what is said is a core platitude about the concept of meaning. But what about the rest of the platitude concerning recognition of error, appropriate reaction and disagreement about meaning? Could those parts of the platitude be merely accidental? I think they belong with the core platitude. Imagine a situation in which there are facts about the correct and incorrect use of the sentence ‘marsupial moles have no eyes’, but where the speakers of the language to which the sentence belongs have no way of robustly disagreeing about whether a given use was correct or not, and where speakers have no way of recognising that a contested use was incorrect, and consequently no idea of what they should do about it. In this case we might be baffled because the idea that there are correct and incorrect uses of a sentence is normally closely linked to the phenomenology of disagreement, recognition and reaction. Of course, we can easily imagine that for some contingent reason we have no epistemic access to the facts about the correct use of a sentence in another language (some ancient script, say), but we would surely expect speakers of that language to have the proper epistemic access to facts about the sentence’s correct use. This suggests that the parts of the platitude come together: for something to count as an error it must be the case that we can detect and acknowledge it and react accordingly. And errors must be possible for those practices to be well-founded. Hence, I think the whole platitude about error and robustness belongs with the set of core platitudes.

Notice that the argument does not say anything like the following: since we have the practices of disagreement, recognition and correction, facts about correct and incorrect use must exist. This would be a non sequitur. Naturally, it may still turn out that we
have to hold an error-theory about meaning: perhaps there are no properties which could establish facts about correct and incorrect use, and so our practices of disagreement, recognition and correction would issue in judgements which cannot be made true in our world. All the argument says is this: the platitude about error and robustness is one of the core platitudes we talk about when we talk about the concept of meaning.

I have described disagreement about meaning as 'robust' disagreement. This is so in order to register what I believe is a platitudinous commitment to the possibility of disagreements which matter to us and which we are interested in dissolving. To see why this is a special concern, consider the following train of reasoning. Consider a language L (L may be an idiolect or a communal language). Assume that there are facts about the correct and incorrect use of the sentences of L, and that therefore, in theory, disagreement about the meanings of sentences of L are possible. But now ask: "what prevents L from going haywire?" That is, what reason do we have for believing that we do not constantly come up with new meanings for the sentences of L? The scenario of the haywire language L is consistent with there being facts about the correct use of the meaningful sentences of L. The problem is just that we constantly change these meanings. In this case disagreements are in principle possible, but they would not be robust. There would be no need to engage in disagreements, and care about reacting accordingly, because the fact of the matter is going to change the next moment anyway. It is not this kind of disagreement the platitude commits us to. Rather, the platitude commits us to a belief in sufficient unity and stability in the facts about correctness and incorrectness. If there is no such unity and stability, then there ceases to be a point in engaging in disagreement about meaning. This is not to say, of course, that the platitude commits us to a belief in a semantics that never change. It only commits us to a belief in a semantics that is stable and united for *long enough* to make sense of the possibility of robust disagreements about meanings.
The discussion of robustness also throws light on the notion of recognition of error and appropriate reaction. The fact that we are interested in resolving disagreements about meaning reflects the fact that we are interested in recognising and correcting errors about meaning. The rationale for this interest is that it is important for us to express ourselves clearly and in a way conducive to communication and transmission of information. The platitude itself doesn’t commit us to any particular way of recognition and correction. However, it is important that both these aspects of the phenomenology are in place. The whole idea of disagreement about meaning and appropriate reaction makes little sense unless we can recognise errors. However, saying that the analysis of meaning must make space for the phenomenology of error-recognition is not to say that our recognition of error is guaranteed—scepticism may after all be a real possibility. The idea of appropriate reaction is also important. Without this aspect of the phenomenology of meaning in place the other aspects will be pointless. Unless a speaker can take appropriate action on recognising an error there is no point in engaging in disagreements about meaning at all. It would be like engaging in a moral debate, recognising that a particular course of action is morally good all things considered, but declining to act accordingly. I shall call this important part of the phenomenology of meaning the aspect of \textit{practicality of meaning}.\footnote{Cf. Smith 1994, Ch. 1, for the moral case.}

In chapter 6 I shall return to this platitude. Spelling it out depends to a large extent on how the other platitudes come out. I hope what I have said is enough to motivate it as a core platitude for the concept of meaning. Commitment to it should be uncontroversial, and it is a platitude which no account of the concept of meaning can afford to ignore.

\textit{e) First-person authority.}

This is the platitude that speakers have privileged, non-inferential knowledge of the meanings of their sentences, they do not have to investigate their environment to know
what they mean. Competent speakers of a communal language, or a dialect or idiolect, are automatically invested with first-person authority about meaning in the sense that they do not need to be able to provide empirical evidence to justify that they know what they mean. Initially this can be defined negatively: speakers do not have to investigate the environment, e.g., they do not have to obtain a third-person perspective on themselves and their previous use in order to know what their sentences mean. Formulated like this it fits the traditional definition of the *a priori* as that which can be known without having to investigate the environment. Therefore I shall speak of it as *a priori* first-person authority.

i) The phenomenology of first-person authority

Let me first distinguish *a priori* first-person authority from an alternative way of talking of meanings which is manifestably not *a priori*. In order to come to possess and gain competence with the sentence ‘marsupial moles are rare’ in English, I do indeed need to investigate the environment: I have to seek an English-speaking community and discover how they use that sentence. Obviously, obtaining such knowledge is not an *a priori* matter at all. But this is not what the platitude intends to capture. Instead, it intends to capture the asymmetry between the way already competent speakers of a language know the meanings of what they themselves say, and the way they know the meanings of what their interlocutors say. It goes against our intuitions about the concept of meaning to say that we know our own meanings by imagining ourselves as ‘interlocutors’ whom we ‘interpret’ just as we interpret any stranger we meet on the street. That is not how it is. We may legitimately doubt whether we have grasped what a normal interlocutor means, and therefore ask “but do you mean *this* or *that*?”, but there is something seriously off-colour about asking that question of ourselves. Other people may ask us that kind of question, but they do so because they have a third-person perspective on us. My claim is that if one’s account

---

22For the discussion of the notion of privilege see e.g., McKinsey 1991; for the focus on the first-vs. third-person cf. Shoemaker 1996, pt. 1; for inferential vs. non-inferential see Wright 1989, p. 236 and Boghossian 1989b, §4.
of meaning cannot make this phenomenology come out true, then something has gone wrong.

However, the platitude could easily seem controversial: the idea that we have infallible a priori knowledge of the meanings of our words might seem objectionable. It might be seen as bad Cartesianism: the picture of infallible access to an inner mental domain. But this worry is unfounded: saying that first-person authority is a priori is not saying that it is infallible. It is not part of the concept of meaning that we have perfect transparency of mind, or that we know everything there is to know about meanings. It is part of the concept that the evidence we have for our authority is a priori, not that the evidence is infallible. So the platitude is neutral in the debate about Cartesianism and perfect transparency of mind.

In this analysis I am interested in the semantic relations between sentences and meanings, rather than in meanings themselves. Therefore, the platitude does not concern that which we have first-person authority about, rather it concerns the semantic relation itself. It requires that the semantic relation expressed by "\( \phi \text{ means } \mathcal{P} \)" should be such that it allows non-inferential a priori first-person authority, and makes sense of the third-person/first-person asymmetry. Meanings themselves are for my purposes abstract objects: the contents of mental states. There are many different theories of content, but such differences are orthogonal to the platitude as I have set it out. For example, commitment to the platitude doesn't force a controversial choice between theories according to which contents are sets of possible worlds, or conceptual roles, or methods of justification.

---

ii) state authority and content authority.

It is possible to describe in more detail the phenomenology of first-person authority about meaning. Consider a distinction between what can be called state authority and content authority. State authority has to do with our knowledge of which mental state we are actually in, thus it has to do with what may properly be called self-knowledge. Content authority has to do with the contents of the states which we are possibly in, we can have content authority even though we have no idea how reliable our state authority, or self-knowledge, is. First-person authority about meaning is content authority, that is, authority about the contents of our possible thoughts. Without content authority we would not be able to ask the further question about the reliability of our self-knowledge. For example, if I do not know the content of the proposition *wobbegongs are nocturnal*, then I cannot intelligibly ask myself whether I really believe that wobbegongs are nocturnal.

It is useful to think of this distinction in terms of first- and second-order beliefs as reflected, for example, in the difference between these two ways of talking: “I believe that wobbegongs are nocturnal” vs. “I believe that I believe that wobbegongs are nocturnal”. I am claiming that first-person authority must have a role to play already at the level of first-order beliefs. This says something about the context in which first-person authority plays its role, but it says nothing about the *a priori* privilege of first-person authority. It is then a further issue how content first-person authority can be *a priori*.

The platitude is neutral between the ways in which the distinction between state authority and content authority can be drawn. The idea that certain mental states are possible-for-me just captures what must be true if it makes sense to ask myself

---

24I thank Philip Pettit for discussion of this point.
25Wright, in a discussion about semantic externalism, argues for a similar distinction between first- and second order beliefs, but doesn't make the general point about possible and actual states, 1992b, p. 79; see also pp. 74-5, 79-81.
whether I actually believe that P or believe that Q, or to ask myself whether I actually believe that R. That simply presupposes that there could be a narrower set of mental states which consists of the mental states that I am actually in. Perhaps one can characterise the broader set of possible mental states as those states which are consistent with my actual overall doxastic state. One can imagine many views about the functions that can take one from the broader set of possible states to the narrower set of actual states: views about broad and narrow content (e.g., it is possible that I believe water falls from the sky and it is possible I believe twin-water falls from the sky, I actually believe water falls from the sky); views about self-deception (e.g., I self-deceptively believe that white tip sharks aren't dangerous, but actually I do believe that they are fierce man-eaters); views about joking (I have the make-believe belief that this cake is poisonous, but actually I don't believe it is poisonous), and so on. The platitude is neutral between these ways of drawing the distinction. It only says that since we can draw these distinctions we must be able to have first-person authority about our possible states.26

iii) the form of first-person authority.

Lastly, let me say something about the form which manifestations of first-person authority take at the first-order level of content authority. At the first-order level, the issue simply concerns our right to use sentences such that it is a priori that we shouldn’t be challenged in doing so. It doesn’t concern our ability to explain the meanings of sentences in other sentences (e.g., saying “the sentence ‘wobbegongs are nocturnal’ means carpetsharks are active at night”), because sometimes we are authoritative even though we don’t know what to say (e.g., “the word ‘wobbegong’ means, you know...wobbegong!”). We can then represent the issue like this:

26Notice that it also doesn’t take a stance on a Cartesian view that holds total transparency of mind. According to such a view there is no distinction between our actual and possible states (the Cartesian must then, it seems, hold an error-theory about the phenomena which I use the distinction to explain (jokes etc.)). Cartesian first-person content authority is then about that one set of mental states. (One might wonder whether a Cartesian can consistently be an error-theorist about mental phenomena. This might show up an internal inconsistency in the doctrine, but that would be independent of the platitude).
establishing first-person authority about our possible meanings and contents comes to
establishing our right to disquote truthfully, and on a non-inferential and \textit{a priori} basis,
as competent speakers of a communal language, dialect or idiolect. But this is
somewhat artificial, as we don't exert first-person authority only when we are
disquoting ourselves: we exert it all the time, whenever we use sentences in thought or
talk. So the point about disquotation comes to this abstract notion: whenever I use a
sentence, I could, in principle, truthfully and on non-inferential and \textit{a priori} grounds,
disquote what I just said.

\textit{f) Remarks on the platitudes.}

I have now tried to motivate the \textit{a priori} platitudes constitutive of the concept of the
relation expressed by "\( \varphi \) means \( p \)". The platitudes count as \textit{a priori} because I have tried
to motivate them on the basis of what someone who possesses the concept is able to
do, without having to investigate the environment. I do not claim that this method is
infallible, or that one can possess a concept without at the same time gaining
knowledge about how it is in actual fact correctly used (as the meaning is use platitude
has it). Nevertheless, this would be knowledge which one comes to possess in virtue
of possessing the concept. I have only motivated the platitudes, I have not given full
philosophical argument for their status as \textit{a priori} platitudes. As we shall see next, the
platitudes are in tension as they stand, so we should expect some alterations in a
complete and adequate analysis of the concept. I shall approach this task in the
chapters to follow. But though I have not given a full account of the platitudes, I have
tried to motivate them in such a way that they should appear \textit{prima facie}
uncontroversial: each platitude is neutral in the main debates associated with it.
3. Three tensions among the platitudes.

I have motivated the platitudes about the concept of meaning in their own right. But doing so of course does not imply that the platitudes are mutually consistent for there might be various forms of tension between them. Indeed it seems as if there are a number of tensions among the platitudes. I discern three such tensions. In what follows, I present and motivate the tensions, and, in the last section of this chapter, I discuss strategies of what to do in the face of the tensions.

a) The first tension: the "meaning is use" problem.

The first tension is between the meaning is use platitude and the explanatory meaning platitude. It arises in the following way. Consider the following lemma: explanation is asymmetrical. That is, \(\neg(A \text{ explains } B \& B \text{ explains } A)\). The lemma is simply a way of saying this: we expect some sort of cognitive value to come from explanations like ‘A because B’, such that there may be something about A which we do not know but which we get information about by the explanation. But if we are also told that ‘B because A’, then our initial ignorance about A hasn’t been removed, in spite of the explanations. That is to say, the explanations weren’t really explanations after all.27

Now consider the two platitudes. Prima facie, the meaning is use platitude says that use explains meaning, and the explanatory meaning platitude says that meaning explains use. Hence, the two platitudes together says that meaning is explained by use and explains use. But now we get an inconsistency:

\[
\begin{align*}
(1) \text{ use explains meaning} & \quad \text{meaning is use} \\
(2) \text{ meaning explains use} & \quad \text{expl. meaning} \\
(3) \text{ use explains meaning} \& \text{ meaning explains use} & \quad \&I \\
(4) \neg (\text{ use explains meaning} \& \text{ meaning explains use}) & \quad \text{lemma}
\end{align*}
\]

27 On explanatory asymmetry see Horwich 1987, p. 147-156.
In the face of such inconsistency one has to either refute one of the platitudes or somehow argue that the inconsistency doesn't arise. (Of course, one can also bite the bullet and accept that our concept of meaning rests on inconsistent core beliefs). In the first instance, this is an example about how considerations about one platitude may interfere with considerations about another, sometimes given known facts (in this case, the lemma). That is, the platitudes may be networked in various ways. This is an aspect of a platitude-based analysis of a concept which should not be ignored: setting out the platitudes individually should be followed by considerations about their networked character.

Of course, I think that this is more than an example of the networked character of the platitudes. I believe this is an interesting problem in its own right. It is interesting because, unless one bites the bullet of inconsistency, one must positively prove that one's platitude-based analysis avoids the problem.

In this particular case it might seem easy to prove that the problem doesn't arise. Firstly, it is easy to prove if is one already, for other reasons, inclined to reject one of the two platitudes. But, secondly, one could also dissolve the problem if one thinks that the meaning is use platitude is a supervenience thesis and the explanatory meaning platitude isn't. The explanatory 'because' is presumably different in supervenience claims and non-supervenience claims. Therefore there can be no explanatory symmetry because (3) above would come out as:

\[(3^*) \text{ use explains}_{\text{superveniently}} \text{ meaning} \& \text{ meaning explains}_{\text{non-superveniently}} \text{ use}\]

However, I have only motivated the meaning is use platitude in supervenience terms, I haven't justified the claim that meaning does in fact supervene on use, nor have I explained in detail what types of explanation of use in terms of meaning there might be. And importantly, I have done nothing to show that the supervenience of meaning
on use is consistent with explanations of use in terms of meaning. This may in fact not be so easy, for the following reasons.

Reflect first that, in so far as we can substitute identicals in explanatory contexts, the lemma about explanatory asymmetry entails a principle of non-identity of \textit{explanandum} and \textit{explanans}:

\begin{align*}
(1) & \text{ A explains B } \rightarrow \neg (\text{B explains A}) & \text{ explanatory asymmetry} \\
(2) & \text{ A explains B } \& A = B & \text{ for reductio} \\
(3) & A = B & \&E \\
(4) & \text{ B explains A } \rightarrow \neg (\text{B explains A}) & \text{ subs. of identicals 1, 3.}
\end{align*}

therefore, by reductio on (2):

\begin{align*}
(5) & \text{ A explains B } \rightarrow A \neq B & \text{ explanatory non-identity}
\end{align*}

which, for the case of the explanatory meaning platitude, is:

if meaning explains use, then meaning \neq use.

The explanations associated with the explanatory meaning platitude could satisfy the non-identity principle by virtue of being explanations of temporally distinct items, \textit{e.g.}, explanations of the form: S accessed the meaning of \( \psi \) and therefore S then uttered \( \psi \). On the other hand, if the meaning is use platitude is a supervenience thesis, then it should be seen as offering non-temporal explanations like: \( \psi \) means \( p \) because use conceptually entails meaning. But now we get an odd result: the explanatory meaning platitude requires that meaning and use come apart temporally, but the meaning is use platitude require that they don’t. It is odd because it seems that meanings must be in place before they can explain use, but they can only be in place if use has conceptually entailed meaning; so, the uses explained by meanings cannot occur after the meanings come into place. Hence, it is not sufficient to posit different
explanatory contexts in order to solve the “meaning is use” problem, for the explanatory contexts must also be shown to be consistent for the case at hand.

One way to achieve this result is to claim that one of the terms involved is ambiguous, e.g., that the notion of ‘use’ occurs with different contents in the different explanatory contexts:

\[
\begin{align*}
(1) \text{use}_1 \text{ explains}_{\text{conceptually}} \text{ meaning} \\
(2) \text{ meaning explains}_{\text{causally}} \text{ use}_2,
\end{align*}
\]

in which case we do not get the odd result. However, this imposes more constraints on the formulation of the platitudes: not only must the explanations offered in them differ, but the notions of use must be shown to differ in such a way that the explanations can be consistent. It is not immediately obvious how this may be done for the two platitudes.

The claim is thus not that the “meaning is use” problem is unassailable, but rather that solving the problem imposes constraints on the formulation of the platitudes. This is potentially problematic because satisfying the constraints may force us to make controversial claims about the platitudes.28

---

28I think the “meaning is use” problem is inherent in much work on the “meaning is use” slogan (for a recent example see Lance and O'Leary-Hawthorne 1997, p.173. They argue that the idea that meaning explains correct use has been underappreciated, and the idea that use explains meaning has been misunderstood). I think that the tension shows up in Wittgenstein’s own works, indeed I think it explains why those works are so hard to interpret univocally. For the idea that meaning is (nothing but) correct use, see e.g., Wittgenstein 1953 §§43, 30, 41; for the case of mathematics see 1984a, I.§4, VI.§10, and, for the idea in the debate about private languages see 1953, §§258-60. For the idea that use explains meaning see e.g. 1953, §199, but also 1984b, p. 397: ‘340. If the language game, the activity, e.g. building a house (as in Investigations §2), fixes [fixieren] the correct use [Verwendung] of a word, then the concept of the correct use is correlated with the concept of the activity. But that is all in the essence [Wesen] of language.’ (my translation). For a nice statement of the tension see 1984a, vol. 6, III. §37: ‘What I keep doing, what it looks like - brings forth a distinction between determination of meaning [Sinnbestimmung] and employment of meaning [Sinnverwendung]’ (my translation).
b) The second tension: the problem from irregular and infallible use.

The second tension concerns the meaning is use platitude and the error and robustness platitude. It comes in two parts, one concerning robustness and one concerning error. They both build on the assumption that there are no, or few, or inadequate, constraints on the notion of use as it is employed in the meaning is use platitude.

i) the problem from irregular use.

The problem concerning robustness is the 'problem from irregularity of use' and it goes like this:

(I) If meanings are robust in the sense required by the platitude about robustness, then meanings must somehow be regular; and meanings cannot be regular if they are determined by speakers' whim [robustness].

(II) If meaning is use, then meanings are determined by speakers' use; speakers' use is determined by their whim [crass conventionalism].

(III) If meaning is use, then meanings aren't regular, and therefore not robust.

The idea that meanings must be regular to be robust comes naturally with the error and robustness platitude. According to that platitude, it is not enough if error is possible: it must be possible in such a way that it makes sense to engage in disagreements about meaning, and to be interested in appropriate reaction when one is shown to be at fault. As a minimum, this requires that the same sentence \( \psi \) retains its meaning for long enough to allow the practices of disagreement and reaction to make sense. There would be no point in engaging in disagreements, with a view to subsequent reaction, about the meaning of, say, 'triggerfish make coral nests', if the normal expectation is that the sentence constantly changes its meaning.

The idea that the meaning is use platitude entails irregularity is based on the assumption that use is irregular because it is subject to speakers' whim. This
assumption builds on a simple perception: nothing prevents me from using ‘triggerfish make coral nests’ now with the meaning *triggerfish make coral nests*, and now with some other meaning. If use in general is so unconstrained, then we have no reason to believe that use of sentences is regular.²⁹

If it is correct that use entails irregularity, and robustness requires regularity, then the two platitudes are in tension. *Prima facie* the meaning is use platitude doesn’t combat the perception that use can be irregular, though it doesn’t suggest that use entails irregularity either. The supervenience claim simply says that if two speakers, in each their possible world, use a sentence in the same way, then their sentences have the same meanings. It doesn’t say that the uses must be regular. Likewise, the robustness platitude does seem to require regularity.

Of course, it may seem easy to resolve the tension: simply rephrase the supervenience claim so that it has meaning supervene on regular use. But it is not clear that such a move would be uncontroversial. Much depends on what ‘regular’ is supposed to mean, on how regular meanings must be in order for disagreements about meaning to be robust (in particular whether they need to be so regular that they become independent of use), and so on. Another reason it may be controversial is that considerations about regularity have to do with what it is for a subject to use sentences, and it is not obvious that the considerations that underpin the supervenience thesis offer any support for any particular view about what is involved in making use regular. In addition, we have seen that the meaning is use platitude is networked with the explanatory meaning platitude: therefore, if we change the meaning is use platitude we would have to ensure that no unforseen consequences result for the explanatory meaning platitude. Likewise, the platitude about first-person authority may be

²⁹Perhaps this objection seems silly. Of course language is not riddled with irregular use (though don’t forget Lewis Carroll’s Humpty-Dumpty!) But the objection is based on the idea that if meaning is use, then our normal intuitions about regularity are put under pressure. Pettit 1991 discusses a similar issue. I return to Pettit’s discussion in Ch. 6 §1.d.
CHAPTER II, FIVE PLATITUDES ABOUT MEANING, AND THE TENSIONS AMONG THEM

sensitive to changes in the meaning is use platitude: if the notion of regularity concerns what it is for speakers to use sentences, then perhaps there is a difference between having first-person authority about irregular meanings and having first-person authority about regular meanings.

Again, the problem from irregular use need not be unassailable, but if we suppose the platitudes are consistent, then one of either involved platitudes must be rejected or reformulated.30

ii) the problem from infallible use.

The next problem also raises concerns about the meaning is use platitude, but this time the concerns are even more serious. In the previous problem the worry was that the platitude may allow meanings, but that these meanings aren’t regular. This time the worry is that the platitude may not allow meanings at all. The argument proceeds by strengthening the assumption about unconstrained use:

(I) If $\text{meaning} = \text{use}$, then $\psi$ has a meaning.

(II) If $\psi$ has a meaning, then it must be possible to make errors about the meaning of $\psi$ [error platitude].

(III) If $\text{meaning} = \text{use}$, then even $\text{erroneous use} = \text{meaning}$, and then it is not possible to make errors about the meaning of $\psi$ [use=correctness].

(IV) If $\text{meaning} = \text{use}$, then it is possible and not possible to make errors about meaning.

30This type of problem (though not in this platitude-based form) receives a great a deal of discussion in Pettit 1991 and 1993, and forthcoming. Pettit is thus one of the few people who distinguish it clearly from the related, but distinct, problem about the possibility of error. Johnston, 1992, discusses a related form of this problem, for the case of colour, not meaning, in terms of the ‘unity’ of colour. I think that, at a deeper level, the worry about irregularity is related to Kripke’s worries about the transtemporal identity of purported meaning-facts, cf. Kripke 1992, Ch. 2, and e.g. Coates 1986.
The second premise is an expression of the error platitude. The third premise is motivated as follows. Assume, for reductio, that an utterance of $\psi$ is an erroneous use of $\psi$, e.g., it is used to indicate mental states with the content $q$, whereas its correct use is to indicate mental states with the content $p$. Now reflect that the erroneous use is itself part of the use of $\psi$, which according to the supervenience thesis constitutes its meaning. But now it follows that the utterance of $\psi$ wasn’t incorrect because if S uses $\psi$ to indicate mental states with the content $p$ and mental states with the content $q$, then $\psi$ means both $p$ and $q$, and then it would be correct to use $\psi$ to indicate mental states with the content $q$. But now we have a contradiction: $\psi$ is used correctly and incorrectly. Something has to give.31

Again we have a number of options: give up or reformulate the error platitude; give up or reformulate the meaning is use platitude. Neither option may be uncontroversial. It is hard to imagine how meanings can be robust if error about meaning isn’t even possible. It is hard to see how the error platitude can be reformulated because it is already rather weak: error just has to be possible. It seems the best option would be to reject the meaning is use platitude altogether, or to reformulate it.

In fact, there might seem to be an easy way out of the problem. I phrased the supervenience of meaning on use in terms of correct use: meaning supervenes on correct use. With that formulation in hand it should be easy to explain how error is possible: incorrect use of $\psi$ is simply when it is used out of accord with its correct use. There will be no temptation to lump the incorrect use in with the subvenient base of correct use, if we know it is incorrect. However, it is this kind of move which raises the suspicions about the meaning is use platitude: the question is on what basis the ‘incorrect use’ is deemed incorrect and the ‘correct use’ is deemed correct. If we identify the correct use of a sentence on the basis of its meaning, then we reverse the

---

31 This type of problem dominates Kripke 1982 (cf. e.g. p. 24), and the enormous literature it has spawned.
supervenience thesis: the correct use of a sentence supervenes on its meaning. This would of course offend against the principle of explanatory asymmetry: meaning supervenes on use, use supervenes on meaning. The crucial point then is whether there is any meaning-independent way to identify the correct use of a sentence, which doesn't at the same time entail the contradiction that $\psi$ is used correctly and incorrectly. Solutions to this problem are unlikely to be uncontroversial. The possibility of error is a contested issue in the philosophy of language, and there is not much consensus on what it would take to show that error is possible.\footnote{Cf. e.g. Kripke 1982, McGinn 1984, McDowell 1984, Wright 1980.}

c) The third tension: the problem of behaviourist and social use.

This is a tension between the meaning is use platitude and the platitude about first-person authority about meaning. Once again the problem stems from raising suspicions about the notion of ‘use’ in the meaning is use platitude. The suspicion is that ‘use’ must be construed such that there is no space for the appropriate first-person perspective associated with the platitude about first-person authority. There are two versions of the problem: one stemming from considerations about behaviourism and one stemming from social externalism.

i) the problem from behaviourist use.

The first version of the problem goes like this:

(I) S can have first-person authority about the meaning of $\psi$ without having to observe or infer anything from a third-person perspective on him- or herself. [First-person authority].

(II) Use of sentences has to do with overt behaviour, it has to do with what words come out of the speakers’ mouths on certain occasions; the only way to gain knowledge about overt behaviour is by adopting a third-person perspective on it (there is no difference between a putative first-person perspective and a third-person perspective because everything that is available
to the first-person is available in the same manner to a third-person observer). [Use is overt behaviour].

(III) Therefore, S cannot have first-person authority about the meaning of $\psi$ without having to obtain a third-person perspective on him- or herself.

This is really an adaptation of the old argument against behaviourism about the mental which accuses it of annihilating the notion of introspection of our own mental states.\(^\text{33}\) If behaviourism is committed to analysing mental states in terms of dispositions for overt behaviour, then our own access to our own mind is in principle no different from any other person’s access to our own mind (though of course I am better informed because I observe much more of my behaviour than others do). This offends against the idea that I know of my own mental states in a different way to others, in particular, a way that does not require that I observe my overt behaviour and/or draw inferences on the basis of such observation. In the above argument this is directed not against my state self-knowledge, but against my content authority.\(^\text{34}\)

It is worth noticing that the argument exerts pressure, firstly, on the idea that the authority must be accounted for in first-person terms, but also, secondly, on the idea that our use of sentences is somehow motivated by grasp of meaning. If all we can do is adopt a third-person perspective on use, then use begins to appear non-authorised—we become disengaged, merely observing ourselves using sentences thus and so. Consequently, if the argument holds up, then it has ramifications not only for the platitude about first-person authority, but also for the explanatory meaning platitude.

But why should the meaning is use platitude be associated with old style behaviourism? I think that the platitude is in fact not committed to behaviourism, but I acknowledge that it is a natural interpretation of it. As stated, it is neutral between

\(^{33}\text{Cf. Braddon-Mitchell and Jackson 1996, ch. 2.}\)

\(^{34}\text{I discussed those terms above in §2.e.ii.}\)
various ways of understanding the notion of correct use. Therefore, this argument at most implies that the formulation of the platitude must be shown to avoid behaviourist commitment, if the analysis is to remain consistent. This may not be easy or uncontroversial to achieve.35

ii) the problem from social use.

There is another version of problem, however, which does not accuse the meaning is use platitude of behaviourism. This argument builds on intuitions about social externalism, pioneered by Burge (1979). It goes like this:

(I) S can have first-person authority about the meaning of ψ in virtue of having a first-person perspective on him- or herself. [First-person authority].

(II) If meaning is use, then other speakers' use go into determining the meaning of ψ. [Social externalism].

(III) If other speakers' use go into determining the meaning of ψ, then what one can be authoritative about via a first-person perspective on oneself may fall short of being the true meaning of ψ.

Again, this is not something which the meaning is use platitude is automatically committed to, but it does represent a possible reading of it. The argument, then, says that if meaning supervenes on communal use, then first-person authority is endangered. The communal reading of the supervenience thesis is in fact quite natural. It seems quite natural to let the communal meaning of ‘cockatoos are precocious creatures’ supervene on the use of a community of speakers, even though one of the members uses the sentence in a slightly different way than the majority (S may fail to use the sentence to indicate beliefs about black cockatoos). But then we have the problem: if S has first-person authority in virtue of his or her own use of ‘cockatoos are precocious creatures’, then S fails to be authoritative about the communal meaning.

35The idea that a use-based theory of meaning, in particular, a dispositional theory, leads to problems for first-person authority appears clearly in Wright 1989, p. 236.
of that sentence. Notice that this only presupposes that the meaning is use platitude is consistent with social externalism. It doesn’t presuppose that the platitude entails social externalism. Other platitudinous commitments may lead us to refute social externalism, or we may allow mixed cases: some meanings are socially determined, some aren’t.36

Once again this problem is also a problem for the intuition that use is motivated by meaning, i.e., that a speaker, as it were, authorises uses of sentences. If what one means by a sentence depends on what other speakers do, then my uses cannot be authorised by me, nor any other particular speaker (unless we all defer to the same one speaker as authoritative, the Pope for example). And it is an open question to what extent it makes sense to say that a community of speakers as such authorises certain meanings.

This problem differs from the previous problem, about behaviourist use, partly in its motivation, partly in what it is a problem for. The former problem concerns the nature of the privilege of first-person authority: the privilege authority I have is seen as in principle no different from others’ knowledge of what I mean. In contrast, this problem doesn’t concern the nature of the privilege, but the very idea that authority about meaning is first-person. It would not solve the problem if we allowed that first-person authority could be achieved by adopting a third-person perspective on ourselves.

In response to this problem one must refute one, or both, of the involved platitudes, or somehow show that on the preferred readings of the platitudes the problem does not arise. This may take various forms: deny that social externalism is true, formulate social externalism and first-person authority so that they come out as consistent,
and/or formulate the notion of use so that the problem doesn’t arise. Again, in the light of the philosophical debates about the nature of first-person authority, the nature of use, and social externalism, neither of these moves seem clearly uncontentious.

\[\textit{d) Networked platitudes.}\]

The existence of the three tensions demonstrates networked character of the platitudes: one platitude may interfere with other platitudes, and we must expect that variations in the formulation of one platitude can entail quite wide ramifications for the remainder. Intuitively, it is clear that they must be networked. The communication platitude says that we use words to indicate mental states with certain contents. In so far as this has to do with our ability to express our thoughts in language, it has to do with first-person authority. Without first-person authority we must make sure that the communication platitude remains intact. In so far as the communication platitude has to do with successful communication it has to do with the possibility of error and robust disagreement: in order to communicate and understand a wide class of contents there must be constraints on what sentences indicate what contents (otherwise I will go to the bakery when you say “there is a hurricane on its way!”). Moreover, the communication platitude is itself formulated in terms of use: ‘sentences are used to indicate mental states with certain contents’. This shows that the conception of use which appears in the meaning is use platitude, and the explanatory meaning platitude, plays a role in our grasp of the communication platitude. Likewise, the explanatory meaning platitude hardly makes sense without the error and robustness platitude: we wouldn’t explain use in terms of meaning if there were no way a sentence could be used incorrectly, or if we saw no point in engaging in disagreements about meaning at all. The tensions show that whereas some of the interrelations serve to mutually support the platitudes, some interrelations may exert pressure on them.

\[\textit{\textsuperscript{37}Cf. e.g., the discussions in Analysis following McKinsey’s 1991 (e.g. Brown 1995, Miller 1997; also Boghossian 1997b and Davies, forthcoming); cf. Burge 1979 as well as Boghossian 1989b, Warfield 1992, and Ludlow 1995 for the debate as it relates to social externalism. Kripke mentions the problem too, cf. 1982, p. 23.}\]
4. Strategies for resolving the tensions.

I have presented three tensions among the five platitudes surrounding the concept of meaning. The first tension was between the meaning is use platitude and the explanatory meaning platitude; the second was between the meaning is use platitude and the error and robustness platitude; the third between the meaning is use platitude and the platitude about first-person authority. Clearly if these tensions represent real problems, then the status of the platitude-based analysis must be questioned. Can the analysis survive the existence of incompatible platitudes, or must the platitudes be revised in the light of the tensions, and can such revisions be uncontroversial?

I shall assume that the analysis cannot retain the platitudes while acknowledging that they are incompatible. This would betray the fact that our concept of meaning is a foil for inconsistent core beliefs. Of course some concepts may be like this, but if the cost of reformulating the platitudes to obtain consistency does not outweigh the cost of accepting inconsistencies in our core belief system, then we should surely go for reformulation.

There are two main strategies for resolving the tensions by reformulation. Firstly, one may reject one or more of the involved platitudes. A second strategy is to hold that the platitudes are compatible, but need to be revised. The first strategy is platitude incompatibilism, the second is platitude compatibilism. (Of course, it would also be possible to adopt a mixed strategy: reject one platitude and revise the remainder).

If we go for platitude incompatibilism we must decide whether one or more of the platitudes can be rejected, and, if so, which. Since the meaning is use platitude is involved in all three tensions the most economical move would be to reject it alone. However, one may just as well argue that the meaning is use platitude must be central
to the concept, since it is involved in all the tensions, and therefore opt to reject the three platitudes that are involved on the other side of the tensions. Since the platitudes are networked only a detailed analysis will show which is the most desirable solution.

Straightoff there is nothing wrong with offering an argument to the effect that a purported platitude doesn’t belong with the set of core platitudes. If we can argue that some platitude belongs in the core set we can argue that some platitudes don’t. But it must be clear on what basis arguments for inclusion or exclusion are put forward, and this basis must be in line with the aim and purpose of this type of analysis. Above I stated that the purpose of the analysis is to give an account of the extent to which human use is involved in determining facts about semantic relations as expressed by \( \varphi \text{ means } p \). The set of platitudes which forms the basis of the analysis must be identified on the basis of a conception of the functions we \textit{a priori} expect the concept to fulfil in our practices. Faced with a tension-ridden platitude we should therefore ask: would its exclusion allow the concept to perform the function in our practices we \textit{a priori} expect of it? If the answer is ‘no’ it shouldn’t be excluded.\(^{38}\)

In the motivations I have given for the five platitudes, I have tried to pre-empt the platitude incompatibilist who wants to reject one or more of the platitudes. I do not think the concept of meaning will be able to do its job if any of the five platitudes is rejected. I think the motivations are strong as they stand, but I have also admitted that we do not have infallible intuitions about the platitudes. There is however a further strategy for showing the incompatibilist wrong: because the platitudes are networked, it may be that rejection of one may influence the others. And I shall in fact argue, in Chapter 3, that one cannot reject the meaning is use platitude without rendering at least one of the other platitudes false. On the assumption that the incompatibilist rejects platitudes in order to retain other platitudes, it follows that one shouldn’t be an incompatibilist of this sort.

\(^{38}\text{Here I follow the general line in Jackson 1998, Lec. 2.}\)
Since I claim that no platitudes can be rejected without disturbing the functional role of the concept of meaning, I am committed to compatibilism. I must therefore revise one or more of the platitudes, without rejecting them, and show that the tensions do not arise on the revised readings. This revisionary approach may be philosophically controversial inasmuch as it excludes certain philosophical readings of the platitudes. Inasmuch as those excluded philosophical readings are reflections of how people ordinarily use the concept it is revisionary of those practices too.39

5. Summary remarks.

In this chapter I have analysed the concept of meaning, i.e., the concept of the relation expressed by "φ means p". The notion of platitude-based analysis I employ has two parts: part one is conceptual and consists in setting out the platitudes surrounding the concept in an a priori manner; part two concerns which property underwrites the concept. I will discuss part two in Ch. 4. I motivated the five platitudes that surround the concept of meaning: viz. communication, meaning is use, explanatory meaning, error and robustness, and first-person authority. Each platitude was motivated in such a way that it comes out as indispensable to the concept of meaning, and as neutral in various philosophical debates. I then set out three tensions among the platitudes: the "meaning is use" problem, the problems from irregular and infallible use, the problems from behaviourist and social use. Finally, I discussed what the appropriate reaction should be to the tensions. I opt for platitude compatibilism: the platitudes are compatible, but not as they stand. In the next chapter I shall argue that the alternative, platitude incompatibilism, isn't a viable option.

39Thus I think that the purpose of the analysis of the concept of meaning is consistent with offering revised surrogates of the concept in question, but only as long as those surrogates allow the concept to have the same, or nearly the same, functional role in ordinary practices, if not in philosophical debate. Here I take my cue from Johnston 1993, p. 108, who in a discussion of response-dependence identifies a position called the Revisionary Protagoreanism which comes to "[recognising] that our concept are not many of them response-dependent as they stand and yet [holding] that they need to be replaced with response-dependent surrogates".
Incompatibilist Approaches to the Tensions Among the Platitudes: Primitivism and Causal Theories

To suppose that someone on some particular occasion allotted names to objects, and that by this means men learnt their first words, is stark madness...if others had not used such utterances among themselves, from what source was the mental image of its use implanted in him? Lucretius, *The Nature of the Universe*

1. Introduction.

a) Against incompatibilism.

In the previous chapter I analysed the concept of meaning by setting out *a priori* five core platitudes for the concept. I called these platitudes: communication, meaning is use, explanatory meaning, error and robustness, and first-person authority. I demonstrated that there are three tensions among the platitudes: a tension between meaning is use and explanatory meaning, between meaning is use and error and robustness, and between meaning is use and first-person authority. If the tensions hold up then the concept is underpinned by incompatible platitudes. We must assume that we cannot just accept the incompatibility. That leaves two main options: incompatibilism, with a concomitant rejection of one or more of the platitudes; or compatibilism, with a concomitant revision (but not rejection) of one or more of the platitudes.
In this chapter I shall critically assess the incompatibilist approach. Since the meaning is use platitude is involved in all three tensions, the most economical incompatibilism would consist in resolving the tensions by rejecting this platitude alone. This is the version of incompatibilism I shall assess (the less economical approach is to reject the three other platitudes involved in the tensions, or all four, but I shall not consider this radical option). Obviously, I believe incompatibilism is a bad option: in the last chapter I have tried to motivate _a priori_ the meaning is use platitude as a core platitude about the concept of the relation expressed by "φ means p". But I cannot rely on this motivation in the argument against the incompatibilist without begging the question in favour of compatibilism.

However, I believe that the type of incompatibilism which rejects the meaning is use platitude can be refuted on independent grounds. No matter how incompatibilism is fashioned, it will offend against some of the remaining platitudes. The reason is that the platitudes are networked: take one platitude out, and there will be ramifications for the remaining set of platitudes. My arguments will show that these ramifications are detrimental to the incompatibilist strategy.

The argument shows, firstly, that incompatibilism of this sort cannot be used to resolve the tensions. Secondly, the manner in which I show this provides indirect support for the meaning is use platitude. Incompatibilism fails precisely because the meaning is use platitude is left out. This has added interest because it transpires that a particular version of the meaning is use platitude is needed to enable the remaining set of platitudes to do their job. This particular version of the meaning is use platitude is _dispositionalism_. Hence, the result is two-fold: not only do we show that incompatibilism fails, but the compatibilist is also given a suggestion, _viz._, that he or she should revise the meaning is use platitude in terms of dispositionalism.
CHAPTER III, INCOMPATIBILIST APPROACHES: PRIMITIVISM AND CAUSAL THEORIES

This is significant because dispositionalism is often seen as the best contender for fleshing out the "meaning is use" slogan.\(^1\) Dispositionalism seems attractive in this context because presumably dispositions can be accounted for in non-semantic terms.\(^2\) This also has the effect that, if a theory is consistent with the reduction of the semantic to the non-semantic, then it will be *prima facie* compatible with a dispositionalist use theory (whether or not such a theory could be shown to work). This point is important because then it can be shown that when an incompatibilist theory is consistent with reductionism, it will be *prima facie* consistent with re-introducing the meaning is use platitude.

**b) Primitivism and causal theories.**

In what follows I discern two main types of incompatibilism: primitivism and causal theories. These types mark different ways of rejecting the meaning is use platitude. I concentrate on these types because they come close to what some people have actually tried to defend, as I shall indicate who below.

Primitivism is the doctrine that there are facts about meaning, that is, there are facts about the semantic relations between sentences and meanings, but these relational facts are primitive. To say that the facts about semantic relations are primitive is to say that these facts are *sui generis* and not reducible to non-semantic facts and, *a fortiori*, not reducible to facts about use. It is also to say that facts about meaning are not dependent in any interesting sense on judgements about meaning and, *a fortiori*, not dependent on how we use judgements about meaning.\(^3\) Since primitivism holds that facts about meaning are independent of facts about use, it rejects the meaning is use platitude.

---

\(^1\) *Cf.* e.g. Kripke 1982, ch. 1.

\(^2\) Dispositionalism has been put under pressure from Kripke 1982. It is this type of criticism which underlies the tension between the meaning is use platitude and the error and robustness platitude.

\(^3\) Here I follow Boghossian 1989, who comes close to defending primitivism, albeit for mental content. He says 'robust realism [about meaning is the doctrine that] judgements about meaning are factual, irreducible, and judgement-independent.', p. 547. Apart from Boghossian, only a few other philosophers have endorsed forms of primitivism. Jerrold Katz’s 1990 ‘new intensionalism’ is some kind of non-natural Platonism. Goldfarb 1985 also seems sympathetic to primitivism, *cf.* pp. 474ff.
Causal theories are somewhat more complex. They likewise hold that there are facts about meaning, but they think that these facts are indeed reducible to non-semantic facts, viz., facts about causal encounters between words and what the words stand for. This amounts to a rejection of the meaning is use platitude, as I stated it, in that it says that meaning isn’t dependent on contemporary use to indicate mental states with certain contents. That is, it says that facts about semantic relations can depend on past encounters with non-mental things.4

I begin my argument against incompatibilism by considering various versions of primitivism. They are all found wanting. Either they fail to account for the remaining platitudes, or they import the meaning is use platitude (often in dispositionalist form) and therefore cease to be primitivist. I then pass to a discussion of causal theories. There the details are different, but the overall structure the same: causal theories fail to make the remaining platitudes come out true, or begin to re-introduce the meaning is use platitude.5

2. Contingent Primitivism

a) A Moorean parallel.

The first brand of primitivism I will consider is contingent primitivism which can be described like this: which meaning is assigned to a sentence is not dependent on any

---

4Devitt and Sterelny 1987 canvass such a theory. In fact a causal theory can be interpreted as a compatibilist theory in as much as 'causal encounters between words and referents' may be interpreted as a revision of the meaning is use platitude rather than as a rejection of it. As we shall see however, causal theories in effect also reject the explanatory meaning platitude.

5Notice that the arguments do not turn on the possible queerness (cf. Mackie 1977, p. 38–42) of the facts about meaning posited by the primitivist and causal theorist. I acknowledge, for the sake of argument, that there could be such facts, but I deny that they could fulfil the role they were intended to fulfil.

Under some readings McDowell’s 1994 ‘re-enchanted’ world seem primitivist. Some deflationists about truth and reference, such as Schiffer and Johnston, also seem primitivist. Schiffer 1987, ch. 10; Johnston 1988. Horwich 1995 is an example of dispositionalist deflationism about meaning. It is interesting to notice that Boghossian ends up defending a type of primitivism in despair over the failure of use-based theories to account for error.
facts about the use of the sentence, for it is a primitive fact that this meaning is related to that sentence. We can accept, for the sake of argument, that there could be such primitive facts, but then ask how could the existence of such facts help to explain the remaining platitudes about meaning?

For the sake of argument we can also allow that the Contingent Primitivist’s facts make occasional error possible, and we can also grant the primitivist an epistemology to come with the facts. Otherwise primitivism wouldn’t get off the ground in the first place. Thus the platitude that occasional error be possible and recognisable is assumed to be satisfied in some primitive way.

But even allowing this much, primitivism still fails to account for the aspect of practicality I associated with the platitude about error and robustness. We have in effect a position which is somewhat similar in spirit to G.E. Moore’s brand of primitive moral factualism. He argues that goodness is a primitive, non-natural and indefinable quality. What is good and what isn’t is knowable by a faculty of intuition. Thus there are primitive facts which make a distinction between good and bad action possible, and which are such that agents can have knowledge about what is good and what is bad. The problem with Moore’s account is that though it might be that by intuition we can obtain true beliefs about the good, the account fails to explain why those beliefs should motivate action. Why should we care about the primitive good? It seems there is nothing odd about refraining from feeding Fido, even though one knows feeding Fido would be a good thing to do. Moore’s account doesn’t make sense of the practical aspect of morality.

---

6Cf. e.g. Moore 1903 §§6-10, 26. This is Moore without his claim that the good in effect supervenes on the natural. I discuss supervenience and primitivism in §4.f. below.

7This type of objection to Moore is inspired by Smith’s 1994 so-called ‘moral problem’.
The point of this brief analogy with Moore is that it helps to bring out why contingent primitivism can't satisfy the platitudinous requirement about the practical aspect of meaning. This requirement is motivated like this: it is part of the notion of error that recognition of error has potential practical consequences. To be credited with understanding, say, 'there is a fine magpie', a subject must be willing to withdraw erroneous applications of it (say in cases where she calls a crow a 'magpie'). It will be rational for her to respond by withdrawing her application when she is reproached by her peers.

But the present account of primitivism leaves out this practical aspect: the fact that a subject comes to have a true belief about what was an error and what wasn't, doesn't indicate why, even if she believes the reproach is justified, she should feel compelled to withdraw any erroneous utterances. According to this primitivism, there would be nothing odd about refraining from withdrawing an erroneous utterance, even if one knows it is erroneous.

The contingent primitivist can respond to this quite easily. She can simply argue that acknowledging that an utterance was an error necessarily comes with acceptance that the utterance be withdrawn. That is, the primitivist can simply argue that it is characteristic of primitive meaning-facts that they bear a necessary connection to the subsequent actions of competent speakers. Accordingly, we let it be part of the primitive meaning-facts that they are linked to subsequent use, i.e., that they make the subject withdraw utterances when reproached. The meaning-facts can be conceived as simultaneously fulfilling two roles: making error possible and making the subject withdraw erroneous utterances.

This version of contingent primitivism can now be phrased like this: the primitive meaning-facts ensure that the subject knows the meanings of her sentences, and by thus knowing meanings she is disposed to respond in certain ways when she comes to
recognise facts about error. If simply knowing meanings in this way compels the subject to undertake the appropriate revisions of her use, then it seems straightforward to say that the primitive facts dispose her in that way. But now, of course, it seems we cannot distinguish primitivism from a kind of primitive dispositionalism.

In order to satisfy the platitudinous requirement concerning practicality, the meaning-facts must be of a certain ontological kind: they must be factual and at the same time be linked to subjects' practices. That is the kind of role dispositions are normally held to have, and so primitivism becomes indistinguishable from dispositionalism. If the incompatibilist opts for this move then he or she is no longer an incompatibilist because a version (in particular, a dispositionalist version) of the meaning is use platitude has been re-instated.8

b) A Lockean parallel

The primitivist might want to resist this line of argument, viz. by holding that it exaggerates what it takes to satisfy the requirement of practicality. She can claim that there is no need for anything other than the following: S knows that $\varphi$ doesn’t mean $q$, and so, if she were to use $\varphi$ to mean that $q$, she would be making a mistake, and it would be rational of her to withdraw. In this way the primitive facts are not conceived as dispositional, but rather as external sui generis facts which we, in some unexplained way, manage to aim at.

If the primitive meaning-facts are sui generis and totally independent of use, then nothing prevents the possibility of a very unwelcome scenario: global error. We found, as we should, that according to this kind of primitivism, for any given application of a term or concept, possibly, the subject, S, is in error. But it follows, if

---

8But couldn’t there be a primitive dispositionalism? That is, couldn’t one hold that dispositions can be underwritten by counterfactuals with no categorical basis? Well, yes, but that wasn’t the problem about dispositionalism in the first place. The problem was that dispositionalism cannot accommodate the error platitude and a primitive dispositionalism doesn’t address that problem.
meaning-facts are *sui generis* in the above sense, that, possibly, on all occasions S, and all other subjects, are in error. Nothing in this primitivist position prevents this result. But this shows that something has gone wrong for this kind of primitivism, because when we are talking about our concept of meaning this radical kind of global error should not be possible. The point here is not that actual correctness should be guaranteed, but that the possibility of communication should not be *incidental* to the concept of meaning. If we construe the concept of meaning such that it is *just* a lucky coincidence if we manage to communicate, then something must have gone wrong. This is reflected in the fact that it matters greatly to us that we use sentences such that our language can be used for communication in various ways. Whether, and to what extent, that use is in accordance with a primitive semantics is not relevant. The question simply boils down to this: why should we *care* about there being a primitive semantics about which it is possible that we are all wrong all the time. Perhaps we should care about primitivist *values*, if there are any, but a primitivist *semantics* is utterly uninteresting.9

Perhaps it is useful here to think of Locke's attack on the Cartesian notion of personal identity as consisting in an underlying non-material mental substance. For Locke personal identity is a 'forensic' notion, and he thought that the Cartesian idea of personal identity gives us no way of controlling our normal ideas of personal identity in the situations we care about, *e.g.*, the attribution of responsibility for past action.10 Similarly, the brand of primitivism under consideration is too divorced from what we care about, *e.g.*, transmitting information in communication.

Let me give an example that shows why allowing global error is problematic, and why this kind of primitivism cannot do anything about it. Imagine two isolated

---

9 The question of global error has been discussed in Fricker 1983, p. 50. Fricker sets it up as an absurd consequence of 'semanticalism' which is a position akin to primitivism.

10 Locke 1690, Bk. 2, XXVII, §§6-26, especially §18.
communities, C and C*, which both have linguistic practices with the sentence 'there is a kookaburra'. In C the use of the sentence is almost always in accordance with some *sui generis* primitive meaning-fact, although C-speakers make occasional errors. In C* the use of the sentence is always out of accordance with the primitive meaning of the sentence. According to the kind of primitivism under consideration, it would be rational for the speakers in C* to revise their use of 'there is a kookaburra' because it is out of accord with the primitive meaning-facts. But this claim seems utterly implausible. It is very easy to imagine a scenario where it would not be rational for the speakers in C* to change their practice, even though it is out of accord with the primitive meaning-fact. Say the primitive meaning-facts determine that 'there is a kookaburra' should be used for all and only kookaburras, but that the C* community uses the sentence for kookaburras plus a couple of very closely related species of birds, perhaps because those kinds of birds all taste deliciously and it is useful to have a way of talking about them. In that case it would be irrational for the C* speakers to begin to change their linguistic behaviour for it would make it more difficult for them to trap and eat what they call 'kookaburras'.

In sum, this kind of primitivism is certainly *possible*, but it has implausible consequences for our understanding of the concept of meaning. Of course, there is a way to remedy the situation exemplified in the case about C and C*. The primitivist could admit that what is rational, and what isn't, is connected to how the sentence is used in C and C*. But this kind of move immediately turns the account into a use-based account because it presupposes that there is a substantial story about how use determines meaning.
3. Back to basic primitivism.

a) Paring down the doctrine

In the previous section one version of primitivism was scrutinised. It appeared that substantiating primitivism in certain ways exposed it to roughly anti-Moorean and pro-Lockean considerations. It ends up as either dispositionalism or uninteresting Platonism. So the primitivist needs to pare down the doctrine and start again. Let us accordingly take as the starting-point a more mundane primitivism where the meaning-facts are simply characterised as irreducible to the non-semantic physical/functional facts about use.11

In order to be in a proper contrast to the previous brand of primitivism, this basic primitivism is not allowed at the outset to satisfy the requirement that occasional error be possible. This proves crucial for basic primitivism because it is very hard to see how, without being amended, it could come to satisfy this requirement. Let me explain why.

b) The problem for basic primitivism

Basic primitivism's problem with accommodating error derives from the debate about rule-following. One of the central doctrines in that debate is that whatever is posited as the meaning-fact for an expression or concept $\varphi$, must allow us to distinguish between that to which $\varphi$ correctly applies, and that to which it doesn't. It just doesn't make sense to say that someone "follows" a rule unless it is possible for her not to follow it. The mere possibility of error will be in place if our account of meaning allows that, on any occasion of application, by a subject S, of an expression or concept $\varphi$, possibly, the application of $\varphi$ is out of accord with $\varphi$'s meaning.

---

11 In addition we can allow the facts to cover an indefinite number of applications of the sentence (e.g. the meaning-fact for 'that is a dog' covers all possible applications of that sentence). Hence we sideline the issue about the finiteness of the mind.
How is this accomplished by basic primitivism? Presumably like this: When S says 'Fido is a dog', then some primitive fact entails that the application of 'Fido is a dog' has certain correctness-conditions, to the effect, e.g., that 'Fido is a dog' means *Fido is a dog* on that occasion. Crucially, it will follow that if S says 'Fido is a dog', and 'Fido is a dog' means *Fido is a dog*, then, if Fido is a dog, S has said something true. On the other hand, if S says 'Fido is a dog' and still means *Fido is a dog*, but does it in a case where Fido is a cat, then S has said something false—she has said that the cat is a dog.

Now consider another situation. Grant that, for basic primitivism, meaning-bearers will get their meanings assigned *contingently* by the primitive meaning-facts. (For example, it will not be an essential property of the expression 'Fido is a dog' that it means *Fido is a dog*). Reflect that this makes the following situation possible: S says 'Fido is a dog', and some primitive meaning-fact entails that 'Fido is a dog' means *Fido is a cat* on that occasion, and Fido is a cat. In this situation, according to primitivism, S has said something meaningful and true. But, now, what could possibly distinguish this case from the case where 'Fido is a dog' in S's language means *Fido is a dog*, and where she consequentially said something false? There is nothing in *this* primitivist doctrine that could ensure that, when we want to say that S said something false, she didn't really just shift languages and in fact say something true.

As a consequence, the requirement of occasional error will not be satisfied. On each occasion of application it will be indeterminate, according to basic primitivism, whether S got it wrong or right. Clearly, basic primitivism must be amended, otherwise it will not give us an account of meaning. But such amendment had better not repeat the mistakes of contingent primitivism.
4. Essential Primitivism.

The only plausible kind of move which doesn't lead back to contingent primitivism is this: meaning-bearers get their primitive meanings assigned essentially by the primitive meaning-facts. That is, take basic primitivism, but add that a given meaning-bearer cannot be assigned another meaning and remain the same meaning-bearer. Adopting this kind of primitivism prevents the alternative story from being told. It will not be possible for the same sentence to have an alternative meaning.

a) Problems for primitivism with essential relations.

Primitivism with essential relations seems implausible for linguistic meaning. Of course the expression 'Fido is a dog' could have had another meaning than it actually has in English. So the primitivist presumably must move to mental content and claim that a content-bearer has its content assigned essentially by primitive content-facts. Prima facie essential relationality seems much more plausible for mental content.12 But the move will only work if mental content really can be essentially related to its content-bearers by primitive content-facts, in a way palatable to the primitivist's other commitments. I will suggest that none of the suitable candidates for mental content are compatible with essential relationality. The pattern of argument will be to show on a case-by-case basis that primitivism cannot both have essential relations and be non-reductionist. The general problem is that, on the one hand, the kinds of facts which could ensure essential relationality seem to be the kinds of facts which do not come apart from the physical/functional facts, and, on the other hand, in the cases where we allow the meaning-constituting facts to come apart from the physical/functional facts

12The move to mental content is favoured by Boghossian: 'A plausible anti-reductionism about meaning would not wish to deny that there is an interesting story to be told about the relation between linguistic content and mental content; what it maintains, rather, is that there is no interesting reduction of mental content properties to physical/functional properties.' Boghossian 1989, p. 541. Notice that Boghossian isn't discussing this in relation to a distinction between contingent and essential relationality. Cf. also Boghossian 1990, p. 172-3.
we are given no reason to believe that essential relationality is in place. The first horn of the dilemma shows that primitivism with essential relations is *prima facie* consistent with re-introducing the meaning is use platitude in a dispositionalist form; the second horn shows that this sort of primitivism flouts the remaining platitudes.

**b) Twin-Earth cases and externalist content.**

Here is what the theory couldn't be. It couldn’t be compatible with one interpretation of Putnam-Burge style anti-individualism. As Burge rightly points out, Twin-Earth cases concern mental content as well as linguistic meaning. Mental contents themselves are to be understood in terms of S’s interaction with her environment and other subjects. In one situation the concept φ has, say, the content *dog*, in another it has the content *dogs-or-wolves*. The semantic relation between the bearer of the mental content and the content it is assigned is therefore contingent, not essential. Consequently, mental concepts susceptible to the implications of Twin-Earth cases are not related essentially with their contents (at least when the Twin-Earth cases are given a fairly weak interpretation).

**c) Primitive informational content - the disjunctive view**

The above objection to primitivism with essential relations works on the assumption that the primitivist isn’t averse to the idea that when something is a contentful state something else is the content-bearer. But the primitivist who believes in *strong* externalism could deny this. She could claim that the Twin-Earth examples show that it is contingent which content a subject ends up having, but hold that that content itself is essentially related to its content-bearer. Thus, in response to the above problem about externalist content, the primitivist could go the other way, and opt for some kind of object-dependence (perhaps employing a notion of primitive informational content from an object). In this way, seemingly, there is essential relationality because a

---

13 Burge 1979. See Ch. 2 §2.b.iii. for examples of Twin-Earth examples.
content-bearer is at least partly identified by its content. One way to capture the overall structure of such a view is by way of a disjunctive notion of perceptual knowledge: either the subject has direct cognitive contact with the object itself, in which case the experience is veridical; or the object of experience is a mere appearance.\footnote{Cf. McDowell 1988, p. 211.}

But what is error on this view? Since there is partial identification between bearer and content it seems there will be no error whenever there is direct cognitive contact between subject and object. Hence error must be what occurs at the other disjunct, \textit{i.e.}, in the appearance situation where there is no object to think about, or where the thought is founded on misinformation. But there are two ways such a theorist could characterise the appearance-situation. Either no thought is entertained,\footnote{As in Evans 1982 (chs. 4, 5, 9). There is no thought to be had if there is no object or if the thought is based on misinformation (such that the wrong object is discriminated).} or it is another thought with the appearance itself as content. If primitive informational content is all we have to judge the alleged error-scenario by, then nothing rules out that what is presumed to be a no-thought case really was a case with another thought having the appearance as content. Therefore we get this problem: was the so-called error (\textit{i.e.} the no-thought-case) really an error, or just the correct (\textit{i.e.} veridical) entertaining of another thought? From the perspective of this brand of primitivism, there is no distinction between the cases.

This is an objection to the purported essential relations between bearer and primitive content. What object-dependent primitivism tells us is that \textit{once} a concept is partly identified with a content, then we should conceive of thoughts involving the concept as object-dependent. But in order to establish essential relations it has to be explained—rather than presupposed—how a concept gets to be partly identified with this rather than that content. Since such an explanation is not forthcoming, the alternative story can be told in the case of purported error. Hence the mentioned partial
identification of the bearer with its content is not the kind of relationality which is needed to satisfy the requirement that occasional error be possible.

d) Conceptual role semantics.

In response to the two previous objections it seems primitivism with essential relations should discard externalist content and go individualist. One way to get essential relations in place would be to opt for a kind of individualised conceptual role semantics (CRS). A given content-bearer $\varphi$ has the content $p$ in virtue of $\varphi$'s conceptual role; were it not for that role, then $\varphi$ wouldn't be $\varphi$. This would give us essential relations. There is no way $\varphi$ could have had another content. Unfortunately, this kind of theory is not straightoff compatible with primitivism because CRS gains credibility precisely because it is a reductionist use-based theory: it is basically a sophisticated dispositionalism.16

What, then, about a non-reductive CRS? Well, in effect it makes no real difference to go non-reductive (as long as we don’t mirror the moves of contingent primitivism). Conceptual role semantics as such is faced with one paramount problem: what is the distinction between the basic, meaning-constituting, conceptual roles and the non-basic, non-constitutive roles? This is familiar,17 but here I want to phrase the problem in terms of essential relationality. So the question is: what is occasional error according to CRS? The answer must be this: it is when a concept is used out of accord with its meaning-constituting conceptual role. The problem is that nothing distinguishes an alleged error from the correct use of a concept in accordance with another conceptual role. This problem is analogous to the problem which basic primitivism was shown to have, so it seems CRS is simply a poor candidate to avoid

16 Cf. Boghossian 1993, p. 30, for an argument that CRS must be reductive to avoid gross circularity.

that problem. The move to a non-reductive CRS does nothing in itself to resolve this problem.\textsuperscript{18}

At any rate, the kind of solution required to overcome the problem is one which will make it determinate which role or use identifies the meaning of the concept in question, and which role or use doesn't. A primitive CRS will be quite pointless if it cannot deliver this kind of solution. But remember that this was precisely the type of solution which dispositionalism required (and which contingent primitivism ended up providing in its particular setting). The upshot, then, is this: if you think conceptual role semantics works, then you should also think some kind of dispositionalism works. Therefore this positions ends up re-introducing the meaning is use platitude and so ceases to be incompatibilist.

de) Individualised content primitivism.

Rather than opt for object-dependence or CRS we could attempt to formulate a theory of 'individualised' content modelled on something like Evans' class of non-Russellian descriptive names. This would give us a general class of descriptive contents,\textsuperscript{19} for the understanding of which no knowledge of which world is really the actual world is required.\textsuperscript{20} For the sake of argument we could grant that this theory is non-reductionist, but then we ask: is there any reason to think it has essential relations between content-bearers and contents?

Such a primitivist should hold that non-Russellian descriptive concepts have their narrow contents assigned in virtue of primitive content-facts. But does it really have

\textsuperscript{18}As I think Boghossian also indicates, 1997, p. 353.
\textsuperscript{19}With descriptions including causal encounters and linguistic information such as e.g. "the watery stuff of our acquaintance" or "the x called 'Julius'".
\textsuperscript{20}I should pause to notice that it certainly isn't non-reductionist in spirit (indeed, it can be tailor-made to suit physicalism), and it also seems able to incorporate some of the intuitions behind semantic externalism (viz. by rigidifying the descriptions), cf. e.g. Frank Jackson 1998. Moreover, if such contents are modelled along the lines of Kaplan's 'characters', then they themselves are not narrow. Cf. e.g. Kaplan 1978.
essential relations? Again we ask: what is occasional error? The answer must be that error is when a concept is employed out of accord with its descriptive content. But then we get the problem of whether the alleged error really was an error, or just a correct application of another concept. The reason we get this problem is that there is always going to be an alternative descriptive content to fit an alleged erroneous application. It seems essential relations cannot be sustained and that the requirement of the possibility of error therefore cannot be satisfied. When considered as an acolyte to primitivism this kind of theory doesn’t work.

f) Supervenience primitivism.

Lastly, consider a very general form of primitivism which embraces non-reductive supervenience such that primitive meaning-facts supervene on, but are irreducible to, physical/functional facts.21 This could be a method of allowing the physical/functional facts to have a place in the theory without relinquishing primitivism. Of course, this move depends on the outcome of the as yet unresolved debate about the possibility of non-reductive supervenience, and I shall therefore not deal with the question of supervenience primitivism in sufficient detail here.

I will note, though, that this debate shares some features with the dilemma-structural of my arguments above. In general, there are three types of supervenience-thesis: weak, global, and strong supervenience.22 I think that the supervenience primitivist faces a dilemma concerning the choice of supervenience thesis. On the one hand, if she chooses a weak or global supervenience thesis, then the modal force of the theory will be so weak that the semantic relations between contents and content-bearers will

---

21Perhaps this is more in accordance with Boghossian, cf. 1989, 549.
22Here is a brief gloss on them: Weak supervenience: in any possible world, if x and y are P-indiscernible in that world, then they are M-indiscernible in that world. (‘P-indiscernible’ stands for “indiscernible in the set of physical properties”, ‘M’ stands for “mental”). Strong supervenience: for any individuals x and y, for any possible worlds w₁ and w₂; if x in w₁ is P-indiscernible from y in w₂, then x and y in those worlds are M-indiscernible. Global supervenience: any two P-indiscernible possible worlds are M-indiscernible. Cf. Kim 1994a, p. 154ff. Also Kim 1994b, p. 579-80. For global supervenience see Kim 1994a, pp. 82-90, 154-55, 276-78.
be contingent (shown, for example, by the fact that we can run Twin-Earth cases for such theories). This indicates straightoff that weak and global supervenience primitivism is susceptible to the problem I raised for basic primitivism. On the other hand, if she chooses strong supervenience, then there will be strict psychophysical laws between the subvenient physical facts and the supervenient semantic facts, and, as Kim has shown, this entails reduction.\(^{23}\) This indicates that strong supervenience primitivism cannot be non-reductive.

So, supervenience primitivism is faced with the same kind of dilemma as the other types of primitivism: if its semantic relations are too weakly constituted, then it ceases to cope with the platitudes about meaning, but if the semantic relations are more strongly constituted, then it ceases to be primitivist.\(^{24}\)

I will not go through any more candidates for mental content. The structure of argument would repeat itself. In the cases where we are given reason to think that there could be essential relations we also get potential reductionism. And in the cases where we grant that a position could be non-reductive we are given no reason to think that there are essential relations.

\(^{23}\)Cf. Kim, \textit{op. cit.}

\(^{24}\)I think Van Cleve 1992, p. 361 suggests that Kripke’s worries about \textit{sui generis} meaning-facts (Kripke 1982, p. 50ff) are underpinned by similar kinds of worries about supervenience. The debate about non-reductive supervenience is far from closed, see \textit{e.g.}, Lepore and Loewer 1987 and 1989. But see also Leiter and Miller 1994, Kazez 1995, and especially Kim 1995 who gives, I think, many arguments that support my general line. Lepore and Loewer suggest a global supervenience thesis such that mental causation is weak, \textit{i.e.} underwritten by counterfactuals, not strict laws. Kim notes that if strong causation entails weak causation, then we have overdetermination, and if strong causation doesn’t entail weak causation then we violate the closure of the physical. My own general view is that mental causation is best dealt with by a functionalist approach, \textit{e.g.}, in terms of Jackson and Pettit’s 1988 notion of programme causation. But note that this notion is compatible with reduction.
5. Causal theories.

a) Primitivism and causal theories.

There is another class of very reductive theories of meaning which have some of the same problems as primitivism, but which are driven by different considerations. These are what I will call purely causal theories of meaning. According to such theories the semantic relations between expressions and meanings are constituted by causal links between words and their referents.

Such theories qualify as incompatibilist, in my sense, because they are purely causal: their causal focus excludes the idea that meaning is determined by contemporary use to indicate the contents of mental states. Though causal theories are reductive, and therefore quite different from primitivism, they have one crucial element in common with primitivism. They do not allow meaning to be dependent of use in the sense that when contemporary use changes then meanings change. They do have an aspect of use, but they make a mockery of it because once the causal link has been fixed, the aspect of use fades out and therefore facts about use and facts about meaning may begin to come radically apart. I shall argue that this aspect proves detrimental to causal theories. The problem is that causal theories only pay lip-service to the meaning is use platitude and thus don’t acknowledge that the platitude is networked with the remaining platitudes. As before, my argument will not beg the question in favour of the meaning is use platitude, but will rely on some of the other platitudes we associate with the concept of meaning.

b) Pure causal theories.

Devitt and Sterelny (1987) have canvassed what I call a pure causal theory of reference. Though they find occasion to amend it themselves, I will take their version of the pure position as my example.25

25It is not easy to find examples of pure subpersonal causal theories. Perhaps Adams and Fuller 1992 are as close as Devitt and Sterelny.
According to their theory, the reference of a word is fixed once and for all in a causally described *grounding* incident:

The name is introduced ostensively at a formal or informal dubbing. The dubbing is in the presence of the object that will from then on be the bearer of the name. [...] As a result of this causal interaction, a witness to the dubbing, if of suitable linguistic sophistication, will gain an ability to use the name to designate the object. Any use of the name exercising that ability designates the object in virtue of the use's causal link to the object: ostension of the object prompted the thoughts which led to the use of the name. In short, those present at the dubbing acquire a semantic ability that is causally grounded in the object.26

After the grounding, further understanding is made possible by virtue of a causal theory of reference *borrowing*:

People not at the dubbing acquire the semantic ability from those at the dubbing. This acquisition is also a causal, indeed perceptual, process. The name is used in conversation. Hearers of the conversation, if of suitable linguistic sophistication, can gain the ability to use the name to designate the object. The exercise of that ability will designate the object in virtue of a causal chain linking the object, those at the dubbing, and the user, through the conversation.27

The *sense* of a word is fixed by the causal link in the grounding incident: give the reference of the grounding incident, and the sense will be fixed too.28

Reference and understanding are explained in purely causal terms, and hence there is no need to posit any kind of cognitive access to meanings. The speakers of a language need know nothing about the grounding incident, and they need know nothing about

---

26 Devitt and Sterelny 1987, p. 55-56.
27 Ibid.
28 Ibid.
the object they purportedly are referring to in virtue of the grounding incident and the reference borrowing: it suffices if there are the mentioned causal links.29 This means that description theories of reference can be disposed of. This is an important result for Devitt and Sterelny because they have independent reasons for believing that description theories are inadequate.30 And this is what most defenders of causal theories will have in common: they think that cognitively based theories, such as description theories in their various forms, are unable to fix reference.31 I will not go into these arguments here, instead I will consider the causal theory in its own right.

c) Problems for the causal theory.
I will mention three points on which the causal theory offends against our platitudes about the concept of meaning, and point out what may be a common source for these offences. Then, in the next section, I will discuss a problem for the theory which Devitt and Sterelny themselves raise.

The first platitude the causal theory offends against is the communication platitude: we normally use sentences to tell each other how we take things to be, or how we want things to be. If we operate with a concept of meaning which doesn't allow this trivial aspect, then we can be sure we are wrong. But how can a causal theory get this right? It may indeed be true that I am in a position to say that "Bohr is a fine scientist" because there is some causal link between me and the original dubbers of Bohr, and Bohr himself. But that fact doesn't even begin to capture what it is for me to successfully communicate my beliefs about Bohr to you, or what it is for you to understand my words. I may have no idea about that causal link, but how can I transmit anything in communication if I don't know anything? On the other hand, maybe I do have some kind of knowledge of the causal link, but that is quite

---

29Devitt and Sterelny 1987, p. 63.
30Devitt and Sterelny 1987, ch. 3.
31This debate was initiated largely by Kripke 1980.
consistent with me not knowing that the link concerns Bohr, the guy I am trying to talk about (perhaps I used to play with his nieces and nephews in the garden in Copenhagen and knew of their uncle Niels). And yet I manifestly manage to transmit trivial beliefs about Bohr, irrespective of whether I know of the causal link or not. Nothing in the causal doctrine explains this fact.

This leads to the second worry which concerns the platitude about first-person authority. Putting aside the question of how I let others know how I take things to be, how do I myself know what I mean? It is a platitude about meaning that we have some kind of first-person authority about meaning. Otherwise it is hard to see why we should be taken seriously in communication about anything. This authority need not be full-blown infallible Cartesian authority: suffice that it is the kind of authority which is very easy to come by, if one is a language-user, and which can be obtained without our having to investigate the environment. It is not clear how the causal theory can make sense of this aspect. If I don’t know of the causal link, then I don’t know the sense and reference of what I say. If I do know the causal link, then we are owed an account of how this causal knowledge can be obtained without having to investigate the environment.

We are also owed an account of why this knowledge is indeed semantic knowledge. What is special about knowing a causal link between words and objects, as compared with knowing the causal link between two objects, or between a word and any object which is not its referent? Our causal interactions are manifold and we need to know which one of them is relevant for constituting semantic relations. Now, we can begin to do this by identifying the functional role which incorporates our use of meaningful expressions, and then deem the causal realiser of this role the relevant causal link. But we have no reason to think that this will yield a causal theory: the functional role for the concept of meaning surely involves the notions of first-person authority and
transmission of knowledge in communication, and I have just explained why these notions cannot be captured by purely causal links.

Of course, we may adhere to the letter of the quoted definitions—it concerns only subjects who are ‘of suitable linguistic sophistication’—and use that as a leverage to begin to distinguish between all our countless causal interactions. Perhaps that would be the beginning of a more adequate account. But now the suspicion will be that the notion of ‘suitable linguistic sophistication’ is really a cover for ‘knows the meaning of the sentence in question’, and hence the causal theory will presuppose an account of meaning, rather than be giving one.

The third notion which subpersonal theories offend against is the platitude that sometimes we can explain use in terms of meaning. This is a fairly trivial point: there are a lot of ordinary occasions where we may wish to explain someone’s use of a word by referring to its meaning, rather than to how the relevant matters of fact are. We can explain someone’s use of the word ‘billabong’ by saying “she said ‘there’s the billabong’ because that sentence means there is the waterhole”. But we do not explain like this in terms of causal chains that we may or may not know anything about. If we don’t know the causal chain, then the explanation is not available at all. If we do know the chain, on the other hand, it is still hard to see how the desired explanation is forthcoming: why is it the relation between billabongs and the use of the word ‘billabong’ which matters, and not one of the other countless causal relations involving ‘billabong’?

The above remarks are not intended to suggest that there are none of the causal links adverted to by the causal theory. There might well be. But the remarks do suggest that these links are irrelevant for the concept of meaning: they do not help explain what it is the concept does for us. This is of course consistent with accepting that the theory of meaning needs a causal element. So I concur with Evans when he says:
I would agree with Kripke in thinking that the absurdity [of the description theory] resides in the absence of the causal relation between the item concerned and the speaker. But it seems to me that he has mislocated the causal relation; the important causal relation lies between that item’s states and the speaker’s body of information—not between the item’s being dubbed with a name and the speaker’s contemporary use of it.32

d) The qua-problem and the possibility of error.

Devitt and Sterelny themselves raise a problem for the pure causal theory, which they call this the qua-problem.33 Consider the causal chain going backwards in time from an utterance of the word ‘Fido’. Assume it ‘ends’ in Fido the dog. The qua-problem is this: in virtue of what property of Fido does ‘Fido’ come to mean what it means? Is it in virtue of the whole object, Fido, or in virtue of part of Fido, or everything else but Fido? Nothing in the causal story seems to deliver an answer to this question. And the problem is worse,34 because it is likewise uncertain how the causal story determines a non-arbitrary end to the causal link: which considerations in purely caused terms could determine which situation in a never-ending causal chain is the original ‘dubbing’?

The problem is not that there are no facts of the matter about these two issues: we know that we mean the whole of Fido by ‘Fido’, and that the dubbing happened when we got him as a little puppy. The problem is that it is hard to see how any purely causal story could provide such answers. Devitt and Sterelny acknowledge this:

In virtue of what does the grounder intend the whole object? It seems that the grounder must, at some level, “think of” the cause of his experience under some general categorical term like ‘animal’ or ‘material object’.35

33Devitt and Sterelny 1987, p. 63-65; 72-75.
34As is remarked by Braddon-Mitchell and Jackson 1996, p. 68f.
35Devitt and Sterelny 1987, p. 64-65.
That is, they argue that the *qua*-problems may be solved by amending the purely causal story with something the grounder thinks on the dubbing occasion. This seems intuitively correct: it will be settled whether ‘Fido’ refers to Fido or a part of Fido if we associate the property of being a whole animal with the name.

Now a question arises: what if the grounder of the meaning of ‘Aristotle’ thought of only part of Aristotle, not the whole object? What would that make of our use of the word ‘Aristotle’? It would be absurd to hold that though we associate the property of being a whole human with the name, the name in our mouths only refers to part of him. We would have been in error about the meaning of ‘Aristotle’ all the time! So we must conclude that ordinary ‘linguistic consumers’ must associate some descriptive knowledge with the name too. But if we admit this then we admit that the aspect of use that matters to meaning isn’t so much the grounding incident as speakers’ contemporary use to indicate the contents of their mental states. And now we begin to see a reversal of the explanatory roles of grounding use (i.e. the dubbing) and contemporary use, according to the meaning is use platitude and the communication platitude: it is the information we use now which helps us determine what the relevant causal link is, and what the relevant dubbing process should be. This is not to say that the causal links are arbitrary, only that the causal links cannot be wholly separated from our contemporary use of the words.

Now the issue begins to converge with the discussion of primitivism. A crude but pure version of primitivism says that meaning-facts are primitive such that we may all always be in error about which sentences are related to which meanings. This shows primitivism to be inadequate because the concept of meaning is not such that global error is possible. Pure causal theories invite the same problematic: what we do now is separated from what happened when meaning was determined, so we may all always

---

be wrong about meaning. This is what the *qua*-problem brings out: we don’t know the
meaning of ‘That is Fido’ until we know which property the grounder intended, and
the causal story by itself hasn’t got the resources to give us this information. So we
cannot exclude that we are all always in error about meaning (as my example with
‘Aristotle’ brought out). I argued that whenever primitivism is confronted with this
problem the only available move seemed to re-invoke aspects of use (dispositionalism
in some form or other). And with pure causal theories the available move is similar:
there is, after all, something in our use which plays a significant part in connecting us
cognitively with the meaning-constituting facts, and which thus is supposed to enable
us to avoid global error. But, as with primitivism, this move just re-introduces a
strongly networked meaning is use platitude which pushes the primitivist, or causal,
intuitions into the background.

e) Further remarks.

Pure causal theories of meaning seem inadequate on different grounds than
primitivism: they allow no clear notion of how meaning helps us tell other people how
we take things to be; they leave out the idea that speakers have privileged knowledge
of what their words mean when they use them; and they make a mockery of the notion
that sometimes meaning explains use. These differences should not make us overlook
the common element of primitivism and subpersonal theories: even though
subpersonal causal theories are very different from primitivism in outlook, they offend
against a core aspect of the concept of meaning, viz. that occasional, but not global,
error must be possible. Moreover, they do this in much the same way: by letting the
meaning-constituting facts come too far apart from the facts about our use of words.
Once again we get the result that if the incompatibilist leaves out the meaning is use
platitude, or only pays lip-service to it, then the other platitudes are flouted.
6. Conclusion.

In this chapter I have argued against the incompatibilist who reacts to the tensions among the platitudes about meaning by rejecting the meaning is use platitude. I have considered types of incompatibilism which have actually been canvassed: primitivism and pure causal theories. For each version of these types it was shown that either it entails, or is at least consistent with, re-introducing the meaning is use platitude or it flouts the functional role of the concept of meaning as given by the remaining platitudes.

Apart from discrediting the incompatibilist approach, this also demonstrates the networked character of the platitudes: once the meaning is use platitude is taken out it becomes hard to see how the other platitudes could retain their functional roles. But the exercise also suggests which version of the meaning is use platitude is needed for the other platitudes to do their work: viz. a dispositionalist meaning is use platitude. It hasn't been shown that adopting such a theory will avoid the three tensions I presented for the platitude-based analysis of meaning, but the suggestion resulting from the discussion of incompatibilism is strong enough to be tested as a hypothesis: if compatibilism is the right strategy, then some platitudes must be revised, and probably the meaning is use platitude must be revised in dispositionalist terms.
CHAPTER IV

TOWARDS COMPATIBILISM:
USE-DEPENDENCE

Was ich immer tue, scheint zu sein - zwischen Sinnbestimmung und Sinnverwendung einen Unterschied hervorzuheben.
Wittgenstein, Bemerkungen über die Grundlagen der Mathematik, III.37.

1. Introduction.

In chapter 2 I gave a conceptual analysis of the concept of meaning by motivating the five platitudes that surround the concept. I then detailed three tensions among these platitudes. In chapter 3 I demonstrated that one cannot resolve the tensions adequately by rejecting the meaning is use platitude. In this chapter I shall begin to defend the position that the platitudes are compatible and that no platitude should be rejected. My claim is that a particular conception of the property that could underwrite the platitude-based analysis will be necessary and sufficient for showing that the tensions do not arise. There is only a small price to be paid for this: a few of the platitudes will be revised accordingly, but these revisions will allow the concept of meaning to fulfil the overall functional role we expect of it. To this end, I shall attempt to show that the concept of meaning, that is, the concept of the semantic relation expressed by \( \phi \) means \( p \), could be the concept of a response-dependent property, in particular, the concept of a dispositional use-dependent property. I shall then, in the subsequent three chapters, demonstrate how the tensions are resolved, given use-dependence.
In this chapter I will introduce my notion of use-dependence. This notion is a version of the more general notion of response-dependence, which I will discuss in some detail. Here I assess what the purpose of a response-dependent account of a concept could be, and accordingly set out constraints that fit such a purpose. I use Crispin Wright 1993 as the basis for this, but suggest an amended version of one of his four constraints. Some versions of response-dependence are closely linked with dispositionalism, i.e. the idea that response-dependent concepts are response-dispositional. I endorse this reading of response-dependence and discuss the notion of dispositionalism in some detail. Finally, I return to my notion of use-dependence and re-state it according to the discussion of response-dependence.

2. Introducing the use-dependence of meaning.

To say that the concept of the semantic relation expressed by ‘\( \varphi \) means \( p \)’ is the concept of a use-dependent property is, initially, just to say the following is *a priori*:

\[(1) \ \varphi \text{ means } p \text{ in S's language } \equiv \varphi \text{ as used by S has the property U.}\]

Where ‘\( \varphi \)’ is a declarative sentence type, ‘\( p \)’ is a proposition, S is a speaker, and U a property that must be unpacked such that \( p \) indeed comes out as the meaning of \( \varphi \). The use-dependence of meaning must exclude the possibility that \( \varphi \) means \( p \) irrespective of S’s use of \( \varphi \). One way to do this is to demonstrate that use-independence is not an option (which I did in the previous chapter); another way is to show what it is about use that makes something like (1) true.

The background for suggesting something like (1) is the following. In the discussion of incompatibilism it became clear that meaning-facts are not primitive, or purely causal. This provides support for the meaning is use platitude and motivates a
compatibilist strategy. But it also suggests where revision of the platitudes must be undertaken in order to ensure compatibility among them. If primitivism isn’t feasible, then the platitude about explanatory meaning (which says that meanings sometimes can explain uses) cannot be understood in a primitivist manner. That is, it is not the case that primitive meaning-facts can explain uses, and the explanatory meaning platitude must be understood to exclude this primitivist reading. Instead, the explanatory meaning-facts are underpinned by a “meaning is use” doctrine which can explain uses. This suggests the biconditional account: meanings explain uses, but are themselves to be explained by use. This tells us something about the general form of the account, but it also brings into the forefront the tension I called the “meaning is use” problem.

Since we know something about the meaning is use platitude, it is in fact possible to begin the unpacking of U. In my discussion of the meaning is use platitude it transpired that, if meaning is use, then the notion of use should be understood as ‘use to indicate mental states with certain contents’. Hence, it is a priori that:

(2) $\varphi$ means $p$ in S’s language $\equiv$ S uses $\varphi$ to indicate all and only mental states with the content $p$.

However, until we have clarified the notion of ‘use’ in (2), we haven’t done anything to substantiate the use-dependent character of meaning. To achieve this I shall eventually opt for this dispositional version:

(3) $\varphi$ means $p$ in S’s language $\equiv$ S is disposed to use $\varphi$ to indicate all and only mental states with the content $p$.

For example, it is a priori that:
(3.1) 'there is an echidna' means *there is an echidna* in S's language ≡ S is disposed to use 'there is an echidna' to indicate all and only mental states with the content *there is an echidna.*

My hope is that by defining 'use' as 'disposed to use', I will be able to substantiate the claim that (3) exhibits the particular use-dependent character of meaning in such a way that all the five platitudes come out as true and compatible. So, at this stage, the *a priori* analysis of the semantic relation expressed by "φ means p" comes out like this:

\[
\text{UD v.1:} \\
φ \text{ means } p \text{ in S's language } ≡ \text{ S is disposed to use } φ \text{ to indicate all and only mental states with the content } p.
\]

In what follows I shall provide the background and motivation for this theory.

3. Use-dependence as a response-dependent notion.

Use-dependence (UD) takes the form of an *a priori* biconditional because, as I hope to show, the concept of meaning belongs with the larger family of so-called response-dependent concepts, which are all characterised by commitment to such *a priori* biconditionals. By aligning UD with this broader notion, I shall bring out the sense in which meaning is dependent on use.

a) What is response-dependence?

The notion of response-dependence is helpful in capturing the *a priori* difference between concepts we feel have different degrees of mind-dependence, *i.e.* which have different degrees of involvement of human response in the determination of their extensions. For example, response-dependence can begin to capture the traditional difference between primary qualities, such as the shape of things, and secondary
qualities, such as the colour or smell of things. A response-dependent theory about colour would say that truths about colours are dependent on the responses and/or opinions of subjects when they perceive and/or judge about colours, when confronted by coloured things. The contrast would be with a more intuitively robust, mind-independent primary quality such as shape. Here we would say that truths about shape are not in the least dependent on the responses and opinions of suitable subjects.

The basic idea is that, for the less robust, mind-dependent concepts, the responses and/or opinions of suitable subjects somehow enjoy a privileged status to such an extent that subjects may, in the right circumstances, be infallible. For example, a response-dependent account of the concept of the sweet-smelling would posit a close constitutive relation between what it is to think, judge, sense etc. that something is sweet-smelling, and what it is for that something to be sweet-smelling. Thus, in general:

\[(1) \text{ if a suitable subject responds to } x \text{ by thinking, judging, having the sensation etc. that it is } F, \text{ then } x \text{ is } F.\]

But (1) isn’t interesting as it stands because it doesn’t exclude the possibility that something can be \( F \) without it being the case that anyone would judge it so. To begin to capture the intuitions about the less robust discourse we need a conditional that goes the other way:

\[(2) \text{ if } x \text{ is } F, \text{ then a suitable subject responds to } x \text{ by thinking, judging, having the sensation } etc. \text{ that } x \text{ is } F.\]

The two conditionals can be combined in one \textit{a priori} biconditional:

\[(3) x \text{ is } F \equiv \text{ a suitable subject responds to } x \text{ by thinking, judging, having the sensation } etc. \text{ that } x \text{ is } F.\]
Biconditionals of (very) roughly this type form the common point of departure for the various kinds of response-dependent accounts. The refinements and constraints are many, but the basic intuition is that if it is \textit{a priori} that something like (3) is true of the characteristic concepts of a discourse, then it is a serious possibility that that discourse should be classed among the less robust, mind-dependent discourses. If, on the other hand, it is easy for us to conceive of something’s being \( F \) independently of the responses or opinions of subjects, then (3) presumably isn’t true of the discourse in question. The biconditionals are then, in general, intended to mark off the response-independent discourses from the response-dependent ones. My notion of UD is intended to fit this general response-dependent mould.

Much of the background and detail of different notions of response-dependence have been worked out by Johnston, Pettit and Wright.\textsuperscript{1} I think it is fair to say that they all share the brief commitments I have set out above, though that doesn’t convey much about the accounts they each defend. Before I go on to define the version of response-dependence which fits UD, it will be useful to settle on a fairly neutral biconditional template: \( F \) is a response-dependent concept if and only if it is \textit{a priori} that:

\[
(4) \; x \text{ is } F \equiv S \ [\text{a suitable subject} ] \ Rs \ F\text{-ly } \ [\text{is disposed, judges, reacts etc. in a certain } F\text{-like way} ] \ to \ x \ in \ C \ [\text{the conditions under which the response takes place}].
\]

in short, it is \textit{a priori} that:

\[
(T) \; x \text{ is } F \equiv S \ Rs \ F\text{-ly to } x \ in \ C.
\]

Where the subjects, responses and conditions all await substantial specifications. To a large extent it is these specifications which determine the status of the final response-dependent notion. Though one can formulate response-dependence for different things I shall discuss it for concepts, and, in a broader sense, for discourses identified by their characteristic concepts. Thus the concepts of a discourse D may be concepts of response-dependent properties, if something resembling the template (T) holds a priori for them. Now for some more detail.

\section*{b) Global or local response-dependence?}

One can hold that response-dependence is a global phenomenon, holding for all discourses. Or one can hold it locally, for particular discourses. These two strategies must be backed up by quite different considerations. A local response-dependence theorist should be prepared to admit that some discourses could be response-independent, or else he or she will not be able to draw much of a distinction. Hence, the arguments that drive a localist must be fairly topic-specific, they must concern the particular content of the discourse in question.\footnote{Johnston has canvassed such local response-dependence theses for discourse about morals and about colour; Johnston 1989 and 1992.} In contrast, the globalist must provide arguments to the effect that no discourse could be response-independent, or alternatively, that all possible discourses have something in common that makes them inherently response-dependent. Arguably, Hilary Putnam’s arguments against the doctrine he calls metaphysical realism yield a global argument of the first type; Pettit’s solution to Kripke’s rule-following problematic produces a global thesis of the second type.\footnote{Putnam 1981, passim. Pettit 1990.}

Since my arguments are focussed on what is particular about discourse about meaning, I shall only pretend to defend a local thesis of response-dependence. Though I am sympathetic to Pettit’s claim that a complete (non-primitivist) refutation of...
Kripke-Wittgenstein content-scepticism entails global response-dependence, I can afford to dodge that question here for the following reason. I am only giving an analysis of the concept of meaning, i.e. of the concept of the semantic relation expressed by \( \varphi \text{ means } p \), where \( \varphi \) is a declarative sentence type and \( p \) is a proposition. Hence I am not offering an analysis of the right-hand relatum, i.e. of what gives \( p \) its particular content. Some property other than use-dependence may underwrite the concept of propositional content. In contrast, Kripke-Wittgenstein scepticism works across the board, for linguistic meaning as well as mental content, and therefore a sustained defence will have to account for both. This defence, whatever it is, will quickly globalise because across-the-board meaning/content-scepticism concerns everything we can ever say or believe. I am only providing a response-dependent defence against scepticism about semantic relations for linguistic meanings, and thus my thesis only concerns what we can possibly say, not what we can possibly believe. Therefore my account is prima facie consistent with a response-independent account of mental content. It requires an extra, very controversial, premise to reach global response-dependence from my thesis, viz. the premise that facts about the semantic relations of linguistic meaning are constitutive of facts about mental content. Such a premise is controversial because it takes a stance in the debate about the relative priority of language and thought.

c) **Successful response-dependence: four constraints.**

Crispin Wright puts versions of the biconditionals associated with response-dependence to use in clarifying a distinction between discourses whose characteristic best judgements are extension-determining and discourses whose best judgements are extension-reflecting.\(^4\) The idea is to clarify, as I discussed above, the distinction between response-dependent concepts and response-independent concepts, or, as Wright terms it, concepts which are judgement-dependent and concepts which are

\(^4\)Wright 1992, Appendix to ch. 3.
judgement-independent. There is much detail in Wright’s discussion, corresponding to his particular concerns, but I wish here to focus on what comes to the forefront in his discussion, viz. the notion of order of determination. This is the question of where, as it were, our explanation of a concept should begin: do our responses or judgements determine truths about the extension of the concept in question, or does the extension of the concept determine our responses or judgements such that they merely reflect the extension of the concept? We might agree that (T) is true \textit{a priori} for some concept without knowing which direction the order of determination has across the biconditional (where a left to right reading is extension-reflecting, and a right to left reading is extension-determining). Wright suggests four constraints which constitute a directional test to answer this question. In what follows I shall discuss the significance and purpose of the idea of order of determination, with a particular eye to the idea of use-dependence, and then formulate Wright’s four constraints accordingly.

\textit{i) the significance and purpose of the notion of order of determination.}

I see the main purpose of pressing a response-dependent account of a concept as two-fold: i) it rules out that the concept in question could be response-independent. That is, on the basis of conceptual analysis of the characteristic content of the concept, it should be \textit{a priori} that it is not the concept of a response-independent property; ii) it should show wherein consists the nature of the human involvement in the determination of the extension of the concept in question. That is, it should not only show that the concept isn’t response-independent, but also in what way this particular concept could be different from other response-dependent concepts which may have other styles and degrees of human involvement.\footnote{There is a third and further question, which I am not so interested in here: viz. the question of where response-dependence fits in the debate about realism and anti-realism. This is a further question because one can agree that a given concept is response-dependent but disagree about its status as realist or anti-realist. This debate occurs in the writings of Wright 1992 and Pettit 1991 but I shall not enter it here.}
The notion of an order of determination feeds into these two purposes: if it is \emph{a priori} that the responses associated with a concept are extension-determining, then they are not extension-reflecting, and hence the concept is not response-independent. And the \emph{manner} in which it is shown \emph{a priori} that the responses are extension-determining should reveal wherein consists the nature of the human involvement in this determination. Of course, it is also possible to argue, by \emph{reductio}, that response-independence isn't an option for a particular concept, much as I did for the concept of meaning in the previous chapter, but this doesn't directly address the concern about human involvement, and differences within the class of response-dependent concepts.

However, biconditionals, as the one in (T), have no 'direction', so the idea of a directional reading does not apply readily. In order to find a place for the idea of a directional reading of response-dependent theses, the biconditionals must be associated with accounts of the concept in question that \emph{do} have 'directions'. Here are two such possible \emph{a priori} accounts:

(1) \(F_X = \text{def } S \; R_s \; F-ly \; to \; x \; in \; C\).

and:

(2) \(S \; R_s \; F-ly \; to \; x \; in \; C = \text{def } F_X\).

Since (1) and (2) are definitions they are directional. Now reflect that the biconditional template:

\[(T) x \; is \; F \; \equiv \; S \; R_s \; F-ly \; to \; x \; in \; C,\]

is entailed both by (1) and by (2). This is because (1) says that it is \emph{a priori} that whenever \(F_X\) is the case it is the case that \(S \; R_s \; F-ly \; to \; x \; in \; C\), and because (2) says
that it is \textit{a priori} that whenever it is the case that S Rs F-ly to x in C it is the case that Fx. So though we can see how ‘directional’ readings may play into the understanding of (T), (T) on its own appears to be neutral between these readings.

However, this suggests how the response-dependent theorist\textsuperscript{6} can after all earn the right to talk about an order of determination. Firstly, it must be perspicuous in the account that (2) is false, \textit{i.e.} it should be shown that the analysis of the concept is inconsistent with assuming something like (2). But secondly, it must also be perspicuous in the account that the \textit{a priori} biconditional in (T) isn’t rendered true by tacitly assuming the response-independent account in (2). It is one thing to show that (T) isn’t entailed by (2) because (2) isn’t available, another to demonstrate that one’s proof of (T) doesn’t in fact rely on (2). Of course, these two demands on perspicuousness dovetail with what I identified as the main purposes of a response-dependent account of a concept. If both demands are satisfied we should be satisfied that the concept is response-dependent.

Let me re-state UD such that this target is clear. The \textit{a priori} analysis of the semantic relation expressed by ‘\( \phi \) means \( p \)’ implies \textit{a priori} definitional biconditionals of this form:

\[ UD \, v.2: \]
\[ \phi \text{ means } p \text{ in } S’s \text{ language } \equiv_{\text{def}} S \text{ is disposed to use } \phi \text{ to indicate all and only mental states with the content } p. \]

With this version of UD I have tried to be perspicuous about the role of response-dependence: the ‘def’ is there to remind us of the directional reading (though I haven’t

\textsuperscript{6}By ‘response-dependent theorist’ I don’t mean that the theorist depends on responses! I use it instead of the correct, but more awkward ‘response-dependence theorist’.
proven positively that it holds), the ‘≡’ is there to remind us of the a priori biconditional normally associated with response-dependence.

ii) four constraints on response-dependence.

Wright proposes four constraints which must be satisfied in order to show that a concept is response-dependent. I suggest that a version of those four constraints is what must be satisfied in order to achieve the two goals of perspicuousness I identified for UD.

Wright’s four constraints are as follows.\(^7\) In order to establish judgement-dependence, which is Wright’s version of response-dependence, for a concept, say \([F]\), one must endorse a version of (T) which satisfies (I)-(IV) below:

(I) The apriority constraint. (T) must hold a priori. That is to say, (T) must hold as an a priori function of the characteristic content of \([F]\). If (T) is only contingent, then there could be hitherto unobserved cases that do not fit (T). For example, the constraint is not satisfied if the biconditional ‘x is square ≡ S judges that x is square in C’ is true as a matter of contingent fact, i.e. if it just so happens that up until now all the square things are judged to be square. The reason is that it would be possible for there to be square things which we would not so judge about.\(^8\)

(II) The substantiality constraint. The C-conditions mentioned in (T) must be characterised substantially, not in a trivial whatever-it-takes-to-get-it-right fashion. Whatever-it-takes conditions can be formulated for all concepts, including response-independent concepts. Thus allowing whatever-it-takes conditions will obliterate the


\(^8\) Miscevitz 1998 argues that the apriority constraint is unlikely to be satisfied for manifest properties, because, he argues, we need to go to science or philosophy to demonstrate response-dependence for such properties. I agree that this is often the case, e.g. for the case of colour, but I disagree that the apriority constraint can never be satisfied. For some concepts, e.g. the concept of meaning, there is a priori reason to believe response-dependence holds.
distinction which response-dependence was intended to draw. For example, the following biconditional is true \textit{a priori} for squareness: ‘\(x\) is square \iff S judges that \(x\) is square in whatever conditions would be optimal for judging about squareness’. But that biconditional does not show that the concept of squareness is response-dependent because it is consistent with something like (2) above. Notice that the substantiality constraint is there to ensure that the distinction between response-dependence and -independence can be drawn, not because there is a separate requirement that the C-conditions be stated in a particularly informative way. This implies that the C-conditions just have to be stated in a way that is substantial \textit{enough} to draw the desired distinction, irrespective of how intrinsically interesting that statement is.

(III) The \textit{independence} constraint. Whether the C-conditions are satisfied in a particular case must be specifiable in terms that are logically independent of any truths about the extension of the concept in question. The motivation for this constraint is the following. In setting out the C-conditions for, \textit{e.g.}, redness, we might have to avail ourselves of such conditions as \(x\)’s ‘being stable in colour’; that is, we might have to presuppose that there are determinate facts about the extension of the colour predicate in question. This is risky because then the response-dependent theorist becomes vulnerable to the criticism that those occurrences of the concept in question implicitly rely on a response-independent conception of the concept. Satisfying the independence constraint ensures that there will be no such potentially response-independent occurrences of \(F\) in the specification of the C-conditions. I shall return to this constraint below.

(IV) The \textit{extremal} constraint. Even though there may be a response-dependent account of why (T) is true \textit{a priori}, there must be no better response-independent account available of why (T) is true. (T) may, as we have seen, be true even though the concept in question is response-independent, \textit{viz.} if \(S\) happens to be an infallible \textit{detector} of F-ness in \(C\). In that case the order of determination would go from
independently determined facts about the extension of F to the explanation of S's responses, given a story about the special epistemology of the case. Hence, the concept F would not be response-dependent after all. For example, assume the concept of pain is response-independent, i.e. we can conceive of facts about pain independently of a capacity for judging about pain. It might nevertheless be plausible that we are infallible detectors of pain such that, for subjects who can judge about pain, they are in pain if and only if they judge they are in pain. The constraint is there to ensure that the concept of pain would not come out as response-dependent even though the biconditional is in fact true. This is because, for the concept of pain, we can, in this case, conceive of the facts and their occurrence such that subjects come out as infallible about pain. Notice that in arguing, in the previous chapter, against the incompatibilist strategies of primitivism and causal theories, I have gone some way to provide an argument that no such better account of the biconditional for the concept of meaning can be given.

iii) weakening the independence constraint: uniformity.

I agree with Wright that constraints (I), (II), and (IV) are necessary, but not sufficient, to ensure a response-dependent order of determination, as I expounded that idea above. In order to gain sufficiency something like (III), the independence constraint, is also needed, but it is too strong as it stands. Johnston has argued forcefully against the constraint because he thinks it is linked with an ambition to provide reductive analyses. To attempt to provide a ‘directional’ reading of response-dependence which satisfies the independence constraint:

would be to have a distinction on one side of which fell only those concepts which allowed for a reductive definition or analysis in terms of concepts of our responses [...] We have good reason to believe that such reductive accounts have very dim prospects.9

---

Johnston's thought is that we can ditch the independence constraint without completely ditching the explicatory value of the biconditionals. With the remaining three constraints satisfied, and free use of the contested concept on both sides of the biconditional, we can see them as offering notions of response-interdependence. As Johnston notes, this would imply that talk of an order of determination would be entirely out of place. There would be no correct answer to the question of which comes first in the conceptual order of determination: the occurrences of [F] in the C-conditions on the LHS, or the occurrences on the RHS. Johnston claims that the point of offering response-interdependence as an account of a concept would be to highlight that facts about Fs cannot outstrip our idealised capacities for response.10

But Johnston's complaint seems to me to miss at least part of the point of the independence constraint. The constraint might be so strong that it requires reductive analysis, but the motivation for it has nothing to do with reduction. The real problem is that without something in the same vein as the constraint, we will have no guarantee that the occurrences of [F] in the specification of the C-conditions are really response-dependent. This was the point of the example with redness: if 'stability of colour' occurs in the C-conditions, then it will be an open question whether those occurrences implicitly rely on a response-independent conception of the concept. Something extra must be said to provide this guarantee. All Johnston has pointed out is that Wright's response clashes with the idea that reductive analysis is not to be had, but this should not imply that no other constraint could be construed. Johnston introduces the notion of response-interdependence as the best we can hope for, and concludes that it at least shows that facts about Fs cannot outstrip our idealised capacities for response. But we can now see that this is a non sequitur: without a suitable extra constraint we have no

10Notice that the debate about order of determination isn't about the notion of a temporal order, as Wright also notes, cf. 1992, p. 134. It has to do with the conceptual order of determination, i.e. with supervenience or constitution as it appears in philosophical explanations.
guarantee that the concept F is uniformly response-dependent and hence no guarantee that facts about Fs cannot outstrip our idealised capacities for response.

In Wright’s discussion, he suggests that the response-dependent theorist must be able to respond to this imagined challenge:

“Show that the way that you have implicated those concepts is consistent with your overall thesis, that their extension is, at least partially, constrained by [our responses in C].”

The answer Wright provides is the independence constraint, which he does admit is a ‘conservative response’ to the difficulties. So the question arises: is there a less conservative constraint which can nevertheless provide a guarantee that no occurrence of the concept F is response-independent? That is, is there a constraint that ensures that wherever the concept in question occurs it will uniformly be response-dependent?

The Ramsey-Carnap-Lewis style analysis can provide just this sort of constraint. The point of this style of analysis is that it allows us to derive explicit definitions from implicit interdefinitions in the following way:

Let $M$ be the conjunction of all the a priori platitudes for the concept $M$, where the occurrences of $M$ are in property-name style. Then replace all occurrences of the property-name, ‘$M_i$’, by a distinct variable such that we get:

\[(1) \ M(x_1,...,x_n),\]

then ramsify such that:

---

12 Wright 1992, ibid.
Now it is possible to give the conditions under which $\alpha$ has $M_i$ as:

\[ (3) \; \alpha \text{ has } M_i \equiv (\exists x_1)(\exists x_2)\ldots [\alpha \text{ has } x_i \& \ldots M(x_1, x_2, \ldots)], \text{ where } x_i \text{ replaces } M_i. \]

The immediate point is that even though the *analysandum* occurs in the implicitly defining platitudes, those occurrences may be stripped away so that suspicions of circularity are alleviated. For my purposes the point is the following: by ramsifying, we ensure that all the potentially problematic occurrences of the *analysandum* occur within the scope of the existential quantifier. Therefore, if the whole analysis is *a priori*, we can be sure that all the occurrences are to be understood uniformly.

This gives us a weaker version of the independence constraint, *viz.* the uniformity constraint:

\[(\text{III}^*) \text{ The uniformity constraint. The C-conditions must be specifiable in terms that are logically independent of any truths about the extension of the concept in question, or it must be such that it is } a \text{ priori } \text{ that any occurrence of the concept can be within the scope of an existential quantifier that governs the entire definiens.}\]

Here is what the uniformity constraint achieves: if it is satisfied, then no occurrence of the concept in question, say the concept of redness, can be response-independent if the analysis as a whole shows the concept to be response-dependent. This constraint, together with the three other constraints, are necessary and jointly sufficient conditions for response-dependence.
The uniformity constraint is weaker than the independence constraint because it doesn’t require the *definiens* to be *logically* independent of truths about the extension of the concept of redness. We could not set out the platitudes in \( M \) without presupposing grasp of such truths, but once \( M \) is fixed we can overcome the dependence on truths about the extension of concept of redness by ramsifying.

To sum up, with the uniformity constraint replacing the original independence constraint we should have enough to re-gain the right to talk about an order of determination such that, though facts about the extension of the concept and facts about the C-conditions are not logically independent, we can be sure that the concept is indeed the concept of a response-dependent property. That is, even though the biconditionals don’t themselves have directional readings, we can be sure that they can be derived on the basis of a response-dependent, and therefore directional, definition of the concept of redness, rather than a response-independent definition.

It should be clear, though, that since the requirement on a Ramsey-style definition isn’t *logical* independence, there is no univocal method of assessing the success of such a definition as informative and interesting. This will instead be a matter of how well the platitude-based analysis comes out. The better the platitude-based analysis, the more convincing the claim that a concept is response-dependent or not. I shall therefore not spell out how UD satisfies these constraints until I have done more to expound UD as a response-dependent thesis (I return to the constraints in Ch. 6).

d) *Dispositions and response-dependence.*

In the first version of the doctrine of use-dependence, I suggested that UD analyses semantic relations between sentences and meanings in terms of *dispositions*:
CHAPTER IV, TOWARDS COMBATIBILISM: USE-DEPENDENCE

UD v.2:
φ means p in S’s language \( \equiv_{\text{def}} \) S is disposed to use φ to indicate all and only mental states with the content p.

There are three initial reasons for this. Firstly, it is fairly common to explicate the “meaning is use” slogan in terms of dispositional use.\(^{14}\) The reason for this is that the concept of meaning cannot be analysed in terms of actual use for actual use is too unconstrained to deliver robust meanings (I can actually use ‘there is a wombat’ in all sorts of ways without that determining any meanings for the sentence). In contrast, dispositions are counterfactually constrained: if use is dispositional, there must be truths about how I would use a sentence irrespective of how I actually use it on a particular occasion. Secondly, the notion of dispositions has a natural place for the notion of normal and standard conditions—C-conditions—which fits the response-dependent template (T) well. Thirdly, I think that the three tensions among the platitudes can only be resolved with a dispositional reading of UD. I believe that the ordinary notion of dispositions is far more resourceful than it may seem in the light of the forceful criticism it has been subjected to by people who are suspicious about dispositionalism in the theory of meaning.\(^{15}\)

\(^{i)}\) counterexamples to the counterfactual analysis of dispositions.
Dispositionalism is a controversial doctrine, not only in the debates about meaning and response-dependence, but also because the standard counterfactual analysis of dispositions is subject to serious counterexamples. The standard analysis goes like this:

\[ (1) \text{Something } x \text{ is disposed to } R \text{ in } C \equiv \text{if } x \text{ were in } C, \text{ then } x \text{ would } R. \]

\(^{14}\) Notably, the critical discussion of dispositionalism in Kripke 1982, Ch. 1; also Boghossian 1989; McGinn 1984. For positive attitudes towards dispositionalism see Forbes 1984; Ginet 1992.

\(^{15}\) E.g. Kripke 1982 and Boghossian 1989.
For example, for the dispositional property of fragility:

\[(2) \, x \, \text{is fragile in} \, C \equiv \text{if} \, x \, \text{were struck in} \, C, \text{then} \, x \, \text{would break.}\]

But now we get the following counterexamples. Sometimes we want to say that \(x\) is fragile even though the counterfactual is false, and sometimes the counterfactual may be true even though we would say that \(x\) is not fragile. Here are three types of counterexample, taken from Johnston:\(^{16}\)

A cup is fragile, but if it were struck an angel would make sure it hardens and doesn’t break. Here the actual obtaining of the \(C\)-conditions change the way the cup is disposed, even though, as long as the \(C\)-conditions do not obtain, the cup is disposed to break when struck. This is a case of \textit{altering}: the very dispositions of the cup are altered when the \(C\)-conditions obtain.

A piece of metal is not fragile, but an angel makes sure that if it were struck it would break. Here the counterfactual for fragility is true, but the piece of metal is not fragile. This is a case of \textit{mimicking}: \(x\) has not got the disposition, and neither does the coming about of the \(C\)-conditions give it the disposition, but something else about the conditions make sure the counterfactual is true and thus makes it seem as if the piece of metal is fragile.\(^{17}\)

A cup is fragile, but an angel makes sure that if it were struck then it wouldn’t break. The counterfactual is false even though the cup is fragile. This is a case of \textit{masking}: the cup is fragile, and the coming about of the \(C\)-conditions doesn’t alter its fragility, but it just so happens that something else in the \(C\)-conditions prevents the characteristic manifestation of the disposition.

---

\(^{16}\)The counterexamples were initially put forward by C.B. Martin (as published later in his 1994); cf. also Shope 1978. I follow the discussion in Johnston 1993, appendix 2.

\(^{17}\)I think there is a variant of this case, call it ‘true mimicking’, such that the cup has the disposition to break when struck, and the counterfactual is true of the cup in virtue of an angel’s interference, such that the manifestation of the disposition in \(C\) isn’t \textit{explained} by the possession of the disposition.
In each of these three cases the standard counterfactual analysis of dispositions fail.\textsuperscript{18} However, this only shows that the standard analysis is inadequate. It doesn’t show that there is no analysis to be had. Lewis has suggested an amended counterfactual analysis which sustains the notion of dispositions while allowing the possibility of the scenarios described in the purported counterexamples. The central step in arriving at the amended counterfactual analysis of dispositions is to focus on the intrinsic properties of the object in question. That is, when we ascribe dispositions we ascribe some intrinsic property to the object in question. And we do that, in effect, on the principle that as long as this property is in place the object will have the disposition, even though, in fact, the disposition will be altered, or its manifestation mimicked or masked in C-conditions. This is the core of Lewis’ suggestion:

\begin{align*}
\text{Something is disposed at time } t \text{ to give response } r \text{ to stimulus } s \text{ iff, for some intrinsic property } B \text{ that } x \text{ has at } t, \text{ for some time } t' \text{ after } t, \text{ if } x \text{ were to undergo stimulus } s \text{ at time } t \text{ and retain property } B \text{ until } t', s \text{ and } x's \text{ having of } B \text{ would jointly be a [cause] of } x's \text{ giving response } r. \textsuperscript{19}
\end{align*}

This certainly gives us something to work with, and it tallies quite well with the ordinary idea that ascriptions of dispositions in general have \textit{ceteris paribus} clauses.

The point, for my present purposes, is not that I think Lewis’ definition is the final word. Rather, it is that, given something like Lewis’ suggestion it would premature to say that the notion of dispositions has \textit{no} analysis, or that we shouldn’t talk about dispositions at all. This much is sure: cups are fragile, sugar-cubes are water-soluble, people are choleric, and so on. It cannot be up for discussion that there \textit{are} dispositional properties (unless we are radical sceptics). In particular, that there are

\textsuperscript{18}The status of the counterexamples is not uncontroversial, \textit{cf. e.g.} Blackburn (1994, §2). I return to this issue in Ch. 6§3.d.

\textsuperscript{19}Lewis 1997, p. 157.
dispositions at all cannot be up for discussion in the debate about the dispositionalist reading of the "meaning is use" slogan. What can be up for discussion is this: whether a dispositionalist reading of the "meaning is use" slogan can be employed in an analysis of the concept of meaning. We must require, therefore, that arguments against a dispositionalist reading of the "meaning is use" slogan be specific to the debate about the concept of meaning. Otherwise those arguments will rob us of the right to talk about fragility, water-solubility and cholerics.

I conclude that, even though there are counterexamples to the simple counterfactual analysis of dispositions, there is nothing wrong, prima facie, with a dispositional reading of the "meaning is use" slogan. This reading must of course make sense of the possibility of the counterexamples, but it need not endorse any particular analysis. My claim, as we shall see in Ch. 6, is that there are no arguments against a dispositional rendering of UD which are not also arguments against the very notion of dispositions.

**ii) Wright on dispositionalism and provisional equations.**

Because Wright is worried by the counterexamples to the counterfactual analysis, he promotes a non-dispositional version of response-dependence. Instead of working with a priori biconditionals, he works with provisional equations such as the a priori:

\[(PE) (x)(S) (CxS) \rightarrow (x \text{ is F iff } S \text{ judges } x \text{ is F}),^{20}\]

which says, for example, that if S is confronted with x in C, then (x would be red iff S would judge it to be red). Provisional equations circumvent any disquiet one may have about dispositions, but they bar us from making judgements about how things are in non-C-conditions.\(^{21}\) We lose, for example, the right to say that a chameleon is green when the lights are switched off. I do not wish to deny that some concepts could

---

\(^{20}\)Wright 1992, p. 119.

be accounted for in terms of (PE), but such accounts should be suggested on the basis of the analysis of the concept in question, not on the basis of general disquiet about the notion of dispositions. We know that there are dispositions and, if an a priori analysis of a concept shows that it is the concept of a dispositional property, then we should go ahead and endorse a dispositionalist analysis of the concept.

Wright's worry about the counterexamples is that they show that purely contingent facts about the obtaining of the C-conditions, and the categorical bases of the dispositions, influence the truth of the supposedly a priori counterfactuals in question. Therefore, they cannot be shown to hold on a priori grounds, as they should to count as establishing response-dependence. This seems to me to be wrong. What the counterfactuals show is that it is part of the analysis of dispositions, however it may eventually come out, that it is a priori that contingent facts may interfere in various ways with the manifestation of the dispositions. But incorporating this aspect into the a priori analysis doesn't require us to have a priori opinions about particular contingent matters.22

e) What type of dispositional response?

Having settled on a dispositional reading of the response-dependent biconditionals, we must now settle on the type of response suitable for UD. Which responses are semantic relations dependent on? The simple answer is that semantic relations are dependent on a subject's dispositions to use declarative sentence types \( \varphi \). For example, the semantic relation in S's language between the sentence type 'wombats are cute' and the meaning wombats are cute is dependent on S's dispositional use of the sentence. This is motivated by the meaning is use platitude which states that 'use'

---

22 Wright does have specific concerns about dispositionalism in the theory of meaning. There are concerns both about the apparent behaviourism and concomitant holism associated with dispositionalism, and concerns about how the dispositionalist can make the phenomenology of first-person authority come out true (cf. Wright 1989, p. 236, §V, and 1993a, p. 23). I discuss the first concern in Ch. 6, §2, and the second concern in Ch. 7.
is ‘use to indicate mental states with the content $F$’. What it means is just this: if $S$ is disposed to utter tokens of $\varphi$ when and only when $S$ has mental states with the content $p$, then $\varphi$ means $p$ in $S$’s language.

This of course sounds terribly behaviouristic, as if we find ourselves mouthing off sentences when we find ourselves with certain mental states. I do not think this is the case, and in ch. 6-7 I will show why this isn’t so. So there is an open question about the phenomenology of meaning until then. However, we can say something about what the phenomenology is not like.

The phenomenology of meaning is not like the phenomenology of colour. The responses that determine meaning are not elicited by experiential scenario, such as being confronted with a chameleon. This picture seems wrong in the case of responses to contentful mental states: we do not have an inner sense, like a mental gaze, which we turn on our own minds. Indeed, the whole idea of confrontation with our own minds seems off-colour. This is why I want to avoid the picture of experiential confrontation and elicited response. I prefer talk of use-dependence.

But the phenomenology of meaning should not be cashed out in intellectualist terms either. We do not have to deliberate and judge about meanings in order to mean things by our sentences—once we are competent with a language its phenomenology is much more immediate. Therefore, Wright’s notion of judgement-dependence would not be appropriate for the concept of meaning. According to this notion the $a$ priori biconditionals in UD would be like this:

\[
\text{JUD: } \varphi \text{ means } p \equiv S \text{ is disposed in } C \text{ to judge that } \varphi \text{ means } p.
\]

---

23 As for example in Johnston 1992.
24 Wright's distrust in dispositionalism notwithstanding.
But not only does JUD seem out of accord with the phenomenology of meaning, it would also be a bad analysis of the concept of meaning. Imagine that we ramsify JUD by stripping away all occurrences of \("\varphi\ means p\) and replace them with distinct variables. Now ask: does what is left of the platitudes give us enough of a network to show what is response-dependent about the concept of meaning? I think the answer is no: we do not have enough of an independent grasp of locutions like \("\text{judge that } q\) to justify the claim that judgements about meaning are extension-determining rather than extension-reflecting. How can we understand \("\text{judge that } q\) without understanding what it is for someone to \text{mean } q when issuing a judgement? The problem is that uniformity isn't guaranteed as long as occurrences of \("\text{judge that } q\) are not stripped away from the platitudes. In response to this objection, the occurrences of \("\text{judge that } q\) could be deemed versions of \("\varphi\ means p\) and hence stripped away from the platitudes too. But \("\text{judge that } q\) plays such a central role in JUD that stripping it away from the platitudes would leave us with too little content in the \textit{definiens} to form any reliable opinions about the concept we wanted to define by JUD.\textsuperscript{25, 26}

\textsuperscript{25}A worry which dovetails with this is also reflected in Boghossian's criticism of Wright's notion of judgement-dependence for the concept of mental content: 'An intuitive difficulty should have been clear from the start. A 'judgement-dependent' conception of a given fact is, by definition, a conception of the fact according to which it is \textit{constituted} by our judgements. The idea is [...] impossible as a conception of facts about mental content. For it cannot in general be true that facts about content are constituted by our judgements about content: facts about content, constituted independently of the judgements, are presupposed by the model itself', 1989, p. 547.

\textsuperscript{26}I think that a similar criticism can be mounted against Pettit's theory (as presented in 1990, 1991, 1993 and forthcoming). Pettit doesn't endorse judgement-dependence, but does nevertheless rely on substantial normative terms in his account of the functional role of C-conditions (which are crucial for the possibility of error). Thus, subjects are such that, in cases of discrepancy, they negotiate with each other to find out who is in error. The conditions under which subjects, on the basis of such negotiations, turn out not be in error are the normal or ideal conditions. The problem concerns the status of these negotiations in the account of the C-conditions: either mention of them should be stripped away from the specification of the C-conditions or not. If we leave them in, then we leave in substantial normative vocabulary and thus we don't get a reductive analysis, and, more seriously, we do not get a guarantee that the account is uniform, \textit{i.e.} that those normative terms are themselves response-dependent. If we strip them away, then there will not be enough left in the specification of the C-conditions for them to play the role Pettit assigns to them: \textit{viz.} that of allowing us to aim at the true and rational.
4. Refining and restating the theory and conception of use-dependence.

*a) Refining and restating the theory of use-dependence.*

My claim is that, if the concept of meaning is the concept of a dispositional use-dependent property, then the five platitudes will come out as compatible and the three tensions among the platitudes will be resolvable. To this end I have aligned use-dependence with the broader notion of response-dependence. In particular, I have discussed the general purpose of putting forward a response-dependent account of a concept, and I have set out four constraints on such an account. I argued that use-dependence should be a dispositional type of response-dependence, where the 'response' is 'being disposed to use sentences to indicate contentful mental states'. With these considerations in mind, as well as the general platitude-based considerations, I will now go on to refine and then restate my preliminary first version of use-dependence, *viz.* that the *a priori* analysis of the semantic relation expressed by 'φ means p' implies *a priori* definitional biconditionals of this form:

\[
UD_{v.2}:
\]

φ means p in S's language ≡ _def_ S is disposed to use φ to indicate all and only mental states with the content p.

Consider first the LHS of the *a priori* biconditional:

(1) φ means p in S's language...

We must understand meanings of sentences as relative to a speaker, S, and his or her language. It would not make sense to claim that the way S uses the sentence, 'marsupial moles are born blind', constitutes the meaning of that sentence in some language other than the language S speaks. Not only would such a claim be ludicrous, it would also be in breach of the meaning is use platitude as it would entail that the
meanings of sentences in my language L, could be constituted by the way someone else uses the sentences in his or her language L*.

The platitudes are also neutral in the debate about the relative priority of idiolect and communal language so UD should make space for both views. Hence we get:

(2) φ means p in S's communal language or dialect or idiolect...

That is, *prima facie*, the way I use a sentence may constitute its meaning in my communal language, or my idiolect or dialect.

Next, 'φ' ranges over declarative sentence types. I settle on *declarative* sentences for reasons of convenience. The debate about the concept of meaning, and the issue of meaning-scepticism, is keyed to the idea of whether meaning-facts are such that what is said can ever be false. It is therefore easier to conduct the debate for sentences, which can be true or false. Moreover, we have a fairly clear and uncontroversial idea of what the meanings of declarative sentences are, *viz.* propositions (though it is controversial how to further account for what propositions are), but not so clear an idea of the meanings of other types of sentences (*e.g.* interrogatives and imperatives). I settle on sentence *types* because I wish to bracket the contribution of context to the determination of meaning. The meanings of token sentences are often sensitive to context: *e.g.*, the utterance "Jones is hungry" may be assigned different propositional content depending on the time of utterance. But there are many ways in which context can enter interpretation so, in order to keep things manageable, I will only talk of the meanings of sentence types, *e.g.*, the context-independent meaning which is in common for all utterances of the sentence 'Jones is hungry' in S's language. Hence, the LHS of the *a priori* biconditional concerns φ's meaning in S's communal language or dialect or idiolect, where φ is a declarative sentence type.
The RHS of the *a priori* biconditional must be overtly dispositional:

\[(3) \ldots S\text{ is disposed in } C \text{ to use } \varphi \text{ to indicate all and only mental states with the content } p.\]

That is, if S were in C, then S would use \( \varphi \) to indicate all and only mental states with the content \( p \).\(^{27}\) The motivation for this is that declarative sentences are used to express what one believes is the case or desires to be the case, often with the intention to inform others about this or to remind oneself. Therefore, sentences are uttered when we have this or that contentful mental state. The idea is simply that a sentence type's meaning is determined by which contents of mental states it is used to indicate the presence of. To give this notion of 'use of sentences' sufficient stability and determinacy, 'use' is characterised in dispositional terms: the idea is that such a dispositionalist understanding of use will ensure that it is *all and only* mental states with the content \( p \) that are indicated by use of the sentence. The account would fail if \( \varphi \) could be used to indicate all sorts of other contents too. Relatively few sentences may be ambiguous (e.g., 'platypuses are electrifying'), but the majority of sentences are not. Above I defended my employment of the general notion of dispositions, and in subsequent chapters I will explain why dispositionalism is apt for the particular concept of meaning. Here I shall also substantiate the important notion of C-conditions.

A worry quickly arises: we do not always say what comes to our minds, so it cannot be the case that we are disposed to say, e.g., 'platypuses are electrifying' every time we have a mental states with the content *platypuses are electrifying*. We may have the mental states without saying anything, and yet the sentence is meaningful. Whether or not we say anything depends on our beliefs and desires about any given conversational situation: are there people to talk to? am I talking to myself or others? do I desire to participate? would it be appropriate to participate? and so on. This is an

\(^{27}\text{Modulo, of course, the correct counterfactual analysis of dispositions.}\)
important aspect of linguistic phenomenology but for present purposes we can set
such mental states aside, and simply stipulate that the dispositions in UD are
dispositions to use sentences to indicate contentful mental states given normal
background beliefs and desires about the conversational situation. This issue will
surface again in Chapter 6 because this stipulation really concerns the specification of
the C-conditions for the dispositions in question.

We can now restate UD in its refined version. The a priori analysis of the semantic
relation expressed by ‘φ means p’ implies a priori definitional biconditionals such that,
where ‘φ’ is a declarative sentence type and p is a proposition:

UD v. 3:
φ means p in S’s communal language or dialect or idiolect ≡def S is disposed in
C to use φ to indicate all and only mental states with the content p.

Where C contains a specification of normal and standard conditions for the disposition
which include a specification of the speaker’s beliefs and desires about the
conversational situation, and how he or she wants to participate in it.28

b) Clarifying the UD-based conception of meaning.
There are various dimensions of UD which must be clarified. Johnston, for example,
specifies three versions of the claim that a concept is the concept of a dispositional
property. So, for the concept of redness:

(A) it is a priori that: the property red = the standardly realised disposition to
look red to standard perceivers under standard conditions.

28For the sake of brevity I shall suppress the distinctions between communal languages, dialects and
idiolects in later statements of UD
(B) it is \textit{a priori} that: the property of being red = the disposition to look red to standard perceivers as they actually are under standard conditions as they actually are.

(C) it is \textit{a priori} that: the property red for subject $S_i$ under conditions $C_i$ = the disposition to look red under conditions $C_i$.\textsuperscript{29}

Here (A) is supposed to be neutral between (B), which says that it matters for redness how standard perceivers and standard conditions are in the actual world, and (C), which says that it matters for redness how perceivers and conditions are at the world under consideration. I think that if colour is indeed a response-dispositional concept, then something like (B) will have to hold.\textsuperscript{30} We are fairly committed to the idea that colours remain constant even if our optical faculties and the conditions were to change. Colours are keyed to how we and the standard conditions actually are.\textsuperscript{31}

Compare this with the corresponding issue for the concept of meaning:

(A) it is \textit{a priori} that: the relation expressed by "$\varphi$ means $p$" = the standardly realised disposition of standard speakers in standard conditions to use $\varphi$ to indicate all and only mental states with the content $p$.

(B) it is \textit{a priori} that: having the relation expressed by "$\varphi$ means $p$" = having the disposition of standard speakers as they actually are, in standard conditions as they actually are, to use $\varphi$ to indicate all and only mental states with the content $p$.

\textsuperscript{29}Johnston 1992, p. 230.

\textsuperscript{30}Cf. also Wright 1992, p. 113.

\textsuperscript{31}For example, we are committed to the idea that the colour of a piece of coral is constant no matter how it appears at increasing depths with less and less red light (if we want to see the real colour we bring torches). (Though this is not univocal: I enjoy looking at coral fishes at between 5 and 20 meters depth because they are extremely colourful at that depth; but once they are pulled out of the water and into normal conditions they cease to appear so beautiful).
(C) it is *a priori* that: the relation expressed by ‘φ means p’ for a subject $S_i$ under conditions $C_i$ = the disposition of $S_i$ in $C_i$ to use φ to indicate all and only mental states with the content $p$.

Here I think we are more inclined to go with (C). What matters for facts about the meanings of our sentences is how we are disposed to use them here and now, not how standard speakers are actually disposed. If our dispositions change, then the facts about meanings also change. We normally have very little commitment to the idea that if our dispositions for use change over time, then we should be interested in how the dispositions for use, and hence meanings, were in the ‘old’ world. What matters to us is what our words mean here and now because this is what we exploit when we express thoughts and transmit information. Of course, we might, when doing etymology or hermeneutics, be interested in what sentences meant in the past; or when interpreting foreign languages be interested in how things would be if usage patterns were different. But it is hard to see how we could do this without be interested in our own language in the first place (e.g., when transmitting information about how we believe the word ‘transubstantiation’ was used a thousand years ago).32

These considerations put another aspect of UD into its right perspective. In UD meanings are fixed relative to S’s actual language. This is a way of rigidifying our talk of meanings: the meaning of ‘kookaburras are meat-eaters’ in S’s language L, and in S*’s idiolect L*, may not be the same. Relativising meanings to languages is just a way of saying that we cannot divorce S’s meanings from his or her language. This is consistent with the non-rigidified (C)-reading of UD because which language or dialect or idiolect S has depends on S’s dispositions for use, no matter how S is.

---

32The intuitions I have sought to bring out are related to the considerations about primitivism and causal theories in the previous chapter. If it did matter more to us how the Romans and Greeks used words than how we use words, then a causal theory could perhaps be feasible; and if it did matter to us how words are used in some ideal world, then primitivism would be attractive.
When we consider the status of the C-conditions, matters are complicated somewhat. If we go with the (C) reading of UD, then we should be prepared to say that the normal and standard conditions, C, under which S is disposed to use a sentence, are whatever conditions are standard in the world under consideration. This seems intuitively right for it doesn’t matter for facts about meaning if the normal and standard conditions change over time. But here we should be careful. The analysis of the concept of meaning has brought out that it is the functional role of the concept that matters to us, not what realises this role in different situations. Thus it doesn’t matter for facts about meaning what realises the C-conditions for a disposition to indicate contentful mental states in this or that world. But it would matter to us if changes in C-conditions alters the functional role of the concept.

Assume that a desire to deceive became part of the standard conditions under which we are disposed to use sentences to indicate mental states. In that case we would no longer recognise the property of being disposed to indicate mental states as the concept of meaning: it would neglect the platitudes about communication and expression of thought, as well as the platitude about purposeful disagreement about meaning.

Such a change in C-conditions would be in conflict with the platitude about communication and would thus change the very content of the concept of meaning. Therefore, whereas we can countenance changes in the dispositions of S, and in the realisers of the functional role of the concept of meaning, we cannot countenance changes that alter its functional role. Accordingly, the platitude-based functional role which I have identified for the concept of meaning must be constant. Apart from that we do not need to rigidify.

Let it be implicit in UD, on the basis of the above considerations, that meanings in S’s communal language, dialect or idiolect are relative to S’s normal and standard
dispositions in the world under consideration, in the normal and standard C-conditions that enable the functional role of the concept of meaning to be fulfilled.

c) Remarks on the UD-based conception of meaning.

UD gives natural expression to the platitude that meaning is use because it explains how 'wobbegongs are awesome' comes to mean *wobbegongs are awesome* in S's language in terms of the use S makes of the sentence. If S had used the sentence to indicate mental states with the content *blue gropers are fearful*, then the meaning would have come out correspondingly. Thus, if UD is true, and satisfies the four constraints on a successful response-dependent account of a concept, then the meaning is use platitude comes out true. It is important to note that the constraints must be satisfied because UD as it stands is in fact consistent with the primitivist idea that meaning doesn't supervene on use. A primitivist could claim that UD is indeed necessary for meaning because meanings are what explains S's dispositions for use of sentences, but *add* that UD isn't *sufficient* for an account of meaning because it must also hold that meaning-facts are primitive. In the previous chapter I have shown that the primitivist's position is untenable. But this is not the same as showing that UD is necessary and sufficient for an account of meaning. To do that I must show that the four constraints are satisfied in such a way that the three tensions are resolved, and the platitudes come out true, without having to undergo revisions that will threaten the overall functional role of the concept of meaning.

The communication platitude says that sentences are used to indicate how things are believed or desired by the speaker to be. This platitude was motivated by concerns about the possibility of communication in language by expression of our thoughts. If an account of the concept of meaning shows that communication is not possible, or that we cannot express our thoughts, then something has gone wrong. UD is formulated in terms of dispositions to express contentful mental states so it should allow this platitude to come out true. It doesn't say anything about what it is for us to
express our thoughts, or what kinds of thoughts we can express, and neither does it say anything about how communication may be possible. These are highly complex and controversial matters which go beyond an analysis of the concept of the relation expressed by "\( \varphi \) means \( p \)". It is important to see that whereas UD does nothing to show what else has to be in place for expression and communication of thought to be possible, it is nevertheless designed such that this is a prima facie possibility, given the contribution of use-dependence to this issue.

UD’s contribution to the remaining three platitudes (explanatory meaning, error and robustness and first-person authority) is not so clear. This is a reflection of the three tensions I discussed in Chapter 2. The meaning is use platitude is in apparent tension with all of these three platitudes. Since the meaning is use platitude is given primacy in the formulation of UD this is what we should expect. The burden of a defence of UD is to show how the use-based conception of meaning can address the tensions and thereby make it clear how the remaining three platitudes may come out true too. I shall do this in the three subsequent chapters.

5. Concluding remarks.

In this chapter I set out to define a strategy for platitude-compatibilism, that is, the position which acknowledges that an analysis of the concept of meaning must respect all five platitudes surrounding the concept. I suggested that an account in terms of use-dependence, UD, will be able to do the job. UD is a version of the broader family of response-dependent accounts of concepts which were discussed at some length. I argued that if the purpose of UD is to show \textit{a priori} that and how meaning is dependent on our use, then an amended version of Wright’s four constraints for an

\footnote{Though for matters of expository convenience I have settled on expression of thoughts in declarative sentences.}
order of determination for response-dependent concepts is needed. In addition, UD must be given a dispositionalist reading. This should not be terribly controversial because we know that there are dispositions, so UD should be allowed to avail itself of this general notion. It is a further question, of course, whether this is a good choice in an account of meaning. Finally, I restated UD with the considerations from the discussion of response-dependence in mind, and made some further remarks on the use-dependent conception of meaning.

Now the task for the platitude compatibilist is set out clearly. The compatibilist must show that UD is necessary and sufficient to make the platitudes, or functionally equivalent revised versions of the platitudes, come out true.
CHAPTER V

RESOLVING THE FIRST TENSION:
THE "MEANING IS USE" PROBLEM

ARGAN: Mihi by docto doctore
Domandatur causam and rationem, quare
Opium facit dormire.
To which respondeo,
Quia est in eo
Virtus dormitiva
Cujus est natura
Sensus stupifire
CHORUS: Bene, bene, bene, bene
respondere...
Moliere, *The Imaginary Invalid.*

1. Use-dependence, explanatory meaning and the "meaning is use" problem.

In Chapter 2 I offered an analysis of the concept of the relation expressed by "φ means p". Part of the analysans were two platitudes: 'the meaning is use platitude' and the 'explanatory meaning platitude'. I discussed a tension between these two platitudes which I called the "meaning is use" problem. In this chapter I shall demonstrate how the theory I call use-dependence (UD) can resolve this problem while preserving the platitude-based analysis. First I show how UD provides asymmetrical explanations of use and meaning. Then I substantiate this result by discussing the phenomenology of the explanatory meaning platitude in more detail. I then discuss a problem raised by Mark Johnston, the so-called Missing Explanation Argument. This problem bears some resemblance to the "meaning is use" problem, and I show how my solution to that problem also solves Johnston's problem.
The "meaning is use" problem is that, as they stand, the two platitudes (i.e. meaning is use and explanatory meaning) are in tension. The first platitude says meaning is explained by use, and the second says that use is explained by meaning. This is a problem because it straightoff seems to flout the principle that explanations be asymmetrical (i.e. the principle that if A explains B, then B cannot explain A). I argued, in Ch. 2, that the problem need not be unassailable because the two platitudes may endorse two different types of explanation. I noted, however, that specifying these types of explanation can force us to adjust the formulation of the platitudes accordingly. This may be quite substantial: specifying a particular explanatory context may have ramifications for the conceptions of the explananda and explanantia. That is, the adjustments may have ramifications for the conceptions of the very notions of meaning and use.

I think that UD shows how this problem can be resolved in a way that is faithful to the motivations behind the two platitudes. Here is the theory of use-dependence. The \textit{a priori} analysis of the semantic relation expressed by \textit{`φ means \( p \)} implies \textit{a priori} definitional biconditionals such that, where `\textit{φ}` is a declarative sentence type and \( p \) is a proposition:

\[
\begin{align*}
\textit{UD v. 3:} \\
\text{φ means } p \text{ in S's language } &\equiv_{\text{def}} \text{S is disposed in C to use } \varphi \text{ to indicate all and only mental states with the content } p.
\end{align*}
\]

My claim is that the dispositionality of UD provides the means to resolve the tension. I must then show that UD sustains both:

(1) use explains meaning

and:
(2) meaning explains use,

without flouting the principle of explanatory asymmetry.

UD sustains (1) in this sense: use explains meaning because meaning supervenes on correct use, i.e. if S possesses a disposition to use φ to indicate mental states with the content p, then φ means p in S’s language. For the sake of simplicity let us say that, for an S at a world, w_i, the following identity holds:

(3) φ means p in S’s language = S possesses the disposition D to use φ to indicate the content p.

This entails that whenever we want to talk about the meaning of φ (for S, at w_i) we talk of S’s possession of the disposition D. This seems problematic if (2) requires that we employ the meaning of φ to explain S’s possession of the disposition D to use φ. It is problematic because it violates the principle of explanatory asymmetry like this:

(4) (meaning explains use) & (meaning = use).

That is, the meaning of φ cannot explain possession of a disposition D because talk of the meaning of φ is already talk of possession of D.

However, this problem is easily circumvented because the explanatory meaning platitude should not concern explanations of possession of dispositions for use. Rather, it should concern manifestation of dispositions for use. What we want to explain in terms of meaning are occasions where φ is used by S. That is, in dispositional terms, we want to explain manifestations of dispositions for use of φ.
And UD, via the principle that possession of a disposition explains its manifestation, can sustain just this kind of explanation. From

(3) \( \varphi \) means \( p \) in S's language = S possesses the disposition \( D \) to use \( \varphi \) to indicate the content \( p \),

together with the principle:

(5) S's possession of the disposition \( D \) to use \( \varphi \) to indicate the content \( p \) explains S's manifestation of the disposition to use \( \varphi \) to indicate the content \( p \),

we get:

(6) (the fact that) \( \varphi \) means \( p \) in S's language explains S's manifestation of the disposition to use \( \varphi \) to indicate the content \( p \).

(6) does not violate the principle of explanatory asymmetry because possession of a disposition is not identical to, or explainable by, its manifestation (compare: the fragility of a glass is not explained by the fact that it breaks, it could break without being fragile). That is, considerations of different explanatory contexts lead us to disambiguate the notion of use such that there is no clash with the principle of explanatory asymmetry. This resolves the "meaning is use" problem. Thus, from:

(7) possession of \( D \) explains the meaning of \( \varphi \),

and:

(8) the meaning of \( \varphi \) explains the manifestation of \( D \)

we only get:
(9) possession of D explains the meaning of φ, and the meaning of φ explains the manifestation of D.

Of course, the principle in (5) may be controversial: it is not transparent how possession of a disposition may explain its manifestation. But I think this is immaterial to my argument. What matters is that the notion of use can be disambiguated via the distinction between possession and manifestation of dispositions. This is what is needed to resolve the “meaning is use” problem.

2. Getting the phenomenology right: varieties of explanation.

My claim is that the “meaning is use” problem can be solved because the notion of use occurs with two different senses in the explanatory contexts associated with, respectively, the meaning is use platitude and the explanatory meaning platitude. But is this claim justified? That is, is explanation of the manifestation of dispositions for use really the kind of explanation which the explanatory meaning platitude requires? To answer this question we must look at what the platitude requires: which type of explanation is involved, what kind of use is getting explained, in what sense is meaning doing the explaining?

a) The type of explanation.

The explanatory meaning platitude says that sometimes meanings can explain uses of sentences. The notion of explanation here is empirical explanation. There is an empirical phenomenon, a use of, say, ‘that is a kookaburra’, which we wish to explain by citing beliefs about what the utterance “that is a kookaburra” means in the

---

1There may be some unease with saying that the disposition causes its manifestation. Notice that, irrespective of which type of explanation we are talking about, the principle seems a priori: it is a priori that a disposition may explain its manifestations. In contrast it is not a priori which dispositions explain which manifestations.
mouth of the speaker. It is empirical in the sense that it is contingent which fact about meaning provides the best explanation of the phenomenon: it could be the meaning *that is a kookaburra*, but it could also be the meaning *some birds are carnivores* (in Australian English the first one is the best explanation, I believe).

It is not empirical explanation in the sense that one must necessarily investigate the environment in order to provide such explanations. This is because the platitude is consistent with explanations of use in terms of first-person authority. For example, “I used ‘that is a kookaburra’ on that occasion because it was a kookaburra and I mean *that is a kookaburra* by that sentence”. If we do have first-person authority, then such explanations should not necessarily proceed by investigating the environment. To say that it is empirical explanation, then, is to say that it must be potentially informative. This is not to say much. An explanation may be potentially informative even though it looks somewhat trite and trivial. For example, explaining a doorstop’s being a doorstop by adverting to its being used as a doorstop may not strike us as particularly informative, but things could have been otherwise: the doorstop could have been a paperweight, if it had been used as a paperweight. What is ruled out, then, are explanations of facts described as A by citing facts described as B, where A and B are necessarily equivalent and have the same senses and literal meanings. I believe this requirement is an expression of explanatory asymmetry, *viz.:

(1) If A explains B, then A ≠ B,

and as such the above solution to the “meaning is use” problem satisfies it.

It is of course a fairly weak requirement on what can count as empirical explanation. It rules in trivial sounding explanations like:
(2) S said 'he is a bachelor' because that sentence means \textit{he is an unmarried male of the marriageable age} in S's language.

This is not wholly trivial because it conveys empirical information about a relation between two distinct states of affairs: the utterance and facts about meaning. That they are distinct is shown by the fact that the utterance could have been related to some other meaning, \textit{e.g.} to the meaning \textit{he has a three year university degree}. Such explanations may sometimes be easy to come by—as in, for example, the introspective case—but this doesn't detract from their informativeness. What is easy for me to obtain may be interesting for someone speaking another language.

The reason explanations like (2) look trivial is that they look a bit like what Johnston calls 'conceptual articulations', \textit{e.g.}:

\begin{quote}
(3) he is a bachelor or is rightly called 'a bachelor' because he is an unmarried male of the marriageable age.\footnote{Johnston 1991, p. 124.}
\end{quote}

Conceptual articulations are probably not genuine empirical explanations because they are a way of specifying meaning in terms of correct use. And, according to the meaning is use platitude, and UD, talking about the correct use of a sentence is already a way of talking about its meaning. That is, I think conceptual articulations really have this form:

\begin{quote}
(4) 'he is a bachelor' means \textit{he is an unmarried male of the marriageable age} $\equiv$
the sentence 'he is a bachelor' is correctly used for all and only unmarried males of the marriageable age.
\end{quote}

In so far as they have that form they are not concerned with explaining particular utterances of a sentence at all. It is, of course, something like (4) which \textit{grounds}
explanations of particular uses of sentences. That is, we should accept explanations of the form:

(5) S said ‘he is a bachelor’ because something like (4) holds of S’s language.

This is quite consistent with my solution to the “meaning is use” problem because (5) is of course just a more elaborate version of (2).

b) What is getting explained by meaning?

The explanatory meaning platitude says that sometimes meaning explains uses of sentences. But we can have various conceptions of uses of sentences: is it my use which I seek to explain, is it your use which I seek to explain, or do we in some sense seek to explain our use? These distinctions are important because different conceptions of the explanandum may require that we draw on different explanatory resources. For example, I look at your behaviour and history when I want to explain your use of ‘that is a kookaburra’ (it would for example make sense to ask you “would you call that a kookaburra?”). But if I seek an explanation of why I myself said something I may introspect (whatever that means) and discern my beliefs and desires, or seek some plausible explanation of, say, why I made a slip of the tongue. In the communal case we might seek an explanation of why we at all have a word such as ‘kookaburra’, that is we might be interested in its etymology, or its social function (e.g., “why do we have that word when they haven’t?”). Again, these different conceptions may overlap when it comes to explaining particular uses: I may be helped in my self-scrutiny by someone offering psychoanalysis, perhaps I draw on communal patterns in my explanation of why you said what you said, etc.

I do not think the explanatory meaning platitude divides sharply between these conceptions of use. The motivation that underlies the platitude seems neutral between them and therefore they should all be compatible with the platitude. It seems to me,
however, that it is the first-person perspective on use which in some sense must be basic: it arguably makes sense on its own, and the other two perspectives hardly make sense without it. The communal perspective is pervasive, but can in principle be left out. Sometimes we wish to explain idiolectic uses in terms of idiolectic meanings (e.g., when we try to explain a child’s idiosyncratic use of words). And though it is possible for someone to obtain a third-person perspective on someone else’s use, it seems unwarranted to claim that it is a perspective on a semantic phenomenon at all if the user herself cannot obtain a first-person perspective on her own use.3 Lastly, there is nothing prima facie wrong with someone in isolation from other speakers obtaining a first-person perspective on her own use of sentences (though the scenario is perhaps a bit odd as we normally use sentences when there are other people around).4

How does the solution to the “meaning is use” problem fare if it is correct that first-person conceptions of use are basic? I think it fares well: if UD is right then what is explained are manifestations of an individual’s dispositions for use of sentences. This should be consistent with all three conceptions of use, as well as with the idea that the first-person perspective is basic: my individual use may (or may not) be in convergence with the use of the other members of my community. My individual use may (or may not) be accessed by someone obtaining a third-person perspective on me. And, prima facie, I may have a first-person perspective on my own use whether or not that use is in accordance with communal use, and whether or not someone else can obtain a third-person perspective on it.

3This is presumably the intuitive point behind Davidson’s argument that we must have first-person authority to be interpretable, cf. Davidson 1990. 4All this, modulo, of course, the scope and validity of the Wittgensteinian ‘private language argument’.
c) **What is doing the explaining of use?**

Finally, we must discuss what is doing the empirical explaining of an individual’s use of sentences. The explanatory meaning platitude says that meanings explain uses, but which conception of meaning is this?

While all uses of sentences must have causes, the explanations we are after cannot be wholly causal. Wholly causal explanations could be explanations that explain uses in terms of, say, neurological states of the individual in question. This cannot be the type of explanation which is required to make the platitude true. Firstly, there are neurological causal explanations of uses of expressions which are not explanations in terms of meanings (e.g., someone with Tourette’s Syndrome using ‘hideous’ repeatedly in a meaningless way). This shows that wholly causal explanations are ill suited to distinguish between uses that are motivated by meaning and uses that are not motivated by meaning at all. Secondly, only neuroscientists would be able to give this kind of causal explanation, and this is in conflict with the platitudinous idea that we can all more or less effortlessly engage in such explanations.

Naturally, it must be consistent with UD that wholly causal explanations of uses are possible, but the point is that these explanations are not what we are after. In so far as causation is involved in the explanations at all, it must be clear that uses are caused by meanings in some sense that is consistent with the speaker’s use being motivated or intentional. This indicates an important aspect of the explanatory meaning platitude. The unease we feel about wholly causal explanations has to do with the concomitant picture of subjects that are being pushed and pulled in various ways by causal forces. This doesn’t clearly leave room for the idea that it is grasp of the meaning of a sentence, and thereby of its correct use, that explains a speaker’s use of the sentence. Without an aspect of grasp of meaning there will be no way of sustaining the idea that uses of sentences are intentional or motivated.
I think this leaves us with two candidates for what explains use: a communal conception of explanatory meanings, and a dispositional conception of explanatory meanings. A communal conception of meaning would provide explanations of use that go like this:

(6) She said ‘there is a kookaburra’ because in her linguistic community ‘there is a kookaburra’ means \emph{there is a kookaburra}. 

This is an important type of explanation, and one I think we in fact often employ. But I think it is at one remove from the core of the explanatory meaning platitude. (6) really has this two-stage form:

(7) i) ‘there is a kookaburra’ means \emph{there is a kookaburra} in S’s linguistic community; 

ii) S means \emph{there is a kookaburra} by ‘there is a kookaburra’; 

iii) S said ‘there is a kookaburra’ on this occasion because of ii), and ii) is the case because i) is the case.

That is, we explain an individual’s utterance by adverting to what they mean, and we explain why individuals mean this rather than that by adverting to the linguistic community they happen to belong to. This suggests that there is a type of explanation which cuts closer to the bone, \emph{viz.} explanations that take it for given that S means something by an utterance, and then explains uses of the sentence without taking a stance on why the sentence has come to have that particular meaning in S’s language. Thus, the explanations which I think are required to make the platitude come out true are shortened versions of (7):

(8) a) S means \emph{there is a kookaburra} by ‘there is a kookaburra’; 

b) S said ‘there is a kookaburra’ on this occasion because a) is the case.
This is in fact a bit easier to state in dispositional terms: we advert to linguistic communities when we want to explain why S possesses this rather than that disposition for use of a sentence; but when we want to explain S’s manifestations of her dispositions we advert to her dispositions, no matter how she has come to possess them. This is what I call the dispositional conception of explanatory meaning, and it captures what I think is the core notion of the explanatory meaning platitude.

Apart from the fact that the communal and dispositional explanations can be split apart (as in (7) and (8) above) there is some evidence that we do in fact split them apart. There are cases where we recognise some use of a sentence as a semantic phenomenon, but where we do not find it appropriate to embark on a communal explanation of it because we suspect it is an idiolectic use and we want to explain it as such. Such an explanation may go like this:

(9) S used the sentence ‘there is a kookaburra’ because for her it means there is a furry bird.

This type of explanation is intended to reflect the idiolectic nature of the use, and therefore it would not be relevant in this case to advert to the communal meaning of the sentence. Such explanations are clearly possible and, one’s theory should accommodate them.

If all this is correct then it fits my solution to the “meaning is use” problem. If the core notion of the explanatory meaning platitude is what I have called a dispositional conception of explanatory meaning, then the uses which are getting explained are best conceived of as manifestations of dispositions for use of sentences, irrespective of why S has come to possess this or that particular disposition for use. Thereby we can forge the disambiguation on the notion of use which the solution to the problem requires. S’s possession of dispositions for use of a sentence explains its meaning in
S's language or idiolect, and meanings, conceived in that way, explain manifestations of the dispositions.

This should be no more mysterious than this case: we may explain a glass's being fragile by adverting to its possession of a disposition to break when struck (in contrast to, e.g., its disposition to form a puddle over time); and we may explain the manifestation of the disposition, viz. its breaking, by adverting to the disposition (in contrast to its disposition to break when there are nuclear explosions nearby, which the glass shares with a lot of non-fragile things).

Of course, it is legitimate to ask how this is a story about how grasp of meaning explains use, and thereby how it is different from a wholly causal explanation. It is different from a wholly causal story in the following way. On a wholly causal story we lose sight of the very idea that it is meanings that should be doing the causing. This is because wholly causal explanations are non-normative, in as much as they do not distinguish between correct and incorrect uses. If I say "echidnas eats ants" then that may be a correct or an incorrect use of that sentence, but in either case it will have a wholly causal explanation which doesn't reflect whether it was correct or incorrect. In contrast, if UD generally is on the right track, then a dispositional explanation would have the resources to make this distinction. Then the correct use of 'echidnas eats ants' is given by how S is disposed to use the sentence in C. Therefore, giving dispositional explanations of manifestations will bear on whether the manifestation is in accordance with the meaning or not. So if UD as such holds up, then dispositional explanations will be explanations in terms of meanings (I substantiate this in Ch. 6). It is then a further question as to what extent we can grasp and hold beliefs about dispositionally characterised meanings. That question concerns the epistemology of meaning and I will address it in Ch. 7. Notice, for now, that once we see that

---

5This point mirrors some of the arguments I presented to the wholly causal theory of meaning in Ch. 3: wholly causal theories make us lose track of what it means to make errors.
dispositional explanations are not wholly causal explanations, it should become more plausible that an adequate epistemology can be found. Concepts of dispositions are concepts of functional properties and as such they leave room for requirements that mention, e.g., epistemology, and this is indeed what is reflected in the five platitudes I have argued surround the concept of meaning.

In this section I have discussed the phenomenology of explanations of use in terms of meaning. The purpose was to see whether the dispositional solution I provide to the "meaning is use" problem fits with the motivation behind the explanatory meaning platitude. I found that the solution does indeed fit the phenomenology. The core notion of explanatory meaning fits a dispositional mould nicely. I distinguished this from wholly causal explanations (because meaning dispositions are dispositions for correct use), and from explanations of use in terms of communal meanings (because communal explanations have dispositional explanations as parts). Both causal and communal explanations are consistent with dispositional explanations. I argued that if UD is right, then dispositional explanations can indeed be explanations in terms of meanings (because they are explanations in terms of correct use), but left a question-mark over the epistemology of meaning.

3. The Missing Explanation Argument.

In a series of articles on response-dependence, Mark Johnston has raised a problem for the notion, based on what he calls the Missing Explanation Argument (MEA, hereafter). MEA is supposed to be a general problem for any response-dependent account of a concept, and thereby also for UD. Moreover, it can be seen as related to the "meaning is use" problem because it says that response-dependent accounts make a certain kind of explanation goes missing (in the case of UD: explanation of use by

---

meaning goes missing). I believe my solution to the "meaning is use" problem can be worked into a solution to the problem raised by MEA. In order to show this, I will discuss MEA in general and then demonstrate how explanations of manifestations of dispositions in terms of possession of dispositions will eliminate the problem raised by MEA.

My response to MEA is intended to be in line with the response of Menzies and Pettit (1993), though I motivate my solution in a slightly different way. Menzies and Pettit focus on a response-dependent account of the concept of redness. They argue that the a priori biconditional for redness concerns things which possess certain dispositions, whereas the empirical explanations (S judges that x is red because x is red) concern the manifestations of the disposition to look red on particular occasions. Once we acknowledge this, they argue, we can paraphrase the biconditional and the explanations accordingly and thereby prevent MEA from applying. I agree with this line of argument, and in what follows I will motivate it for UD and the concept of meaning.

a) Response-dependence and MEA.

Assume we agree that for a certain concept the following holds a priori:

\[(T) \, x \, \text{is} \, F \equiv S \, \text{Rs} \, F-\text{ly to} \, x \, \text{in} \, C.\]

for example, it is a priori that:

---

7Menzies and Pettit 1993, p. 103-5.
8Wright 1992, p. 129-30, presents related arguments. I also find Blackburn's discussion in his 1993, §3 congenial, partly because he allows distinctions based on what contingently happens on particular occasions. In a recent article Johnston (1998, p. 26) credits Wright for proposing (in an unpublished manuscript) the manifestation-possession distinction as a solution to MEA. I thus agree with Wright too. However, such a proposal must be defended by drawing on the detail of the platitudes surrounding the concept in question; it is not given that it will work for all response-dependent concepts (that is why I spent time trying to get the phenomenology right in §2 above). Johnston argues that the distinction isn't sufficient to avoid MEA because we would still want the relevant facts to explain possession of dispositions (1998, p. 17-18). I view my discussion in section b.i-ii. below as a defensive answer to Johnston.
(1) \( x \) is pious \( \equiv \) \( S \) judges in \( C \) that \( x \) is pious.

That is, assume that piousness is the concept of a response-dependent property.\(^9\)

Now, we might have a strong intuition that it should still be possible to give true explanations of this kind:

(2) \( S \) judges that \( x \) is pious \textit{because} \( x \) is pious,

and that we are potentially getting empirical information when someone offers us explanations like (2). But now assume the principle that \textit{a priori} equivalents (like (1) above) can be substituted in explanatory contexts.\(^{10}\) Then we get the following:

(3) \( S \) judges that \( x \) is pious because \( S \) judges that \( x \) is pious.

And (3) is not particularly enlightening. In Johnston’s opinion (3) is \textit{a priori} false—it is an ‘explanatory solecism’—because (3), when the clauses are read with their literal meanings, can never convey any empirical information (as I defined it above).

The implication is that the explanation we thought we had in (2) is drained of cognitive value, it ‘goes missing’. We assumed that such explanations are possible, but it follows on assuming the response-dependent thesis (T) that they are \textit{not} possible, so,

---

\(^9\)This is the stock example. \textit{Cf.} Johnston 1993, appendix 1. It is derived from Plato’s \textit{Euthyphro} and it makes best sense if we think of \( S \) as a god. \textit{Cf.} my chapter 4 for more on (T) and response-dependence.

\(^{10}\)There is some discussion about which substitution-principle is concerned: the argument doesn’t work if the principle requires substitution \textit{salva veritate} because there are cases where substituting \textit{a priori} equivalents into explanatory contexts produce falsehoods. So, Johnston acknowledges, the principle must be the weaker: it must not be possible to generate explanations, by substituting \textit{a priori} equivalents in explanatory contexts, which as an \textit{a priori} matter are false; that is, if we could generate explanatory solecisms by substituting, then the MEA can get started. See Johnston 1991, p. 127-131.
by *reductio*, (T) is false. This is, in brief, MEA as applied to (T) for the case of the pious.\(^{11}\)

In the context of UD MEA goes like this. UD claims that it is *a priori* that:

\[
(A) \, \phi \text{ means } p \text{ in } S's \text{ language} \iff S \text{ is disposed in } C \text{ to use } \phi \text{ to indicate all and only mental states with the content } p. \tag{12}
\]

But I have argued that there are possible empirical explanations that go like this:

\[
(B) \, S \text{ is disposed in } C \text{ to use } \phi \text{ to indicate all and only mental states with the content } p \text{ because } \phi \text{ means } p \text{ in } S's \text{ language.}
\]

That is, I wanted to be able to employ meanings to explain use of sentences. But now, by substitution of *a priori* equivalents in explanatory contexts, we get:

\[
(C) \, S \text{ is disposed in } C \text{ to use } \phi \text{ to indicate all and only mental states with the content } p \text{ because } S \text{ is disposed in } C \text{ to use } \phi \text{ to indicate all and only mental states with the content } p.
\]

And (C) appears to be one of Johnston's *a priori* false explanatory solecisms. The problem is that, if (C) is impermissible, then we must refute the assumption that lead to it, that is, we must refute UD.

*b) Critical Assessment of MEA.*

I believe the claim that (C) is an explanatory solecism is sustained by the more general principle of explanatory asymmetry, in this form:

\(^{11}\) Cf. Johnston *op.cit.*

\(^{12}\) Notice that here I have, for the sake of argument, suppressed the definitional claim 'ϕ means p in S's language = def S is disposed in C to use ϕ to indicate all and only mental states with the content p'. I do this because the issue is whether, and in what way, I am entitled to the definitional claim (the bare *a priori* biconditional is entailed by the definition).
(4) If A explains B, then A ≠ B.

In our case: if we want to say that meaning explains use, then we cannot also say things that entail that meaning = use. This shows that MEA is related to the “meaning is use” problem. Below I shall show how my solution to the “meaning is use” problem, which turns on disambiguation of ‘use’, works as a solution to MEA too. If (4) is what must be respected in order to avoid MEA, then my solution should straightoff be in the clear because it says:

(5) (meaning explains use\text{manifest}) & (meaning = use\text{possession}).

In order to show this it will be useful to have a closer look at MEA and its connection to response-dependence in general. My view is that we gain the right to explanations of use by meaning in virtue of endorsing response-dependence, not in spite of response-dependence.

i) the awesome force of MEA.

It should be clear at the outset that there is something wrong with MEA: it is simply too strong. Consider the heady response-dependent theorist who is happy to acknowledge, contra our normal intuitions, that there is no possibility of empirical explanations of the type:

(6) S judges that x is pious because x is pious,

The heady response-dependent theorist may agree that defending response-dependence is indeed so much of a threat to objectivist intuitions that (6) is just ruled out of contention from the beginning. But this kind of theorist will of course still want to admit the possibility of empirical explanations of another ilk, viz.:
(7) x is pious because S judges that x is pious.

That is, this kind of response-dependent theorist believes that some empirical information can be conveyed by the explanation in (7): there may be something about S, or the conditions under which S judges about piousness, which distinguishes the case from other possibilities (in the case of piousness, perhaps, the idea is that S must be a god, not a philosopher, and the gods must love x, not hate it). But this theorist’s own thesis is given by the *a priori*:

(8) x is pious $\equiv$ S judges that x is pious.

And by substitution of *a priori* equivalents in explanatory contexts we get:

(9) S judges that x is pious because S judges that x is pious,

which is an explanatory solecism. So, by *reductio*, a response-dependent theorist who believes in the possibility of empirical explanations, like (7), must refute his or her own thesis, *i.e.* (8). This should show us that something has gone wrong in MEA. Surely someone who believes in response-dependence should be allowed recourse to explanations like (7) for they are just a spelling out of how response-dependence works in particular cases. Without such explanations, the very idea of a distinction between response-dependence and response-independence begins to evaporate. Notice that we can run this style of argument for the concepts which according to Johnston ‘wear their response-dispositional character on their linguistic face—pleasing, shy-making, nauseating etc.’\(^{13}\) Surely we should be allowed to explain something’s being pleasing by citing facts about human responses, and yet, if MEA is right, we cannot even do that.

\(^{13}\)Johnston 1993, p. 118.
To see why MEA has gone wrong we need to discuss how it relates to response-
dependence, and the possibility of explanation, in general.

**ii) MEA, response-dependence and explanation.**

MEA demonstrates a tension between claiming that a particular concept is response-
dependent, and claiming that subjects’ responses can be explained by the appropriate
facts. The tension is so severe that one of the two claims must be rejected: either the
concept is not response-dependent, or the explanations ‘go missing’. But I think there
is something off-colour about the imagined tension: if we care so much about the
explanation that threatens to go missing, then why should we in the first place have
thought that the concept was response-dependent? My suggestion here is this: if the
commitment to the possibility of explanation is a commitment to explanation by robust
facts of the kind in question, then the concept in question should never have invited a
response-dependent approach in the first place; the concept just seems response-
independent. If, on the other hand, our commitment to the possibility of explanation is
not so clearly a commitment to explanation by robust response-independent facts (as it
isn’t in many of the interesting cases: colour, morals, comedy etc.), then we should
expect the possible explanations to reflect this. That is, the possible explanations
should reflect, precisely, that we are dealing with a response-dependent concept.
Therefore, if we suspect that a given concept is response-dependent we should also
expect that the status of the commitment to empirical explanation is up for discussion.
This is part of what it means to endorse response-dependence for a concept.

This implies that we cannot take a commitment to ‘the possibility of empirical
explanation’ at face value. There are different readings of it, pointing us in different
directions. Some readings will suggest response-dependence, some response-
independence. It is of course a good question to ask what guides these different
readings of the commitment. On the type of analysis I favour the strategy is clear: the
various platitudes and substantial commitments of a discourse are often interrelated, or networked.\textsuperscript{14} They cannot easily be accounted for in splendid isolation. The understanding of any particular commitment may be influenced by our understanding of any of the other ones (the "meaning is use" problem itself arose as a product of such interrelations). Hence, if the other platitudes strongly suggest that the concept in question is response-dependent, then the commitment to the possibility of empirical explanation should be read accordingly.\textsuperscript{15}

With respect to ME A the situation is this: ME A correctly shows that some kind of explanation goes missing when a concept is claimed to be response-dependent, \textit{viz.} the kind of explanation the possibility of which would demonstrate that the concept isn't response-dependent after all. But this shouldn't worry the response-dependent theorist because he or she should of course never be committed to the possibility of such explanations in the first place. But ME A shows this with such awesome force that we lose sight of the possibility of other kinds of empirical explanations, \textit{viz.} explanations which are \textit{sustained} by the response-dependent character of the concept in question. The task is then set for the response-dependent theorist: show, on the very basis of the \textit{a priori} biconditionals, that there are such alternative types of empirical explanation.\textsuperscript{16}

\textsuperscript{14}Cf. Ch. 2.
\textsuperscript{15}This leads to a commitment to careful revisions (but not rejection) of some platitudes, \textit{cf.} Ch. 2 §4.
\textsuperscript{16}This suggestion is in the same vein as Miller 1995, p. 362ff. Miller's point draws on works by Alston and C. Wright on ontological commitment whereas mine is motivated by general worries that ME A does more work than it should. I view this as a response to Johnston's insistence that ME A is a problem for response-dependence (about manifest qualities), Johnston 1998, p. 17-18. It might be that Johnston is right about redness, but that would be because redness should never have invited a response-dependent reading in the first place (so I, as it were, bite the ME A-bullet, but only to spit it out again).
iii) *explanations that go missing and explanations that become available.*

In my case this task is in fact easily performed. If we claim that the concept of meaning is response-dependent, then certain types of explanation do indeed go missing, viz.:

\[(B) \text{S is disposed in } C \text{ to use } \varphi \text{ to indicate all and only mental states with the content } p \text{ because } \varphi \text{ means } p \text{ in S's language.}\]

MEA shows that (B)-type explanations are not available. But this should be welcomed because (B) is consistent with regarding the concept of meaning as the concept of a response-independent property. I cannot simultaneously hold ‘meaning = dispositions for use’, and expect that use can be explained by use-independent meanings. But notice that by endorsing response-dependence, and losing the right to (B)-explanations, the possibility of another type of explanation becomes available. By saying that the concept of meaning is the concept of a dispositional property, I gain the right to explanations of manifestations of dispositions in terms of possession of dispositions, just as my solution to the “meaning is use” problem would have it.\(^\text{17}\)

This is what the *a priori* biconditionals divulge:

\[(A) \varphi \text{ means } p \text{ in S's language or dialect or idiolect } \equiv \text{S is disposed in } C \text{ to use } \varphi \text{ to indicate all and only mental states with the content } p.\]

And now MEA will not get a grip because we disambiguate on the notion of ‘use’ and thereby prevent the substitution principle from applying. Let me indicate this by underlining. There is ‘use’ in the manifestation sense:

*Manifestation:* S is disposed in C to use \(\varphi\) to indicate all and only mental states with the content \(p\);

\(^{17}\text{Notice that a non-dispositional response-dependence, } e.g., \text{ based on Wright's basic equations, will not immediately gain the right to this kind of explanation. Indeed, in so far as an account of meaning demands that facts about meaning display disposition-like theorecticity, it seems that basic equations are ruled out of contention from the start. Cf. Miller 1997b, §6.6.}\)
and there is 'use' in the sense of possession of dispositions:

_**Possession**: S is disposed in C to use \( \varphi \) to indicate all and only mental states with the content \( p \).

If the concept of meaning is response-dependent, and UD is its analysis, then I am committed to the possibility of empirical explanations of this type:

(D) S uses \( \varphi \) in the Manifestation sense because S Possesses a disposition D for use of \( \varphi \).

That is, via (A):

(E) S uses \( \varphi \) in the Manifestation sense because S means \( p \) by \( \varphi \).

This is not an explanatory solecism because one can rationally believe that the _explanandum_ is true without believing that the _explanans_ is true, and vice versa (one could believe that S uses 'echidnas eats ants' because S possesses some other disposition for use than the disposition to indicate mental states with the content _echidnas eats ants_, and one could believe that the disposition to indicate the content _echidnas eats ants_ is manifested in other ways, _e.g._ given other background beliefs and desires, or that it is never manifested). Hence the manifestation of a disposition and its possession are not _a priori_ equivalents, and therefore they cannot be substituted in explanatory contexts, and MEA cannot go ahead. The problem raised by MEA is solved because, by going response-dependent, we gain the right to present the _explanandum_ in manifestation mode, and thereby it is brought outside the scope of the substitution principle.
4. Conclusion

In this chapter I have tried to deal with the tension between the meaning is use platitude and the explanatory meaning platitude, that is, the "meaning is use" problem. The problem was that we seem to violate good explanatory principles if we say both that use explains meaning and that meaning explains use. I showed how this problem can be solved, on the basis of UD. The notion of 'use' can be disambiguated so that the problem doesn't arise. I went on to argue that this solution fits the phenomenological motivation for the explanatory meaning platitude. The core kind of explanatory meaning is precisely the kind which ensues once we disambiguate.

The "meaning is use" problem is produced by a tension between two of the platitudes surrounding the concept of meaning, together with assumptions about explanations. As such it arose before I explicated the concept of meaning as the concept of a response-dependent property. Johnston's Missing Explanation Argument is intended as an objection to response-dependence, but it raises a problem similar to the "meaning is use" problem. Therefore it was doubly important for me to defend UD against MEA: i) MEA is not a threat to response-dependence, and ii) my solution to the "meaning is use" problem works also on the assumption that UD represents a response-dependent property—indeed it is the very response-dependent character of UD which gains us the right to the explanations which help solve the problem and avoid MEA.

The price I must pay for this is comparably small: the explanatory meaning platitude must be revised such that (1) it doesn't require the possibility that use-independent meaning can explain use of sentences. This is simply ruled out once we acknowledge that the concept of meaning is response-dependent. This is of course fully in line with the result of Chapter 3 where I provided independent arguments against use-independent—primitive—meanings. And the explanatory meaning platitude must also
be revised such that (2) it doesn't require the possibility that meanings can explain possession of dispositions for use of sentences. Meanings, that is, semantic relations, are constituted by these dispositions, and cannot therefore explain them. This is the kind of explanation that does go missing, but this is as it should be because, if not, then we would have to acknowledge the possibility of use-independent meanings. This is not a bad consequence at all: we are still entitled to explain manifestation-use in terms of meaning, and moreover, we can still explain someone's possession of a disposition for use, *i.e.* someone's meaning something, in communal terms ("S has come to mean *sea-dragons are cute* by 'sea-dragons are cute' because S has become part of the linguistic community D"). So my claim is that, even though we have to revise the platitude about explanatory meaning, we retain the right to explain use in just the ways that matter to overall functionality of the concept of meaning.
In Chapter 2 I discerned three tensions among the platitudes. The second tension was between the meaning is use platitude and the error and robustness platitude. It came in two parts, one concerning robustness and one concerning error. The first part said that if meaning is use, then nothing prevents sentences from being used irregularly, and, if sentences are used irregularly, then we cannot have robust disagreement about meanings. The second part of the tension said that the meaning is use doctrine is inconsistent with the requirement that error about meaning be possible.

In this chapter I shall argue that these tensions can be resolved, given the doctrine that the concept of meaning is the concept of a use-dependent property. I shall begin by discussing the problem about robustness and irregular use, and then turn to the all-important issue about the possibility of error. In my attempt to do this, I align the dispositionalism in UD closely with our ordinary notion of dispositions, and, with that, the common (amended) counterfactual analysis of it.¹ My central idea is that if meaning-dispositions can be shown to do their job by doing what ordinary

¹As advertised in Ch. 4 §3.d.
dispositions (fragility, water-solubility etc.) uncontroversially do when they are not masked, mimicked or altered, then the dispositional account of meaning is just as cogent as our ordinary notion of dispositions.\footnote{Forbes (1984) has a proposal along these lines too, I find it congenial although he doesn’t talk about the counterexamples. I shall indicate as I go along where he and I concur. I also find Coates 1986 congenial. Especially because he allows short-lived, idiolectic use as meaningful. Ginet 1992 is interesting too, but he deals with the possibility of error, and of being disposed to make errors, differently from me and he settles, inexplicably it seems to me, for irreducible (primitivist sounding) dispositions with no categorical bases.} The various objections to a use-based account of meaning are then rejected either because they fail to recognise the dispositional character of UD, or, when they do recognise it, they are objections only against an un-sophisticated analysis of dispositions, specifically, an analysis that doesn’t accommodate masking, mimicking and altering.

1. Robust disagreement about error, and crass conventionalism.

In my discussion in Ch. 2 I gave the first part of the tension in this form:

(I) If meanings are robust in the sense required by the platitude about robustness, then meanings must somehow be regular; and meanings cannot be regular if they are determined by speakers’ whim [robustness].

(II) If meaning is use, then meanings are determined by speakers’ use; speakers’ use is determined by their whim [crass conventionalism].

(III) If meaning is use, then meanings aren’t regular, and therefore not robust.

The core of the motivation for the platitude about robustness is this: if meanings are not relatively robust—if they can change wildly all the time—then there will be no point in engaging in disagreements about meaning with a view to subsequent
correction. Why should we care about getting meanings right if they change all the time?  

The tension with the meaning is use platitude arose because, *prima facie*, saying that meaning is use could entail that meanings could lack robustness. The natural line of thought goes like this: nothing prevents me from using a sentence $\varphi$ now with this meaning, now with that, so if meaning is use a user's whim will determine which meaning a sentence is assigned. I call this the premise from 'crass conventionalism' because it makes a travesty of the common-sense notion that language is conventional. It presents language-users as being virtual semantic dictators: as if we were all—like Lewis Carroll's Humpty-Dumpty—authorised to change the meanings of our sentences without warning. I shall defend robustness by showing that UD isn't trumped by crass conventionalism. I will say more about crass conventionalism and then show that because UD is dispositional, it doesn't entail crass conventionalism.

*a) Crass conventionalism.*

Crass conventionalism is the view that meaning is conventional to such an extent that speakers aren't bound by previous use, or community use, when they use words. Whatever content they use a sentence to indicate goes into determining the sentence's conventional meaning. For example, you and I play a game in which you are supposed to guess what I mean by the sentence 'that is jolly spiff'. I point to different things, tell you about past events, and say of some of them that they are 'jolly spiff', and of some that they aren't. You are trying to guess the content which the use indicates, but there is no one and only content $p$ which my use of 'that is jolly spiff' indicates. Instead, there is the rule: I can freely change the contents which my use of 'that is jolly spiff' indicates. Of course, you have no chance of winning the game

---

3The aspect of robustness is different from the aspect about the possibility of error: error might in principle be possible even if meanings are not robust. The problem is just that lack of robustness makes it hard for us, and pointless, to track errors.
(unless perhaps you guess that I am trying to deceive you, but couldn’t I change what counts as ‘winning’ too?). Crass conventionalism is the view that language in general always has this nature.

The problem with crass conventionalism is not just that it entails that nothing I do is really wrong (as in the game). That problem is more directly concerned with the problem of error. But in the present case there could after all be a show of correctness conditions, even in the game of ‘spiff’. I might say “last week I said x is jolly spiff, but that was a mistake because it isn’t spiff”. A group of speakers may adopt a crassly conventional ‘spiff’-practice: saying that this-and-that is jolly spiff, and correcting each other on certain occasions.4 But such a show of correction and disagreement doesn’t alleviate the problem, of course, because it is itself based on crass conventions.

The problem with this is that robust disagreement about anything begins to seem impossible: if I can legitimately move the goalposts at my whim, then it is pointless to object like this: “but if you say that one is jolly spiff, but this one isn’t, then surely x couldn’t be jolly spiff”. If crass conventionalism is the case, then use is so irregular that I can never rule out that my interlocutors (perhaps including, as it were, my earlier self) constantly change their meanings, i.e. I can never exclude the possibility that our language or our idiolects have gone haywire. But then it will be pointless engaging in disagreements about meaning with a view to adjusting our uses of expressions. Moreover, from such a perspective, it doesn’t make sense to say that the way I use words is motivated by meaning. It follows, of course, that then meanings cannot explain use.

4Such a practice would be somewhat like Pettit’s notion of ‘U’ which I discuss briefly below in §2.d.
b) Dispositions are not necessarily irregular.

Once we see that the concept of meaning is the concept of a dispositional property, any worries about crass conventionalism should fall away. If the “meaning is use” slogan is given a dispositional reading, then it doesn’t entail the kind of irregular use that could give rise to crass conventionalism. The reason is simple: in general, if S is disposed to R in C, then some kind of counterfactual regularity holds true of S. This is what the counterfactual analysis begins to capture:

\[(D) \text{S is disposed to R in C } \equiv \text{ if S were in C, then S would R.}\]

For example, a cup has the dispositional property of being fragile iff were the cup struck in C it would break. The cup wouldn’t correctly be considered fragile if there is no regularity in the way it would behave when struck. If this is how it is with dispositions in general, then we may safely assume that this is how it is with dispositions to indicate contents of mental states. This ensures that the default case of the use-dependent story is one in which some regularity governs the use of a sentence \(\varphi\). If the scenario of crass conventionalism is meant to show that we could never expect regularity in the contents indicated by the use of a sentence, then the dispositionalist UD cannot be its target. If the scenario is meant to show, on a weaker note, that sometimes (as in the guessing game) we are disposed to use \(\varphi\) in an irregular manner (e.g. to indicate what we believe will throw the hearer off the scent), then UD is again out of trouble: we can imagine being regularly disposed to use sentences like ‘that is jolly spiff’ irregularly. So UD can accommodate phenomena of crass conventionality alongside our other normal dispositions to indicate contents of mental states in a regular manner.

---

5Modulo, as always, how the counterfactual analysis is amended. Here I register an accordance with Forbes 1984 because this is what motivates the overall endorsement of dispositions in his paper, cf. section IV. His account is not set in the context of robustness of meaning, but rather turns on the problem of error.

6Though of course someone may be regularly disposed to respond irregularly (as in the game of ‘spiff’).
We can conclude that use-dependence doesn’t entail crass conventionalism, (though it makes sense of the possibility of games of crass conventionalism). The dispositional nature of UD simply prevents irregular use and crass conventionalism. Pressing the point against UD amounts to pressing the point that dispositions as such (e.g., being fragile, water-soluble, and choleric) are not regular. But that is tantamount to saying that there are no dispositions at all, which is surely too strong in a debate about the concept of meaning.

c) Communication and dispositionalism.

I can imagine a staunch opponent to the use-theory of meaning insisting that, even though crass conventionalism isn’t entailed by UD, still we have not been given reason to believe that crass conventionalism doesn’t in fact prevail. That is, such an opponent would say that UD doesn’t give us reason to disbelieve that we always play games of crass conventionalism. However, with this move the opponent has shifted character completely. The opponent is no longer claiming:

(1) If UD is true, then crass conventionalism is true,

Rather, the opponent is simply claiming this:

(2) Crass conventionalism is true.

But (2) amounts to denying the very platitude about robustness in the first place. As such (2) does not bother me because I think the platitude about robustness is indispensable to our grasp of the concept of meaning. Indeed, not only does it fit the phenomenology of linguistic disagreement, learning etc., it also seems indispensable for the very possibility of communication. If our beliefs and desires were always so aberrant that we would always in effect play games of crass conventionalism, then it
would be hard to see how ordinary humdrum communication (e.g., telling you when the train is leaving, letting someone know that I am sad or happy etc.) could be possible. How could you know what ‘the train is leaving now’ means if you cannot safely assume anything about the meaning I attach to that sentence?7

This is a trivial point, and it should immediately silence the staunch opponent. In another respect, it is also a significant point. In my attempt to arrive at a formulation of UD, in chapter 4 (§4a), I mentioned that there would be some beliefs and desires that we should sideline as standard for all speakers in order to concentrate on the idea that sentences are used to indicate particular contents of mental states. These were beliefs concerning the conversational situation and our desires about how to participate in them. We can now see why it was right, and uncontroversial, to do that: because when communication is in fact going on, we have every reason to believe that those beliefs and desires standardly do not allow games of crass conventionalism to prevail. Games of crass conventionalism only ensue when those mental states are not standard, when we believe our interlocutor is pulling our leg, or when we desire to throw a nosy journalist off the scent. This is a significant point because it allows us to define part of the C-conditions for the use-dependent account of meaning: the meaning-dispositions must work in conditions where the speaker’s beliefs and desires about the conversational situation are such as to allow communication. This specification is a priori because it is made only on the basis of the a priori platitudes about robustness and communication.8

**d) Crass conventionalism and U-ness.**

Philip Pettit is concerned to defend a version of global response-dependence, has discussed a scenario like crass conventionalism too.9 He rightly says that this is a

---

7 Also, how could Humpty-Dumpty let Alice know that he is being crassly conventional, if crass conventionalism prevails?
8 I return to the specification of C-conditions in §4 below.
problem which any response-dependent account must face up to: doesn't response-dependence just entail that nothing constrains subjects' responses in whatever discourse is in question, and how could unconstrained responses even begin to capture the content of any of our concepts?

The example Pettit focuses on is a collusive community-based practice with the concept of U-ness (as used by the Sloane Square set). This practice is akin to what I have called crass conventionalism because the extension of 'U' is determined by the whim of the Sloane Square set (e.g., it is 'U' to 'go to the lavatory', 'non-U' to 'go to the toilet'). Pettit, however, focuses on a further aspect of the scenario, viz. that the crass conventionality of the practice with 'U' is a reflection of the fact that there is no relevant unified property in the world that 'U' seems to be predicated of. This is a serious problem for Pettit because he is also concerned to show that global response-dependence is consistent with a reasonable doctrine of realism. If all predicates are like 'U', then we have no reason to expect that our use of predicates ever tracks real properties. Leaving aside the debate about realism, I will sketch Pettit's solution to the problem about 'U'-ness and assess it in the light of UD.

Pettit discusses the tension between response-dependence and what I call crass conventionalism in the familiar terms of the Euthyphro-contrast. Consider attempts to test whether, in biconditionals like this:

\[(3) \text{ for all } x, \text{ } x \text{ is pious } \equiv \text{ the gods love } x,\]

we are entitled to explain the gods' love of x by adverting to x's being pious. Correspondingly for U-ness:

(4) for all x, x is U ≡ Sloanes judge x is U.

There is a clear intuition that, for the case of U-ness, we are not entitled to explain judgements in terms of a relevant unified property of U-ness. On the other hand, if our practices with the concept of redness manage to track a unified property of redness, then response-dependence had better sustain our entitlement to such explanations. This must be a prerequisite for the property to be robust (these considerations should be familiar from the above discussion of crass conventionalism and the previous chapter on explanatory meaning). Pettit argues that there is such an entitlement for redness in contrast to the case for U-ness. Once we see that

the only place for a systematic difference between [redness and U-ness] is in the things that [...] determine [the] responses, [then] we do indeed find a significant difference. U-responses are determined [...] by the efforts of the Sloanes in their efforts to keep in step with one another in their classification of things. But clearly, red sensations do not clearly spring from such collusive machinations [...]. When people see something as red [...] they do so [...] because [...] it is such as to merit the description 'red'—in short, its being red.\textsuperscript{11}

This is backed up by the argument that redness may count as a causally efficacious kind in virtue of possessing the higher level property relevant to red-sensations, which is itself realised by the relevant basic microphysical properties of the thing in question. The argument would then be that the concept of redness passes the test associated with (3) above \textit{because} there is some relevant property (or set of properties) which can appear in the purported explanations. In contrast, for U, there is no relevant basic level property.\textsuperscript{12}

\textsuperscript{11}Pettit 1991, p. 613.
Although I am generally sympathetic to this kind of approach, there is a *prima facie* problem for Pettit's suggestion. The problem is quite simple. Pettit claims that there is a difference between the concepts redness and U-ness in that redness passes the test that entitles it to a robustness-reading in (3) and U-ness doesn't. The reason U fails is, essentially, that what realises U-sensations doesn't form any relevant property.\(^{13}\) Now, as is fairly common knowledge, the reflectance-properties that go into producing colour sensations themselves form a very heterogenous set. The problem I want to raise is this: what, short of begging the question, should make us think that the difference between redness and U-ness is more than a mere difference in degree between concepts on the same heterogenous ontological footing? That is, why isn't U-ness a heterogenous property like redness, only more so?\(^{14}\) The worry in the background is that on Pettit's proposed test we let our pre-theoretic response-independent intuitions guide our judgements of which underlying properties are relevant and which aren't, and that begs the question in favour of the robustness of redness. (This is tantamount to violating the uniformity constraint I discussed in Ch. 4). If we do not beg the question like that, then there seems to be only a difference in degree of heterogeneousness between the properties.

I am not sure how serious this problem is for Pettit's purposes. My point is not that it is unassailable (indeed there might be a defence against it\(^{15}\)). My point is rather: i) to show that I couldn't just adopt Pettit's solution, and ii) to provide a contrast with my own solution (based on aligning the cases with ordinary dispositions) above. My proposed solution is much weaker than Pettit's: I just say, by analogy to other dispositional concepts, that dispositions must be regular. But that doesn't say anything about *which* contents of mental states dispositions for use of sentences.

---

\(^{13}\)See also the discussion of ping-pongness, Pettit 1991, p. 615-6.

\(^{14}\)Pettit does operate with a distinction between roles and realisers, where the roles are higher-order properties. This distinction doesn't help against my objection which is made in terms of realisers, because at role-level redness and U-ness are just different contentful concepts.

\(^{15}\)Building on, for example, the notion of programme causation, discussed in Jackson and Pettit 1988.
regularly manage to indicate. This helps us to pinpoint, for the case of UD, the type of problem Pettit is concerned with: does anything guarantee that there is *unity* in the contents regularly indicated by my use of a sentence like ‘fruit-bats are dangerous’, even given that we do not constantly engage in games of crass conventionalism? If this is a real problem, then my weak solution to the problem about crass conventionalism really seems too weak.

I think it *is* a real problem, and that my solution must be made stronger. Luckily, I think that it can be made stronger, just by strengthening the dispositional aspect of UD. In the next section I will generate the problem, discuss it and provide my solution to it. Since my approach is local, in contrast to Pettit’s global approach, I can avail myself of the particular analysis of the concept of meaning. This will make it easier for me to specify what it takes to have unity in the contents that are indicated by use of sentences, without having to draw on pre-theoretic potentially response-independent intuitions about meaning. That is, I hope to be able to specify what unity requires without violating the uniformity constraint I discussed in Ch. 4.


The problem I have in mind is simply that there is not enough *unity*, in the situations where S responds by tokening \( \varphi \), to warrant ascription to S of a disposition to indicate all and only mental states with the content \( p \).\(^{16}\) The argument is that each utterance of \( \varphi \) is associated with a certain set of background beliefs and desires, and that different utterances need not be associated with the same such set.

\(^{16}\)The objection was raised to me by Crispin Wright.
Thus this problem concerns our right to assume that when S manifests his or her disposition to use φ, then φ is used to indicate the same content of the underlying and potentially diverse sets of mental states. If the objection gives us reason to believe this isn’t the way it is, then UD is in trouble because the central contention of UD is that sentences are used to indicate all and only mental states with the content \( p \). This objection differs from the previous objections (in §1b-c) because those objections didn’t take sufficient notice of the fact that UD is a dispositional account. In contrast, this objection builds on the very notion that dispositions allow that different stimuli can occasion the same response.

\[ a) \textit{Generating the objection.} \]

Consider the set of all the diverse tokens of the sentence ‘this is poisonous’. For rational beings, the same mental states may underpin many different manifestations. Say I believe that a certain substance is poisonous. I may manifest this belief in ever so many ways: I may avoid eating it, I may eat small portions of it if I believe that doing so will immunise me to it, I may ensure my family doesn’t eat it, but I may also ensure that they do eat it, if my beliefs and desires were somewhat different, and so on.\(^\text{17}\)

Likewise, I may manifest different beliefs by similar means, that is, I may have different mental states underpinning tokens of the same sentence. Thus, for example, I may produce a token of ‘this is poisonous’ when I believe that something is poisonous and I desire to warn someone; or I may be in the pub telling a lame joke, so I produce a token of ‘this is poisonous’ believing that some substance isn’t poisonous, but desiring to pull someone’s leg; or I may produce a token of ‘this is poisonous’ believing that some substance is poisonous, believing that my interlocutor

\(^\text{17}\)This is a development of an argument in Evans 1985, p. 336f, directed against a notion of implicit knowledge. The argument has also been discussed by Wright, \textit{cf.} 1986 §I. Its ancestor must be Geach’s (1957, p. 8) argument against behaviourism. A version which comes close to the version I generate can be found in Miller 1997b, §§7.4-7.5. I discuss Miller’s version briefly below.
is a daredevil and desiring him or her to take the poison, and so on. There are also many more wayward examples: I produce a token of ‘this is poisonous’ believing that someone wrote me a nasty letter and desiring to let someone know.

The result is that the beliefs and desires underlying my various tokens of ‘this is poisonous’ may have very little in common—there may be very little unity. Different beliefs and desires underlie different tokenings of the sentence. In the normal run of things, my entire belief state and set of desires change all the time. I constantly acquire and lose different beliefs and desires. Due to the holistic nature of belief-desire explanation, any such change could in principle influence the state underlying a particular token of ‘this is poisonous’. Hence, there is potentially endless diversity in the states underlying my tokens. If this is right, then UD seems to be undermined: the objection shows that \( \varphi \) can mean ‘this is poisonous’ even though its use doesn’t, as UD requires, indicate all and only beliefs and desires about, say, poisonous things.

For my purposes I shall treat this as a threat to robustness. That is, it is a threat which says:

\[(1) \text{ If UD is true, then there is no unity.}\]

This is a threat to robustness because if there is no unity in the contents of the mental states indicated by use of \( \varphi \), then there will be no robust disagreements about meaning either: there will be no unity in that which we appear to disagree about. What I have to show, then, is that we have reason not to believe (1).\(^\text{18}\)

\(^{18}\)Johnston 1992 discusses a problem for the core belief that there is ‘unity’ in colours.
c) Assessing the problem about unity.

There is in fact some tension in the objection as I have presented it. On the one hand, it has the form of (1) above: since UD is dispositional, there is no unity. But, on the other hand, part of the motivation for the objection just turns on the holistic nature of belief-desire explanations, i.e.:

(2) if belief-desire holism is true, then there is no unity.

I think we should be suspicious about the status of (2). If (2) is right then any theory of meaning it is combined with would be in grave trouble. It is true that whenever I interpret someone saying, e.g., "there is a marsupial mole" as meaning there is a marsupial mole, then it is a possibility that I got it wrong: the speaker's beliefs and desires may be different from what I expected them to be. But in general we can interpret other people: we do in fact manage to communicate, so there must be fairly fixed patterns of beliefs and desires underlying our use of words. This is, I think, just common sense. This should tell us that belief-desire holism is no threat to the unity of the contents of the mental states that we use sentences to indicate—(2) is just wrong. From this common sense point of view, the objection about unity just boils down to the question of how in fact we manage to identify the unified patterns of use we know are there in most cases.

The objection stemming from (2) does of course say that there is not total unity in the mental states underlying our uses of sentences: sometimes our interpretations do in fact fail. But this doesn’t entail that there is no unity. All that is needed is that there is enough unity to make sense of the fact that we do communicate successfully most of the time, and (2) doesn’t threat that idea at all. Indeed, the objection is generated by trading on this fact: the very phenomena of telling jokes and daring daredevils presuppose that there is ordinarily enough unity in the beliefs and desires underlying our uses of sentences.
This provides the backdrop for the objection stemming from (1). To refute (1), I must show that the dispositionality of UD doesn’t aggravate the situation by creating too much disunity, and thus undermine the common sense view about communication. Importantly, I must show this in a way that leaves conceptual space for the particular problematic phenomena like joking and daring.

d) An answer to the problem about unity: joking and daring.

I generated the problem about unity (in the form of (1)) by adverting to cases like joking and daring. These are illustrations of the general point that we may produce tokens of the same sentence in response to different input. Notice first that this general point must be limited in its force. It cannot be so strong that we lose the right to talk about dispositions at all. That is, the point cannot be that there is no such thing as fragility (i.e. being disposed to break when struck), because we can tell stories about how objects can break or fail to break in all sorts of circumstances. If the problem about unity takes this general form then it is too strong (given we agree, as I think we should, that things can be fragile, water-soluble or choleric etc.)

This shows the true complexion of the problem about unity: something about the particular cases (e.g., joking and daring) raises the spectre of disunity, but this cannot be generalised without threatening the very possibility of dispositional properties. As I shall simply assume that dispositional properties are possible, I must show, on the basis of the dispositionalist UD, that the particular cases that generate the objection don’t produce too much disunity.

I believe there is a good dispositionalist story about what happens in the problematic cases of joking and daring. We already know of some types of apparently problematic
cases for the ordinary notion of dispositions, \textit{viz.} masking and mimicking.\textsuperscript{19} I believe they provide the key to the solution to the problem about unity. In the case of joking the normally unified mental states are \textit{masked}, and in the case of daring the normally unified mental states are \textit{mimicked}. Just as the standard counterexamples of masking and mimicking should not lead us to reject the very idea of dispositions, so the cases that give rise to the problem about unity shouldn’t lead us to reject the idea that meaning is dispositional.

In the case of joking, it is true of me that:

(3) I am disposed to use ‘this is poisonous’ to indicate all and only mental states with the content \textit{this is poisonous}.

However, when I jokingly say ‘this is poisonous’ I do not have mental states with that content, instead I have mental states with the content \textit{this is not poisonous}. This is what throws doubt on the status of the disposition in (3) and raises the worry about disunity. But the right response to this is not to reject (3), rather it is to say that in the case where I joke I am indeed disposed as in (3), but the disposition is masked. Something, \textit{i.e.} non-standard beliefs and desires, prevents its ordinary manifestation.

In the case of daring, (3) is likewise true of me, but when I dare a daredevil to eat some beef-jerky I believe to be poisonous by saying ‘this is poisonous’, the mental states I have are not what we ordinarily would suppose them to be to have the content \textit{this is poisonous} (in particular, I have the non-standard desire that someone eat poisonous beef-jerky).\textsuperscript{20} This again throws doubt on the status of the disposition in (3) and raises a worry about disunity. But as before it should not lead us to reject the

\textsuperscript{19}I discussed masking and mimicking as counterexamples to the simple counterfactual analysis of dispositions in Ch. 4 §3.d.i.

\textsuperscript{20}The thought is that the belief that something is poisonous normally goes with the desire that no-one should eat it. The story could be told more morbidly as a suicide case.
whole dispositional machinery, rather it should lead us to say that I am indeed disposed as (3) says, but that the disposition is mimicked in the case where I dare someone. Something, \textit{i.e.} a non-standard desire, prevents the ordinary manifestation of the disposition to use ‘this is poisonous’.\textsuperscript{21} (If the problem about unity is more severe, \textit{i.e.} if it says that manifestations and dispositions can come radically apart, then mimicking also seems to be relevant. In that case I may not have the disposition in (3) at all but may still have manifestations as if I were so disposed).

The general approach goes like this: for every particular case that suggests disunity in the beliefs and desires that underpin use of a sentence, there may be a story about masking or mimicking which saves the dispositional character of the account of meaning. This gives us the following general strategy for dealing with the problem about unity: formulate UD such that it is clear that the dispositions it utilises are neither masked or mimicked. This will ensure the default case is unity in the mental states indicated by use of φ.\textsuperscript{22}

I think that this is a reasonable response to the objection—especially when we focus on it in the form of (1) rather than (2). Of course, there might be cases that suggest disunity which cannot easily be fitted in as masking or mimicking. This may be so, for example, for the case where I say ‘this is poisonous’ about a nasty letter someone wrote to me. This raises the general question: what about disunity beyond masking and mimicking? Here there is a common sense answer: disunity beyond masking and mimicking indicates \textit{different} meanings. When I say ‘this is poisonous’ about a letter, what I say has a metaphorical meaning, that is, it indicates mental states with contents about the sender’s malicious intentions. In contrast, when I say ‘this is poisonous’

\textsuperscript{21}This is a special case of what Johnston calls mimicking: normally mimicking is where the disposition is not there but the corresponding counterfactual is true of x. It seems to me mimicking also occurs when the counterfactual is true but is not explained by possession of the disposition. Cf. Johnston 1993, appendix 2.

\textsuperscript{22}I shall find room for the type of counterexample called ‘altering’ below in §4.
about a piece of dried meat, it indicates mental states with contents about how this meat is harmful to the physical health of humans. My suggestion, then, is that such cases don’t really suggest disunity, rather they suggest a dispositional account of how similar sentences may have different meanings. This is surely a common sense and platitudinous explanation of cases like the poison-pen letter. To deny this is to hold that ordinary language doesn’t have ambiguous sentences.\textsuperscript{23}

There might be further cases that suggest disunity, but which seem neither masked or mimicked, nor seem to be plausible cases of different meanings. Such a case might be Pettit’s ‘U’-example. My suggestion is that these cases have \textit{less unified} meanings than other cases. This is how it is with ‘U’ (and perhaps other slang expressions like ‘awesome’, ‘cool’, ‘wicked’ etc.). Their meanings are presumably designed primarily to work more as social emblems than as means of swift communication. Such meanings may be called ‘conglomerate’ meanings because they don’t have a high degree of unity.\textsuperscript{24}

To sum up: My claim is that UD can avoid the problem about unity because it is dispositional. That is, it can avoid it if we assume that dispositions in general have an analysis such that masking and mimicking of dispositions is possible. Beyond that we have to bite the bullet: not all the potential cases of disunity can be explained away in terms of masking and mimicking. But common sense tells us that those cases are not

\textsuperscript{23}There is a caveat to this: I do not intend it as a general theory of metaphors. I think it would be possible to give dispositional accounts of different instances of metaphors that would bring out the differences, for example, between dead and live metaphors. It does, however, look as if it is on collision course with Davidson’s position in 1984, Essay 17.

\textsuperscript{24}These remarks are consistent with retaining the basic intuition that belief-desire explanations are holistic. Compare first with the fragile cup which may be struck in ever so many ways (with this or that force, repeatedly, with this or that instrument etc.); the variety of ways it may be struck doesn’t take anything away from the ascription to it of the dispositional property of fragility (as long as we have a way of dealing with masking, mimicking and altering). Likewise with dispositions for use of sentences: there may be ever so many different belief-desire sets underlying S’s use of ‘\textit{p}’ but this shouldn’t impede our ascription to S of the disposition to use ‘\textit{p}’ to indicate mental states with the content \(p\), as long as those sets aren’t so diverse that the disposition is masked or mimicked (or altered). (See also Ch. 8 §5.a. where I discuss this problem in relation with Horwich’s use-theory of meaning).
serious: they are either cases of ambiguous sentences, or cases of sentences with conglomerate meanings. UD then comes out like this. The \textit{a priori} analysis of the semantic relation expressed by \textquotequotesingle \( \varphi \) means \( p \)\textquotequotesingle implies \textit{a priori} definitional biconditionals such that, where \textquotequotesingle \( \varphi \)\textquotequotesingle is a declarative sentence type and \( p \) is a proposition:

\textit{UD v. 3:}
\begin{itemize}
  \item \( \varphi \) means \( p \) in \( S \)'s language \( \equiv_{\text{def}} \) \( S \) is disposed in \( C \) to use \( \varphi \) to indicate all and only mental states with the content \( p \).
\end{itemize}

Where we are to remember that the notion of dispositions has an analysis—\textit{e.g.} of the Lewis type (cf. Ch. 4§3.d.i.)—so that masking and mimicking (and altering) come out as possible, but without working as counterexamples to the analysis.

e) \textit{Substantiality and uniformity in the specification of the C-conditions.}

If this suggestion is on the right track, then we also get the beginnings of a solution to another serious problem for the dispositional analysis of meaning. Alex Miller has discussed Boghossian's criticism of dispositionalism and proposed that the problem Boghossian sets up can be divided into two parts, one part concerning circularity and the other concerning infinity. Boghossian's problem turns on the trouble the dispositionalist has with formulating the C-conditions for dispositions in substantial and uniform ways, Miller's suggestion is that this is either a problem about being able to formulate the C-conditions for meaning-dispositions without presupposing a grasp of the concept of meaning, or it is a problem about explicitly excluding an infinity of possible beliefs and desires which could, in our terms, create disunity in meanings. The present suggestion deals with this latter problem (I shall return to the circularity problem towards the end of this chapter).

\textsuperscript{25}Miller 1997b, Ch. 6, cf. also Miller 1997; Boghossian's arguments are in his 1989.
The problem about infinity is this: the problem about unity tells us that there is a potential infinity of beliefs and desires that may create disunity in the contents indicated by use of a sentence. It is incumbent on the dispositionalist to specify the C-conditions in a substantial and uniform way. But it is impossible to state an infinity of beliefs and desires in a finite analysis, so, *a fortiori*, the dispositionalist cannot succeed. However, there seems to be no reason to require that the C-conditions be *specified* in such detail for a dispositionalist account to get off the ground. Consider again the fragile cup: we can say what it is for it to be fragile, *viz.* that it be disposed to break when struck in C, barring masking and mimicking (as well as altering) of the disposition. We don’t lose the right to talk about fragility just because we don’t say anything about all the many different ways in which masking and mimicking (and altering) can take place (many different kinds of angels and demons can interfere in the causal order of things). We don’t have to say that the cup must not be masked in this way, or in that way, or... . It is enough that we can say something substantial about what in general it takes for the disposition to be masked, mimicked, or altered (*e.g.* that its intrinsic, or natural, properties are being interfered with such that its characteristic manifestation fails). Hence, if we do assume that the notion of dispositions has a Lewis-style analysis which explains the possibility of masking and mimicking (and altering), then we don’t need to specify all the ways in which masking or mimicking (or altering) can take place. The same goes, *mutatis mutandis*, for the case of the dispositions in UD.26

This is how I envisage a response to Miller’s second part of the Boghossian problem for dispositionalism. I shall return to it again towards the end of the paper. There I shall also discuss the part of the problem concerning circularity. I think this problem also has a solution, given the Ramsey-style analysis I favour (and Miller concurs on this point).

---

26This kind of solution is along the lines of the functionalist account espoused by Pettit forthcoming. He does not however connect it to masking, mimicking and altering of dispositions.
CHAPTER VI, THE SECOND TENSION: ERROR AND ROBUSTNESS

f) Two further versions of the problem from belief-desire holism.

The objection generated by considerations about the holism of belief-desire explanation can in fact be evaluated in two further ways. I will briefly explain these and discuss their relevance to UD.

i) the objection as a problem about compositionality.

Miller (1997b) employs a version of the objection stemming from belief-desire holism to demonstrate the failure of Gricean analyses of meaning.27 He observes first that the Gricean must begin her analysis from sentence-meaning (i.e. whole sentences used with this or that Gricean intention) The main reason for this is that there doesn’t seem to be a good Gricean model of how subsentential expressions express beliefs at all (and even if they do, it would presumably be beliefs about the semantic properties of linguistic expressions (e.g. the belief that the word ‘cow’ has the conventional meaning cow)).28 Miller then, in a discussion of Blackburn, goes on to show that interpretation cannot begin with sentence-meaning, that is, that interpretation must presuppose linguistic structure.29

This last argument rests on the same kind of case as gave rise to the problem about unity. Imagine a radical interpreter trying to interpret a native who utters ‘Lo, monadi kel guro’ whilst pointing at a nearby mountain. There is, as we have seen, an indefinite number of mental states that could underlie this utterance, so how can the interpreter begin to interpret the native speaker? According to Miller, the method for narrowing down the number of possible interpretations proceeds by the assumption that the subsentential parts of the utterance (e.g., the so-called predicate ‘...guro’)

---

27 Miller 1997b, §§7.4-7.5.
28 Here Miller builds on Platts 1979; Evans 1985; Wright 1993a.
29 The relevant passage is Blackburn 1984, p. 128.
make systematic contributions to the meanings of the sentences in which they figure. That is, successful interpretation presupposes the compositionality of language.\(^{30}\)

Miller uses these arguments to conclude that the prospects for the Gricean account seem bleak. On the one hand, it seems the Gricean must make sentence-meaning prior, but on the other hand, it seems that he or she cannot make sentence-meaning prior. This objection may be relevant for UD, in so far as UD resembles broadly Gricean accounts. But I think the objection is harmless: Miller’s conclusion concerning the prospects for the Gricean project seems to be a \textit{non sequitur}, for the following reason.

The fact that we must assume compositionality in order to \textit{know} which interpretation is right doesn’t show that the correct \textit{analysis} of meaning must be compositional. Sentence meaning can be \textit{analytically} prior to subsentential meaning, even though subsentential meaning is \textit{epistemologically} prior to sentence meaning. If indeed something like UD is right, then we can be sure that the problem from belief-desire holism doesn’t show that there is no one correct interpretation of an utterance (as my answer to the problem from unity showed). It is then a \textit{further} question by which method we get to know that correct interpretation. But it would clearly be a \textit{non sequitur} to say that the method by which we get to know the correct interpretation is, at the same time, what constitutes the fact that it is the correct interpretation. (It would be like saying that we have methods for getting to know which mountain is the tallest, and then adding that the method by which we get to know this constitutes which mountain is tallest).\(^{31}\)

\(^{30}\)Miller cites Fricker 1982. Incidentally, I do not think Miller’s and Fricker’s arguments are entirely successful against Blackburn. Surely Blackburn is right that we can begin to do some interpretation of a language, however limited, without having any idea of its articulated semantic structure.

\(^{31}\)Miller in fact uses just this type of reasoning to rebut an earlier objection to the Gricean project, cf. 1997b, p. 234.
ii) the objection as a problem about first-person authority.

The objection about unity can in fact be given another, epistemological slant. I have assessed it in terms of our success at interpreting others. But what about the first-person case? How do we ourselves keep track of the mental states indicated by use of words? From this perspective the objection goes like this: it may well be that we do have dispositions, and that they can be masked or mimicked in various ways. But how could a speaker know whether a disposition is allowed its characteristic manifestation, or whether is it masked or mimicked (or whether a sentence is ambiguous or a conglomerate, for that matter)? The problem is, the objection goes, that dispositionalism is inherently behaviourist: if dispositionalism is true, then we only have epistemic access to our verbal behaviour and on that evidential basis we cannot say anything about the proper manifestation of dispositions. When I say ‘feral goats are tasty’, then that is nothing but a piece of verbal behaviour and as such it may be underpinned by an infinity of different sets of mental states which I don’t know about.

This changes the complexion of the problem from a worry about dispositions to a worry about knowledge of meaning, given dispositionalism. To answer this problem I must explain why dispositionalism isn’t a type of behaviourism. I attempt that in Ch. 7. It is an interesting objection, though, because it ties the problem of robustness in with the third tension among the platitudes: the tension about behaviourist use. My answer will be that the particular epistemology associated with privileged first-person authority about meaning sits particularly well with the dispositionalist character of UD.
3. Dispositionalism, error, and the problem of infallible use.

I turn now to what seems to be the core problem for a dispositionalist account of meaning, viz. the problem of error. This problem shot to prominence with Kripke's (1982) treatment of Wittgenstein's rule-following considerations.32 The core of the problem is really rather simple: a dispositionalist seemingly equates performance, or use, with correctness, and therefore cannot make sense of the possibility of incorrect performance, or incorrect use. In my discussion of the platitudes surrounding the concept expressed by the relation 'φ means p' (in Ch. 2), I set out the problem about the possibility of error like this:

(I) If meaning = use, then ψ has a meaning.

(II) If ψ has a meaning, then it must be possible to make errors about the meaning of ψ [error platitude].

(III) If meaning = use, then even erroneous use = meaning, and then it is not possible to make errors about the meaning of ψ [use=correctness].

(IV) It is possible and not possible to make errors about meaning.

Assume that an utterance of 'echidnas are spiky' is an error. Also assume that meaning is dispositional use and thereby, seemingly, that use equates correctness. We then we get a contradiction because it follows that the utterance cannot be an error after all. In the light of the contradiction, something must go: either the assumption that error is possible, or the assumption that accounts of the UD-type entail the equation of use with correctness. Obviously, I think the latter assumption is false. UD does not entail the equation of use with correctness. In this section I shall state why I think this is the case. I will argue that error is what happens in non-C-conditions. I shall then

defend this claim against three versions of the problem about infallible use. In §4 where I discuss the notion of C-conditions in more detail, I provide the final defence of my claim.

My argument in this section is prefaced by this simple idea. We know that there are concepts of dispositional properties, e.g. fragility, water-solubility, cholerism etc., so I shall compare the meaning-constituting dispositions with these ordinary dispositions. We have a right to talk about dispositions as such, no matter what their final analysis is going to be. My claim is that objections to meaning-dispositionalism cannot be so strong that they are also objections to the ordinary notions of dispositions. The objections must somehow be specific to the debate about meaning. As it turns out, no version of the objection that use equates correctness can be contained so that it only concerns the concept of meaning. This provides me with an intuitively strong defence of dispositionalism about meaning: it is just as strong as dispositionalism about paradigmatic dispositional properties such as fragility and water-solubility and cholerics.

However, by aligning UD with ordinary dispositions a new threat arises: the differences among dispositional concepts become less visible. We analyse the concept of meaning because we want to know what is special about that concept, not just that it shares something with concepts of other dispositional properties. In §4, I overcome this difficulty by giving a more detailed account of the C-conditions of meaning-constituting dispositions.

a) Error and the importance of error.

The standard objection to dispositional accounts of meaning is that they make error impossible. What type of error is this, and why is it important that error be
possible.\footnote{For further detail about these questions cf. Ch. 2 \S2.d.} There are two ways in which we can be in error: either there is falsity in what is \textit{believed}, or in what is \textit{said}. The distinction between what is believed and what is said implies that falsity in what is said can be explained in at least two ways: either as the result of holding false beliefs but getting the meanings right, or as a result of getting the meanings wrong while having false or true beliefs. It is the latter case which is relevant for the criticism of UD because UD only concerns what is said, not what is believed. This implies that the problem about the possibility of error, with respect to UD, only concerns the possibility of falsity in what is said, which is produced by getting the meanings wrong. Falsity in what is said, which is a result of having false beliefs, concerns a problem beyond the scope of this essay.

If the problem about the possibility of error shows that UD entails that there can never be falsity in what we say, then UD must be wrong. We could not use language for communication as we do, unless it is possible to say something false: imparting information to someone implies that certain states of affairs are ruled out as not true. Phenomena such as statements of self-deception and slips of the tongue would also be impossible if we could never say anything false. This is a very serious problem for UD, even though it would be consistent with this to hold that we could make mistakes in what we believe (we could then have true or false beliefs which we could not transmit in communication).

\textit{b) Dispositions and errors.}

UD is the following dispositional account of the concept of meaning: the \textit{a priori} analysis of the semantic relation expressed by \( \text{"} \varphi \text{"} \) means \( p \) implies \textit{a priori} definitional biconditionals such that, where \( \text{"} \varphi \text{"} \) is a declarative sentence type and \( p \) is a proposition:
UD v. 3:
φ means \( p \) in S’s language if and only if S is disposed in C to use φ to indicate all and only mental states with the content \( p \).

Where we are to remember that the notion of dispositions has an analysis—e.g. of the Lewis type (cf. Ch. 4§3.d.i.)—so that masking and mimicking (and altering) come out as possible, but without working as counterexamples to the analysis. According to UD, then:

1. \( φ \)’s meaning, i.e. correct use, is given by S’s disposition for use of \( φ \) in C, which makes space for this idea:

2. an error with respect to \( φ \) is when S uses \( φ \) in non-C-conditions.

For example, \( φ \) is the sentence type ‘there is a kangaroo’ and S is disposed in C to use ‘there is a kangaroo’ to indicate all and only mental states with the content there is a kangaroo. An error with respect to ‘there is a kangaroo’ would be when non-C-conditions obtain, and S utters ‘there is a kangaroo’.34 This seems intuitively right because non-C-conditions include situations where there is no mental state with the content there is a kangaroo (e.g., masking), or, where there is such a state, but something else is going wrong (e.g., mimicking).

Hence, my claim is that UD does not entail the impossibility of error. There is a clear intuitive way in which UD allows error, viz. use in non-C-conditions. So the problem about the possibility of error doesn’t go through and therefore no contradiction.

34The idea that error is what happens in non-C conditions has been canvassed and criticised by Boghossian 1989, p. 529. The proposal is developed and defended in Pettit 1989, 1993, and forthcoming where it lays the foundation for a global response-dependence. I find Pettit’s proposal very congenial, but in contrast to him I stick much closer to a pure dispositionalist vocabulary, and I only intend this to be local account of the concept of the relation expressed by ‘\( φ \) means \( p \)’. Forbes 1984, p. 228, endorses the idea too, in a more narrow dispositionalist clothing.
ensues. I will now defend this claim against three alternative ways of putting the point that use equates correctness. None of these alternatives amounts to a serious problem for UD.

c) First alternative version of the problem of error: an objection about infallible use.

I believe the problem about the possibility of error can be made to seem stronger than it is, when it is focused only on the part of dispositionalism that concerns performance, or use. On this conception of the problem, the criticism of dispositionalism amounts to:

(3) if S uses φ to indicate the content p, then S is correct to do so,

i.e., according to the opponent to UD, whatever S does is correct. This would indeed lead to serious trouble for UD. But it blatantly overlooks the core of a dispositionalist account like UD:

(4) if S is disposed to use φ to indicate the content p, then (if S uses φ to indicate the content p, then S is correct to do so).

That is, UD has the form of a hypothetical statement: it is only given S's disposition for use of φ that S is correct to use φ in a certain way. The simple version of the problem about error in (3) simply overlooks this and states UD as a categorical statement of the antecedent of (4). The effect is that UD is criticised for entailing something it just doesn't entail: viz. that use equates correctness. What UD does entail is that use equates correctness, in so far as S is disposed in C in a certain way.

This emphasises the point that, even before we get into a discussion of the C-conditions etc., the opponent to dispositionalism must come to terms with, as it were, the idea the dispositionalism is dispositional. Without acknowledging this the
objection is just not an objection to UD. But once this is acknowledged, the task seems much harder for the opponent, who must then come up with grounds for premise (III) in the argument, which takes the dispositional status of UD into full account. This has the important consequence that the opponent's arguments cannot be arguments against the very notion of dispositions without begging the question. The arguments must be specific to dispositionalism in the debate about the concept of meaning.

d) Second alternative version of the problem of error: guaranteed correctness.

I can imagine someone objecting like this: even if UD is right, it still guarantees that we are never wrong. On this version of the objection that use equates correctness, the idea is that being disposed to do something somehow entails that whatever one does is correct. Thus:

\[(5) \text{ even on the assumption that UD is dispositional, it follows that error is impossible.}\]

Perhaps we can hear the objection as a Wittgensteinian objection: dispositionalism entails that whatever we do is correct, and that therefore we can no longer talk about correctness and incorrectness.\(^{35}\) Dispositionalism, as it were, places the subject beyond blame.

i) there is no problem about guaranteed correctness.

But this again glosses over the detail of the dispositional account. Remember that the claim is that dispositionalism is entitled to a distinction between correct and incorrect

\(^{35}\)This would be as a paraphrase of the last part of Wittgenstein 1953, §258. Related debates can be seen in Wright 1986b, although he conducts it in non-dispositional terms; and, perhaps implicitly in Davidson's argument that to be interpretable we must be viewed as not actually always uttering falsehoods, cf. Davidson 1984, Essays 9-11.
use, *viz.* in terms of use in C-conditions and use in non-C-conditions. The Wittgensteinian objection does nothing to show UD isn’t entitled to this distinction.

Here is what dispositionalism *does* entail: it is not possible that we would always be wrong. That is, if S is disposed in a certain way, then, necessarily, there is a possible occasion where S gets it right. This is what it means to say that something is disposed in a certain way. Compare again with the fragile cup. A cup is fragile if there are some possible circumstances such that, if the cup is struck in them, then it breaks.

But nothing detrimental follows from this. Firstly, it doesn’t follow that S will get it right on *all* possible occasions. From the fact that there are some possible occasions on which S gets it right, it doesn’t follow that S gets it right on all possible occasions. The C-conditions might not obtain on some occasions, or S may simply not have the prerequisite mental states (just as there are many possible occasions on which the cup doesn’t break: it is never struck, or the C-conditions never obtain, or are masked, mimicked or altered). Still less does it follow that S gets it right on *all actual* occasions. In fact, S can be disposed in a certain way without *ever actually* getting it right (compare: the C-conditions may happen to conspire against the dissolution of a water-soluble sugar-cube every time it is actually put in water).36

Hence, it is just wrong to suppose that dispositionalism entails that correctness is guaranteed. Dispositionalism is fully consistent with basing the notion of error on the distinction between what happens in C-conditions and what happens in non-C-conditions.

36In a recent paper Johnston makes a similar remark (though in a discussion of response-dependence and so-called ‘manifest’ qualities), *cf.* 1998, p. 18.
ii) the flip-side of the problem: meaning based on incorrect use.

We might think, however, that this saddles us with another problem. If dispositionalism is consistent with:

(6) S may never actually get it right,

then we have admitted that S may mean there is a waratah by 'there is a waratah', even though S never actually uses that sentence correctly. Is this a credible account of meaning-dispositions, let alone of dispositions as such? Simon Blackburn has discussed a case which suggests it isn't.37 Sometimes it seems we judge about the possession of dispositions on the grounds of what would actually happen in C-conditions. Consider, for example, the person who has all the characteristics of someone who is a good rockclimber, but who gets sweaty hands when there are rocks nearby. This could be a case of altering or masking of the dispositions to be a good rockclimber, but Blackburn convincingly invites us to judge that the subject was never disposed to be a good rockclimber in the first place, because for rockclimbing it is what happens when there are rocks nearby that matters.

Though Blackburn's example has intuitive force I don't think it should be generalised. Consider again the rockclimber who has all the characteristics of a good rockclimber. But assume that the proximity of rocks doesn't make the hands sweaty, but that the gung-ho macho fraternity of rockclimbers makes the subject freak out and fall off every time he or she gives it a go. It would be unfair, I think, to judge that this subject isn't disposed to be a good rockclimber—here the circumstances really do conspire against the subject to make him or her perform badly. So it seems there are cases that can go both ways.

37Blackburn 1993, §2.
How do meaning-constituting dispositions fit into this? On the one hand, I would like to accommodate the intuition that for some dispositions it is what happens in C-conditions that matters to whether we can be said to possess a given disposition. On the other hand, I would also like it to be the case that I can be disposed to mean something by a sentence, even though I always get it wrong (e.g., I use the sentence 'there is a marsupial mole' twice in my life, and both times there is error in what I say yet that sentence still means there is a marsupial mole in my language), or even though I never actually get a chance to utter it (as with a very long convoluted sentence, for example).

I think there is an easy answer to this: we do after all very often communicate successfully, so we have every reason to believe that our world as a matter of contingent fact is one of those worlds where we get it right, a lot of the time, for a lot of our utterances.\footnote{However, though this may be true for a lot of our sentences it cannot be true for most: a natural language has an infinity of sentences and we never get round to use more than a tiny fraction of them. These comments concern the so-called problem of 'meaning without use', cf. e.g., Lewis 1992. The problem is that it is hard to see how meaning is use, when we never get round to use the vast majority of the infinite number of sentences in any natural language. My comments here mark the beginnings of an answer to this problem (but only the beginnings, I'm afraid): there is in general nothing wrong with assuming that we can be disposed to do something even though we never get round to manifest that disposition. Similarly with use-dispositions. So we should not be surprised that there are some sentences that have a meaning in our language—because we are disposed to use them in a certain way—but which we never get round to use. From this perspective, what is special about the 'meaning without use' problem is that it concerns an infinity of sentences. This tells us something about our functional organisation: we are disposed such that our language has an infinity of meaningful sentences. This is a piece of armchair metaphysics, of course, but it doesn't tell us anything about how in particular we must be disposed.} So we can pay lip-service to the intuition that it is what happens in C-conditions that matters for what we mean: we have every reason to believe that, when the C-conditions obtain, then we get it right; but there is no necessity about it.\footnote{This doesn't imply that the possibility of communication is incidental to the concept of meaning. It is not as if we, as a happy coincidence, happen to latch on to a primitivist semantics. It is still true that we speak the language we speak because we are disposed in a certain way.}

But, still, how should we, in a more systematic way, accommodate Blackburn's intuition in the rockclimber case? He wants to say that sometimes it is what happens in C-conditions that matters for how we are disposed. I would rather say this: different
dispositional properties may have substantially different C-conditions, such that it is only the obtaining of these conditions that enable us to judge whether S possesses the disposition or not (e.g., the enabling conditions for rockclimbing include the proximity of rocks, but not the proximity of the macho rockclimber fraternity; the enabling condition for being disposed to stink is that the thing can be smelled (it is not on the bottom of the sea, for example)). This has the effect that dispositions still can be masked, mimicked, or altered (as in my rockclimber variant), and that our analyses of a concept of a dispositional property should reflect what may be special about the C-conditions for this property. In §4 below I shall show what is special about the C-conditions for meaning-dispositions.

e) Third alternative version of the problem about error: being disposed to make errors. The opponent of UD may be inclined to dismiss the two foregoing sections as mere sophistry: the real problem about the possibility of error is where S’s purported error is a result of a disposition S possesses. If S is disposed to use ‘echidnas are spiky’ in an erroneous way, then we need to know why this dispositional use doesn’t go into constituting the meaning of ‘echidnas are spiky’ too. How does UD justify the claim that certain dispositional uses are errors and certain dispositional uses aren’t? Stated like this the objection comes out as:

(7) UD cannot distinguish the meaning-constituting dispositions from the non-meaning-constituting dispositions; and then it cannot accommodate the possibility of error because some errors are the result of non-meaning-constituting dispositions.

I acknowledge this is the strongest form of the objection about error. But I do not think it is strong enough to refute UD. I shall first argue that we have important reasons not to think that all purported errors are the result of dispositions to make errors. This will show that the objection is, at least partially, misguided: in many
cases, purported errors are not relevantly dispositionally based. Then I shall argue that the remaining class of purported dispositional errors may be explained away.

i) errors and the possibility of dispositions.

The main assumption behind this version of the objection is this: when S makes a purported mistake it is because S is disposed to make it. But pressing this version of the objection is tantamount to overlooking what is involved in defending a dispositional account. Compare the fragile cup which self-destructs without being struck. Need we say that it is disposed to self-destruct? What if its self-destruction is really caused by a one-off unhappy series of events (bad clay, a hot lava stream nearby etc.)? Likewise, if S makes a purported error with respect to 'echidnas are spiky', are we then forced to say that she was disposed to make the error? Couldn’t we say that it was a glitch in the causal bases of her dispositions, that she was inattentive, on drugs, joking or something like that? It seems wrong to ascribe dispositions to her on the basis of something she does under unusual, non-standard circumstances.\(^{40}\) That would make a mockery of the aspect of regularity that we associate with having a disposition to do something. And it would be easy for the dispositionalist to dismiss applications made under such circumstances as nothing but errors. Though they involve use of the contested word, the use is not dispositional. Therefore, according to UD, the purported erroneous application of a word doesn’t go into constituting the meaning of the word.

Let me try to indicate which type of argument I am pressing on behalf of the dispositionalist. Consider a similar debate about the concept of causation. Assume we are discussing what caused a certain event, say, a child’s falling off a swing. We ask:

---

\(^{40}\)Forbes 1984, p. 231 expresses this possibility in terms of simultaneous ascription of more than one disposition: hold the other dispositions stable and determine errors for the remaining one. Forbes employs a holistic quasi-scientific method which seems to me somewhat out of tune with normal straightforward understanding and self-knowledge, but I agree with the basic idea: that the ordinary notion of disposition can elucidate the concept of meaning.
was the baby sitter's failure to keep a constant eye on the child a cause of the accident? That is, would we count events that are omitted from the causal chain as causes of the accident? There is some pull towards an affirmative answer because, if the baby sitter had looked, then the accident could have been averted. But the temptation should be resisted. Allowing omissions as causes opens the door to allowing all sorts of implausible 'causes' of the accident: lots of people all over the world failed to do something that could have averted the accident.41 The problem with allowing such crazy causes is not that we cannot see how they could somehow be counterfactually relevant, but that if we do allow such counterfactual relevance, then we change the subject matter: it is no longer our concept of causation we are trying to elucidate, but some distant cousin of it.

The point about dispositions is similar: if we allow breaking (or non-breaking) in all sorts of circumstances to form part of what it is to manifest, say, fragility, then we lose sight of the idea that fragility is a dispositional concept and thereby of our normal concept of dispositions altogether. The opponent's attack comes across then, not as a specific attack on dispositionalism in the debate about meaning, but as an attack on the general, normal, notion of a disposition. So the argument against the opponent is not that he or she is plainly wrong, but that accepting the opponent's point comes at the cost of giving up on our normal concept of dispositions. So if we assume that there are dispositions, then we must assume that with respect to a given disposition some possible errors aren't dispositional.42

---

41 Thanks to Helen Beebee for discussion on this, see Beebee (forthcoming).
42 I think it is consistent with this to hold that everything that happens, even in non-C-conditions, is dispositional. My point is just that talk of dispositions in general must allow us to fix on one disposition and treat what happens in non-C-conditions as non-dispositional with respect to that disposition (this point comes up in Forbes 1984 too, p. 231ff).
ii) is it possible to be disposed to make errors?

Naturally, I have to spell out what the all-important C-conditions are, in a priori, substantial and uniform terms. I will indicate how this may be done below in §4. Before I do that, I will discuss how the opponent to UD could reply to my argument that error cannot always be dispositional. The opponent could grant that some applications of a contested sentence take place in non-C-conditions and therefore count as non-dispositional. But she could insist that there will still be possible situations where S is disposed to make what looks like errors. This would be what can be called 'normal systematic errors'. Here are two potential examples (i) someone is disposed to say 'there is a kangaroo' whenever he or she sees kangaroo-shaped bushes in the bush in the late afternoon; (ii) someone is disposed not to carry when adding numbers. I think that once we have the result from the foregoing section in mind both examples can easily be dealt with.

The kangaroo-example is initially relatively easy to explain away. The UD-theorist can simply say that this concerns falsity in what is believed and as such is beyond the scope of UD. It might be that the speaker made no error with respect to meaning: it seemed to S that it was a kangaroo, therefore he or she says 'there is a kangaroo'. There was no error relative to meaning, if S means there is a kangaroo by 'there is a kangaroo'. However, the example does prompt an important consideration. What if an interlocutor points out that it isn't a kangaroo, but merely a wattle in the dusk? Then there is disagreement between the speaker and the interlocutor. If we focus on the speaker, there seems to be two interesting scenarios: either he withdraws his utterance of 'there is a kangaroo' in the face of the interjection, or he stands by it. If he withdraws we have good reason to believe that he meant there is a kangaroo by 'there is a kangaroo'. If he stands by it we again have two options: either non-C-conditions obtain (he is on drugs, sleepy, has neural glitches etc.); or he is really disposed to use

---

43The adding example is Kripke's, 1982, p. 30-2.
'there is a kangaroo' to indicate mental states about wattles-in-the-bush too. In the first case UD has no problem: the use of 'there is a kangaroo' was made in non-C-conditions so it was an error about the meaning of 'there is a kangaroo'. In the second case, however, the UD-theorist faces what could seem like a hard choice: is the speaker mistaken, or is 'kangaroo' assigned another meaning in his language?

The choice is hard, but not so hard that it threatens the dispositional theory of meaning. There are two possible explanations. Either the sentence is ambiguous in S's language such that there are two dispositions in play: a disposition to use 'there is a kangaroo' to indicate contents about kangaroos, and a disposition to use it to indicate contents about things that seem like kangaroos. And in so far as there are dispositions in play, there will be occasions of use of the ambiguous sentence which count as errors, no matter which of the two dispositions is in play on that occasion (as my argument in the foregoing section demonstrated). Or, the sentence has a dialectic or idiolectic meaning. Nothing hinders a description of the speaker as having assigned an idiolectic meaning to 'there is a kangaroo'. It might not be particularly helpful for the speaker in his communicative interactions, but, if it is dispositionally constituted, then the sentence has a meaning nonetheless (there are still non-dispositional uses of 'there is a kangaroo' which would fall outside its dispositionally constituted idiolectic meaning). What shows that the sentence is meaningful is that it could, in principle, be used to transmit beliefs. Only, the beliefs it could be used to transmit are not very helpful for us.

Hence, we can choose to regard the speaker's recalcitrant application of 'there is a kangaroo' as in accordance with its ambiguous or idiolectic meaning. Of course, this doesn't preclude us from thinking that S is mistaken relative to another language where the term isn't ambiguous, or relative to a communal language. But such a mistake should not be described as a local error with respect to 'there is a kangaroo' in S's communal language or idiolect. Rather, it takes the form of a reproach for
misunderstanding the sentence 'there is a kangaroo' as it is used in some other language, or having failed to learn the correct communal use of 'kangaroo'.

Now we can see what to do with Kripke's example about the speaker who is disposed not to carry when adding. Why say that she is 'adding' at all? It seems she is following another rule: she has assigned another meaning to the word 'plus' than we normally have. Our task would be to instil the right dispositions in her, viz. dispositions to carry when adding. What the UD-theorist has to say, then, is that one-off miscalculations are (non-dispositional) errors about the meaning of 'plus', but that systematic miscalculations indicate that the subject has failed to learn the common meaning of 'plus' and is instead following some idiolectic rule for 'plus' (or at best uses 'plus' as an ambiguous term, in one sense when she describes other's use of 'plus', and in another sense when she herself uses 'plus').

Where does this leave the objection in (7)? I believe my arguments explain away the class of purported dispositional errors, so they provide a pre-emptive answer to the objection that UD cannot distinguish between the meaning-constituting dispositions and the non-meaning-constituting dispositions. However, I do this by aligning the meaning-constituting dispositions closely with other ordinary dispositions, like fragility and water-solubility.

Although this helps me answer the problem about error, it now seems I cannot after all distinguish the meaning-constituting dispositions from the millions of other dispositions that we may have. What is it about dispositions for use of sentences which make them meaning-constituting, in contrast to any other non-meaning-constituting disposition? Why, to put it candidly, isn't a cup's fragility a meaning-

44My account here is similar to the gist of Forbes 1984, cf. p. 232. He also countenances the possibility of non-standard meanings, but doesn't do much to show that it is consistent with our concept of meaning.
constituting disposition? The initial answer to this is of course that a meaning-
constituting disposition must be a disposition to indicate mental states with certain
contents. But this cannot be the whole answer, because we have many dispositions to
do things when we have mental states (breathing, blinking our eyes, doodling when
on the phone etc.). The full answer must, I think, advert to what is special about the
C-conditions for meaning-constituting dispositions. And in fact we have already seen
the beginnings of this answer above, in my remarks about what may happen when
speakers disagree about meaning: it is characteristic of meaning-dispositions that they
are embedded in a particular practice surrounding disagreements about meanings. I
turn now to a detailed defence of this claim.

4. On C-conditions.

My basic answer to the problem about the possibility of error is this: given UD, error
is what happens when S uses φ in non-C-conditions. In the previous section I tried to
make this a plausible claim. Any reason for not characterising error like this is also a
reason for rejecting the very notion of dispositions. But what about the notions of ‘C-
conditions’ and ‘non-C-conditions’ for meaning-constituting dispositions? I must
show that the use-dependent conception of meaning is entitled to this notion such that
a) UD satisfies the apriority, substantiality and uniformity constraints; and such that
b) UD demonstrates what is special about meaning-constituting dispositions, as
compared to other kinds of dispositions.

I do this by first arguing that, if we are confident that our analysis suggests
dispositionalism, then we should have no qualms about specifying the C-conditions
by interdefinition, and I argue that this actually satisfies the three constraints. I then

---

45As discussed in Ch. 4. The fourth, ‘extremal’, constraint I regard as satisfied by the arguments in
Ch. 3 to the effect that there is no better use-independent story to tell about meaning.
build on this suggestion by saying that, if we are confident our analysis suggests dispositionalism for these particular reasons, then we can employ these particular reasons in an interdefinition of what is special about the C-conditions for the concept in question. Of course, these ‘particular reasons’ had better reflect the fact that ‘incorrect use’ is a normative notion, in contrast to the notion ‘failure to break when struck’. That is, the specification of the C-conditions had better pertain to our interests in getting meanings right.

a) Interdefinition of C-conditions for use-dependence.

I think that there is an easy way to obtain an adequate notion of C-conditions for an account like UD. I’ll state this first as a general case for all dispositions, and then for the case of UD.

i) C-conditions for dispositions.

Assume we regard the following claim as true:

\[(1) \ x \text{ is } G \equiv \text{def } x \text{ is disposed to } F \text{ in } C.\]

For example: x is fragile \(\equiv\text{def } x \text{ is disposed to break when struck in } C.\) In so far as we understand (1), and regard it as true, we can say the following about its right hand side.\(^{46}\)

\[(2) \text{ The disposition to } F \text{ is that disposition } D \text{ which is such that } x \text{ has } D \equiv \text{ if } x \text{ were in } C, \text{ then } x \text{ would } F.\]

For example: the disposition to break when struck is that disposition D such that x has the D \(\equiv\text{ if } x \text{ were in } C, \text{ then } x \text{ would break when struck}.\) But now we can use (2) to define the notion of C-conditions like this:

\(^{46}\)Modulo, as before, an amended analysis of dispositions.
(3) The C-conditions for the disposition to F, D, are those conditions C such that \( x \) is in them \( \equiv \) if \( x \) has D, then \( x \) Fs (i.e. then \( x \) would manifest D).

For example: the C-conditions for the disposition to break when struck, D, are those conditions such that \( x \) is in them \( \equiv \) if \( x \) has D, then \( x \) breaks when struck. That is, (3) is a specification of the role of C-conditions in our normal dispositional concepts. It holds, of course, that we cannot understand (3) without understanding (2), and that we cannot understand (2) without understanding (3). So we have an interdefinition of dispositions and C-conditions. That is, by regarding both (2) and (3) as true, we will arrive at an understanding of what both dispositions and C-conditions are.\(^{47}\) I regard this as no less controversial than interdefinitions of the concepts of point and line in Euclidean geometry, or as interdefinitions of the concepts of belief and desire.

**ii) C-conditions for UD.**

I want to employ the interdefinition of C-conditions for the special case of UD. Thus, assume we regard the following expression of UD as true:

\[
\text{(1*) } \varphi \text{ means } p \text{ in S’s language } L \equiv_{\text{def}} S \text{ is disposed in } C \text{ to use } \varphi \text{ to indicate all and only mental states with the content } p.
\]

We can then say something about the right hand side in (1*):

\[
\text{(2*) The disposition of } S \text{ to use } \varphi \text{ to indicate mental states with the content } p \text{ is that disposition } D \text{ such that } S \text{ has } D \equiv \text{ if } S \text{ were in } C, \text{ then } S \text{ would use } \varphi \text{ to indicate all and only mental states with the content } p.
\]

And then, as in the general case proceed, by interdefinition, with:

\(^{47}\)Blackburn 1993, §3 briefly canvasses an idea reminiscent of this.
(3*) The C-conditions for S's disposition to use $\varphi$ to indicate mental states with the content $p$ are those conditions C such that S is in them $\equiv$ if S possesses D, then S uses $\varphi$ to indicate mental states with the content $p$.

With (3*) we have a conception of the C-conditions in UD, in the form of an interdefinition of dispositions and C-conditions. Is this sufficient to prove that I can use the notion of C-conditions in my answer to the problem about the possibility of error? I think it is an important first step, because it is a conception which satisfies the three constraints and which will provide the backdrop for setting out what is special about C-conditions for meaning-constituting dispositions.

**iii) the apriority constraint.**

This is the constraint that UD must hold a priori, that is, as an a priori function of the characteristic content of the concept of meaning. The interdefinition in (2*) and (3*) satisfies this constraint because it is made on the basis of (1*) which is the representation of an a priori analysis of the concept of meaning, and on the basis of the (amended) a priori analysis of our ordinary notion of dispositions.\(^{48}\)

**iv) the substantiality constraint.**

This is the constraint that the C-conditions mentioned in UD must be substantially characterised, not characterised in a whatever-it-takes fashion. Importantly, this constraint is there to ensure that UD can be distinguished from use-independent accounts of meaning. Thus, 'substantial' doesn't mean that the C-conditions must be characterised in intrinsically interesting terms. The C-conditions just have to be characterised in terms that are substantial enough to distinguish UD from a use-independent account. I think the interdefinition of C-conditions does just that.

\(^{48}\)Thus I disagree with Mischevic 1998: cf. my comments on Mischevic in Ch. 4§3.c.ii.
Firstly, it is substantial because it is made on the basis of (1*) which is a representation of a use-dependent analysis of meaning, not a representation of a use-independent account of meaning. So, if the motivation for (1*) is correct, then there is indeed a distinction between what defines the C-conditions for a use-dependent concept of meaning, and what defines the potential whatever-it-takes C-conditions for a use-independent concept.

Secondly, the interdefinition of C-conditions is substantial because it isn’t a whatever-it-takes definition, it is rather a ‘whichever-they-are’ definition of the C-conditions. That is, the interdefinition yields a conception of the role of C-conditions in dispositional accounts, where this role is the same no matter which possible dispositional account is under consideration. Which C-conditions may fulfil this role in different situations is contingent on the concept and situation under consideration. So they are defined as a specific role to be realised in different ways in different situations whichever-they-are. Though this perhaps sounds trivial, it does exclude the whatever-it-takes problem, because whichever way things are may fall short of ensuring the dispositional status of the concept in question (e.g. things may be such, in a possible situation under consideration, that there happens to be no C-conditions to fill the role as defined in (3*)).

v) the uniformity constraint.

The uniformity constraint says that the C-conditions must be specifiable in terms that are either logically independent of truths about the extension of the concept of meaning, or which are such that it is a priori that any occurrence of the concept can be within the scope of an existential quantifier that governs the entire definiens of UD. The motivation behind this constraint is, again, that to be successful UD must exclude the possibility that the concept of meaning is really use-independent. This requirement is flouted if UD employs the concept of meaning in the specification of the C-
CHAPTER VI, THE SECOND TENSION: ERROR AND ROBUSTNESS

conditions such that it is opaque whether those occurrences are use-independent or not.

I believe the specification of the C-conditions in terms of the interdefinition in (2*) and (3*) goes some way to satisfying this constraint, because the interdefinition is arrived at by virtue of regarding UD as true. That is, by setting out UD in detail, we put ourselves in a position to interdefine the notions of dispositions and C-conditions. This is just a way of saying that UD implicitly defines the functional roles for meaning-dispositions and their C-conditions. If this is right, then we should be able to turn it into an explicit definition by ramsifying and thereby guarantee that all occurrences of the concept of meaning occur within the scope of an existential quantifier that governs the whole definiens of UD.49

vi) Miller's worry about circularity.

In my discussion (in §2 above) of the problem of unity, I mentioned that Miller (1997 and 1997b) suggests that dispositionalism seems to be either circular or regressive. I argued that we need not worry about regress because we don't need to specify, in the C-conditions, all the ways in which dispositions for use of sentences may be masked or mimicked. But what about circularity? Miller points out that this may not be a problem for a sufficiently sophisticated dispositionalist, i.e. for a dispositionalist that avails him- or herself of Ramsey-style definitions: '...circularity...needn't worry the sophisticated dispositionalist. It just shows that he will need to become an ultra-sophisticated dispositionalist and adopt the Lewis style analysis'.50 Since this is precisely what I have done—by satisfying the uniformity constraint—I conclude that the specification of the C-conditions in UD isn't circular.51

49 As I described in Ch. 4§3.c.iii.
50 Miller 1997b, p. 188.
51 Though the circularity I avoid is not the circularity Miller has in mind, because he discusses a reduction of the semantic to the non-semantic/non-normative/non-intensional. I am happy to have contentful mental states appear in my analysans (and that is why I have kept the issue to the terms of uniformity rather than circularity).
b) How meaning-constituting dispositions are set apart from other dispositions.

This, however, may still not be entirely satisfactory, because I have not yet shown what sets C-conditions for meaning-constituting dispositions apart from C-conditions for non-meaning-constituting dispositions. The conception of C-conditions in the interdefinition treats the C-conditions as variables that can take different values. Since UD is supposed to be an interesting analysis of the concept of meaning, not of any other dispositional concept, I must also say something about which values these variables take for the concept of meaning. This is particularly important because I have relied on a distinction between meaning-constituting dispositions and non-meaning-constituting dispositions in my defence against the problem about error.52

Note that if I make sure that this further specification is done purely on the basis of the use-dependent conception of meaning, then the above reflections about the three constraints should be sustained. All I am doing, in what follows, is stating what the use-dependent analysis gives us reason to believe specifies the role for C-conditions I have identified for the meaning-constituting dispositions.

i) the remaining platitudes implicitly define the C-conditions for meaning-dispositions.

In my defence against the problem about error, I began to distinguish between meaning-constituting dispositions for a sentence φ and other dispositions, by discussing how speakers may react when they disagree about what was said. That is, I turned to the other platitudes about meaning in order to explain the conditions under which a meaning-constituting disposition works. That approach can be generalised, such that the remaining platitudes implicitly define the C-conditions for the special case of meaning.

52Forbes 1984 does not address this important point. He is right that the objections to dispositionalism aren't specific to the debate about meaning, but that can only be half the story because it is a very good objection against dispositionalism to say that it doesn't distinguish between fragility and meaning (and I do not find Forbes' comments in §VI convincing).
That is, take the platitudes about the possibility of communication, explanatory meaning, robust disagreement, the aspect of practicality in case of disagreement, and first-person authority (all of which I discussed in Ch. 2), state them in one clause such that the concept expressed by the relation "φ means p", and its close cognates, all appear in property or relation name form, then substitute for variables and ramsify.  

We then get a detailed conception of the C-conditions which is derived purely from the platitude-based analysis of the concept of meaning, and which thereby is such that all occurrences of the concept in question can be within the scope of an existential quantifier that governs the whole definiens of UD. This is then a detailed conception of the value of the variable for C-conditions I interdefined above, and, moreover, it is a conception which satisfies the uniformity constraint.

ii) is this a credible account of the C-conditions special for the concept of meaning?

The suggestion is that the C-conditions for UD are given partly in terms that UD will share with other dispositional concepts, partly in terms that are thrown up by the very platitude-based analysis of the concept of the relation expressed by "φ means p". I think this is a credible account of the C-conditions because it defines the C-conditions along these lines:

(4) The C-conditions for the concept of meaning are the conditions such that when S is in them, then facts about "φ means p" can explain S’s use of φ; & S can use φ to communicate and express his or her mental states; & S can disagree robustly about "φ means p"; & S can react to disagreements about "φ means p"; & S is first-person authoritative about "φ means p".

This characterisation of the C-conditions of UD tells us what distinguishes the meaning-constituting dispositions in UD from other dispositions, and it is uniform in

---

53 As described in Ch. 4 §3.c.iii.
the way explained above. The crucial part of this specification of the C-conditions is the part about disagreement and reactions to disagreements, i.e. the part I have called 'the practicality of meaning'.

If the conditions under which we manifest our meaning-constituting dispositions were not such that we could disagree about meaning, and be prepared to react accordingly to disagreements, then the dispositions that are manifested would not count as meaning-constituting after all. Imagine S having a disposition for use of a sentence, e.g., 'this is a kangaroo', but imagine also that S doesn't flinch in the case of what seems to be inter- or intrapersonal disagreement. That is, S is not disposed to reconsider his or her position in the case of disagreement (not even with an eye to reassert his or her position). Then it seems S does not mean anything at all by the sentence 'this is a kangaroo'. This is not the point that a subject may chose to ignore interjections, which of course happens. It is the point that the subject may not properly appraise the possibility of disagreement—and thereby of error—at all. In that case there might be dispositions to do this-or-that in place, but they are not meaning-constituting dispositions.

Without this aspect in place the whole idea of robustness, explanatory meaning and finally communication and expression of thoughts, cannot get off the ground. Since (4) reflects these crucial aspects of the analysis of meaning, I think it qualifies as an account of the C-conditions which are specific for the concept of meaning.

**iii) is UD entitled to this specification of the C-conditions?**

The specification of C-conditions illustrated in (4) satisfies the uniformity constraint. It is also based on the platitude-based analysis of the concept of meaning, and as such

---

54Here is where I keep my promise, made with regard to the rockclimber-case in §3.d.ii.
55In Ch. 2§2.d.
56Imagine someone repeating a mantra, or someone with Tourette’s Syndrome constantly repeating the word 'hideous', as reported in Sacks 1995, pp. 73-102.
it is \textit{a priori}. However, it is one thing to say that ramsifying guarantees uniformity and ensures against circularity, another to guarantee that the ramsification succeeds in such a way that UD remains informative. When we ramsify over the specification of the C-conditions, we have to strip away all occurrences of the \textit{definiendum} but also of other relevantly related terms.\footnote{As in the colour case we have to strip away \textit{all} of the colour terms before we can go ahead and define any one of them. \textit{Cf.} Miller 1997 for an exposition.} The problem is that we may be required to strip so much away that what remains of the specification isn’t substantial enough to explain what is special about the concept in question. On the other hand, if we do not strip away the relevant terms, then we flout uniformity (and get circularity). I do not think my account of the C-conditions is caught in this problem. I can afford to strip away all the occurrences of the concept of the relation expressed by ‘\(\varphi\) means \(p\)” and yet leave enough in the specification to account for what is special about the concept.

Firstly, notice that the practicality aspect of meaning is itself defined in dispositional terms: S must be \textit{disposed} to do something (retract or not) in the case of seeming disagreements. So there is no occurrence of normative terms that would have to be stripped away. This also reflects the notion that it is not terribly important what S is disposed to do in the case of disagreement, as long as S either retracts or stands by what she said. Of course, it will be most useful for S to do what most other speakers do (\textit{e.g.}, retract uses of ‘there is a kangaroo’ when there is no kangaroo). But if S doesn’t do what most other people do we must say that S speaks another language than most other people in this community, or that S attaches an idiolectic meaning to the sentence in question. The fact that S fails to follow suit doesn’t in itself make S’s utterances meaningless. Meaninglessness only arises when S is not disposed to react to seeming disagreements at all.\footnote{This is one of the prime differences between Pettit’s theory, in 1991, 1993, forthcoming, and the story I have tried to tell. For Pettit, it matters a lot which way subjects go in the case of disagreement, primarily because his account is also an account of mental content. I think this puts pressure on his account because it leaves a thick normative notion in the specification of the C-conditions. If we follow my suggestion and ramsify over the C-conditions to obtain uniformity, we must decide whether to leave it in the C-conditions or to strip it away with other occurrences of the.
Secondly, the recognitional capacity of S is described in deflated reliabilist terms: S should be disposed to do something in cases that appear to be disagreements; they need not really be disagreements. What matters is that S is prepared to engage in revision of his or her use whenever there is something that could be a disagreement. This again ensures that S need not engage in construing justifications to hold views about what is a disagreement and what is not, and again we can keep normative notions out of the specification of the C-conditions.

Thirdly, though this still requires that S be able to pass judgements about seeming disagreements, it doesn’t presuppose that S must have the concept [x is a disagreement about meaning] before UD can be true of S. In so far as this capacity involves the concept expressed by the relation ‘Φ means p’, it is in this form:

\[
(5) \ 'x \ is \ a \ disagreement \ about \ meaning' \ means \ x \ is \ a \ disagreement \ about \ meaning.
\]

And as such it just fits into UD itself. That is, if the question arises as to the status of the concept [x is a disagreement about meaning] in the specification of the C-conditions, we can simply iterate UD itself. This will give us simple explanations like: S is disposed in C to use the sentence ‘x is a disagreement about meaning’ to indicate mental states with the content x is a disagreement about meaning. Such a move ensures that UD doesn’t avail itself of any untoward normative notion other than what appears explicitly in UD.

normative concepts in question. If we follow the first option, then we do get a worry about circularity: we cannot understand the concept of meaning and content without already grasping a thick normative notion. If we follow the other option and strip the normative notion away it becomes doubtful whether there is enough left in the definiens to succeed in analysing mental content. I circumvent this problem my restricting myself to the concept expressed by the relation ‘Φ means p’ as analysed dispositionally by the five platiitudes.
iv) is there a permutation problem?

A final worry attaches to the Lewis-style analysis. It was raised by Smith (1994, 2.11) and in its simple form it says that such analyses may be non-circular, and also say what is special about a particular discourse, e.g., colour discourse, because it sets out the platitudes surrounding the characteristic concepts of that discourse. But this is all at a horrible price: because the discourse’s concepts are interdefined, a Lewis-style analysis will lose track of the differences among concepts within the discourse itself. For example, the *definiens* for ‘red’, ‘green’ and ‘blue’ all come out the same.\(^59\) This is an extremely bad result because those concepts should come out as having different contents.

Does this problem bedevil the use-dependent conception of meaning? The first thing to notice is that the uniformity constraint is motivated by a wish to ensure that UD doesn’t rely on any use-independent material. It is not motivated by a wish to avoid mention of all normative terms in the *definiens*. This is reflected clearly in the idea that sentences are used to indicate *contentful* mental states. So I am not trying to show what sets normative discourse apart from non-normative discourses, rather I am trying to show what sets discourse about the relation expressed by "\(\varphi\) means \(p\)" apart from other discourses, including other normative discourses. Hence, my Ramsey-sentence doesn’t comprise the platitudes surrounding all normative concepts. This tells us how the permutation problem should go for UD. If it is a problem for UD, then it should follow that the *definiens* for, *e.g.*, 

\[(6)\] ‘echidnas are spiky’ means *echidnas are spiky.*

comes out the same as for, *e.g.*:

---

\(^{59}\) I generate and discuss this problem in more detail in Ch. 8 §5.c.
(7) 'marsupial moles are born blind' means *marsupial moles are born blind*.

Just as the *definiens* for 'blue' and 'red' comes out the same on a Lewis style analysis of colour. But this clearly doesn't follow because the *definiens* for (6) mentions:

(8) ...mental states with the content *echidnas are spiky*,

whereas for (7) we get the different:

(9) ...mental states with the content *marsupial moles are born blind*.

And UD is entitled to (8) and (9) because UD isn't in the business of accounting for the contents of mental states. I conclude, therefore, that the use-dependent conception of meaning isn't susceptible to permutation.60

*vi) a place for 'altering'.*

In my discussion of the robustness requirement above, (§2), I argued that the problem from the holistic nature of belief-desire explanation can be avoided, if we specify the C-conditions such that the meaning-constituting dispositions are neither masked or mimicked. I think that the discussion of the practicality of meaning offers a way of letting the C-conditions reflect the third type of potential counterexample, altering. This will have the effect of unifying the specification of the C-conditions neatly, such that its different specific aspects answer to constraints on masking, mimicking and altering.

---

60Miller 1997 has raised the permutation problem for a Lewis-style analysis of content. His problem, as I have shown, doesn't bother UD because UD isn't in the business of giving a reduction of the concept of content to non-normative terms. Apart from that, I find Miller's argument a bit perfunctory; the permutation problem can be shown to apply to the colour case, and Miller argues that normative discourse is saliently similar because it endorses paradigm-cases and interdefinitions. But surely there are many dis-analogies between colour discourse and normative discourse.
Here is a dispositional story about what can go on when we disagree about meaning. Assume S says 'there is a billabong' and S* corrects S by pointing out that 'there is a billabong' isn't correctly used like that in L. Then we seemingly have a disagreement about meaning if, that is, S is disposed to do something in this case (either retract or stand by what was said). But no doubt it often happens that when S is challenged in this way his or her disposition for use of 'there is a billabong' changes, i.e. that the fact that there is a disagreement about meaning interferes with S's dispositions and causes S to acquire a new disposition. This may be a dispositional description of our more or less effortless learning of new meanings: instead of going along with an old, possibly idiolectic, meaning for a sentence, we simply cotton on to a new meaning when we are challenged.

Dispositions come and go all the time as the result of many different things, and sometimes the obtaining of the C-conditions for a disposition can make it go away. This is what Johnston calls altering. In the above case, the obtaining of the C-conditions—a disagreement about meaning—makes the very disposition S has for use of the sentence go away, and a new disposition takes its place. This would be a counterexample to the idea that S be disposed to do something (retract or stand by what was said) in the case of disagreement: the obtaining of the C-conditions will simply make the disposition go away and thereby prevent the relevant counterfactual from being true of S. The remedy is easy though, because we know that there is a solution to this problem for other dispositional concepts, such as the fragile cup. So, as before, we align UD with the ordinary notion of dispositions by claiming that the dispositions in UD are meaning-constituting in so far as they are not altered in C, where we remember that UD entitles us to a particular specification of C. This allows the dispositional account to go through, and yet it leaves room for the phenomenon I described, where we change the meanings of our expressions when challenged.

---

61 Cf. Ch. 4 §3.d., cf. also Lewis 1997.
5. Restating UD - concluding remarks.

We are now in a position to restate UD: The *a priori* analysis of the semantic relation expressed by \( '\varphi \text{ means } p' \) implies *a priori* definitional biconditionals such that, where \( '\varphi' \) is a declarative sentence type and \( p \) is a proposition:

\[
\text{UD, final version:}
\begin{align*}
\varphi \text{ means } p \text{ in } S\text{'s language} & \quad \equiv_{\text{def}} S \text{ is disposed in } C \text{ to use } \varphi \text{ to indicate all and only mental states with the content } p.
\end{align*}
\]

Where we are to remember that the notion of dispositions has an analysis—e.g. of the Lewis type (cf. Ch. 4§3.d.i.)—so that masking, mimicking and altering come out as possible, but without working as counterexamples to the analysis. Furthermore, where part of \( C \) is given by a ramsified statement of the platitudes about communication, explanatory meaning, robustness, practicality and first-person authority.

And we can say what is special about the concept of meaning, as opposed to concept of other dispositional properties, by saying something about the specific ways in which the dispositions in UD cannot be masked, mimicked or altered. That is:

(1) The dispositions in UD are meaning-constituting if \( S \) has no further beliefs and desires that masks or mimicks the dispositions for use of \( \varphi \) (as in the joke and dare cases) & if \( S \)'s dispositions are not altered when \( S \) disagrees with someone about meaning (as in the case of effortless learning).

Moreover, this satisfies the apriority constraint because everything in UD is based on the *a priori* platitude-based analysis of the concept of the relation expressed by \( '\varphi \)
means \( p \). It satisfies the substantiality constraint because it sets the concept of meaning apart from use-independent accounts. It satisfies the uniformity constraint because all occurrences of \( \varphi \text{ means } p \) are within the scope of an existential quantifier that governs all of the definiens. Lastly, UD satisfies the extremal constraint because there is no better use-independent account of the concept of meaning, as shown in Ch. 3.

This will be my final statement of UD. It expresses the use-dependent conception of meaning which asserts that the five platitudes surrounding the concept of meaning are compatible such that the three tensions among the platitudes can be resolved. So far I have shown i) that UD is consistent with both the platitude that use explains meaning, and the platitude that meaning explains use (cf. Ch. 5); and ii) that UD is consistent with the platitude that meanings are so regular that we can have robust disagreements about meaning (cf. §2 above); and iii) that UD is consistent with the important platitude that we can make errors about meanings. In the next chapter I shall show that UD is consistent with the phenomenology of first-person authority about meaning.
RESOLVING THE THIRD TENSION:  
FIRST-PERSON AUTHORITY

1. Introduction.

In Chapter 2, I discussed the five core platitudes surrounding the concept of the relation expressed by "φ means p". One of these platitudes was about our privileged first-person authority about meaning I argued that this platitude seems to be in tension with the platitude that meaning is use. In this chapter I discuss this tension. I believe that given the use-dependent conception of meaning, there is no serious tension.

In my discussion in Ch. 2 I discerned two versions of the tension between first-person authority and the meaning is use platitude.¹ The first version turned on the suspicion that a use-based theory of meaning entails behaviourism, this is 'the problem from behaviourist use':

(I) S can have first-person authority about the meaning of ψ without having to observe or infer anything from a third-person perspective on him- or herself.  
[First-person authority].

¹Cf. Chapter 2§2.c. for more general background for these problems and references to their background in the literature.
(II) Use of sentences has to do with overt behaviour, it has to do with what words come out of the speakers' mouths on certain occasions; the only way to gain knowledge about overt behaviour is by adopting a third-person perspective on it (there is no difference between a putative first-person perspective and a third-person perspective because everything that is available in the first-person is available in the same manner to a third-person observer). [Use is overt behaviour].

(III) Therefore, S cannot have first-person authority about the meaning of ψ without having to obtain a third-person perspective on him- or herself.

The other version of the tension stems from the idea that if meaning is use, then social use constitutes meaning, hence 'the problem from social use':

(A) S can have first-person authority about the meaning of ψ in virtue of having a first-person perspective on him- or herself. [First-person authority].

(B) If meaning is use, then other speakers' use goes into determining the meaning of ψ. [Social externalism].

(C) If other speakers' use goes into determining the meaning of ψ, then what one can be authoritative about via a first-person perspective on oneself may fall short of being the true meaning of ψ.

These are both problems for a compatibilist about the platitudes, but only in so far as behaviourism and social externalism are entailed by the meaning is use platitude. It seems straightforward plausible that one must be committed to some kind of behaviourism when one claims that meaning is use. Likewise, it seems plausible that if meaning is use, then other people's use may go into constituting the meanings of the sentences in one's potentially communal language. So the onus of proof is on the meaning-is-use theorist to show that his or her use-based theory doesn't entail behaviourism, and is consistent with, or explains away, the possibility of social externalism. My proposal for a use-based conception of meaning is this: The a priori analysis of the semantic
relation expressed by \( \varphi \text{ means } p \) implies \textit{a priori} definitional biconditionals such that, where \( \varphi \) is a declarative sentence type and \( p \) is a proposition:

\[
UD: \quad \varphi \text{ means } p \text{ in } S\text{'s language } \equiv_{\text{def}} S \text{ is disposed in } C \text{ to use } \varphi \text{ to indicate all and only mental states with the content } p.
\]

(Where part of \( C \) is given by a ramsified statement of the platitudes about communication, explanatory meaning, robustness, practicality and first-person authority). Accordingly, I shall attempt to show that UD doesn’t entail behaviourism in any problematic sense, and that UD is consistent with an intuitively trivial version of social externalism. I shall discuss these two claims in turn.

2. Use-dependence is not behaviourist in any problematic sense.

Behaviourism is a fairly broad term. It covers both psychological and philosophical doctrines, and within philosophy it is used for theories from Ryle’s dispositionalism, through Wittgenstein, to modern-day functionalism. So the claim that UD entails behaviourism doesn’t really say much. My view is that the problem from behaviourist use is only serious for UD in so far as UD entails some particular type of behaviourism which in some way is detrimental to the prospect of first-person authority. This is an important point, because UD has got to be behaviourist in some sense of the word. UD does after all say that meaning is dependent on our dispositions for use of sentences, which must be some kind of behaviour, or latent behaviour. With respect to first-person authority, this ‘latent behaviourism’ comes, initially, to this: S cannot exercise her first-person authority without using the sentence whose meaning she is authoritative about. I want to claim that this type of ‘behaviourism’ is

\footnote{Ryle 1949, Wittgenstein 1953 (e.g. the treatment of the possibility of a private language around §258). For a discussion in relation to functionalism see Braddon-Mitchell and Jackson 1996, Ch. 2.}
different from the types of problematic behaviourism which do lead to problems with first-person authority.

a) Which types of behaviourism are problematic for first-person authority?

The problem from behaviourist use is directed at two necessary conditions for first-person authority. The first condition says that S could not have first-person authority by inferring something from observing her behaviour, because her overt behaviour can come radically apart from its underlying states. The second condition is that the phenomenology of first-person authority rules out that S has to infer what she means by observing her behaviour, i.e. that first-person authority is a priori in the sense that it doesn’t require S to infer or observe anything from a third-person perspective.³

I’ll show that UD is not behaviourist in any sense that entails that behaviour and its underlying states come radically apart, and that UD doesn’t entail that S should infer or observe anything from a third-person perspective in order to have first-person authority.

b) UD doesn’t entail the problematic kind of behaviourism that threatens the first-person perspective.

Behaviourism is a threat to first-person authority if it requires that S investigate his or her own behaviour, and if this behaviour can come apart from the contents of the underlying mental states. In that case, S will not have first-person authority because the use of sentences she has access to come apart from the sentences’ meanings. So we need to ask whether UD is so behaviourist that it gives us reason to believe that uses and meanings come radically apart. Here we have a quick answer because this question is just the question about unity in meanings, which I discussed in Ch. 6 §2. Since UD is dispositional, it does not give us reason to think that uses of sentences come radically apart from their meanings. In general, phenomena may fail to be proper

³For discussion of these, and references to the literature, cf. Ch. 2 §2.e.
manifestations of dispositions if the dispositions are masked, mimicked or altered in various ways. But if the view is that such phenomena suggest that the manifestations of dispositions may come radically apart from the dispositions they are manifestations of, then one begs the question against the idea that there could be dispositions at all. So, since UD is dispositional, we can conclude that even if S has to access behaviour to exercise her first-person authority, this would not in general threaten her first-person authority.4

c) UD doesn't entail the kind of behaviourism which is a threat to the a priori privilege of first-person authority.

However, even though UD doesn't entail the kind of behaviourism where use and meanings can come radically apart, it may still entail the kind of behaviourism which forces S to investigate the environment, i.e. her use, from a third-person perspective in order to have authority about the meanings of her sentences. This would be problematic because first-person authority is supposed to be privileged in the traditional a priori sense of not requiring investigation of the environment. Also, if it were the case the case that S had to investigate her environment, then perhaps S would have to infer her meanings from her use, i.e. infer that she means cockatoos are precocious creatures by extrapolating from observation of a limited number of her uses of 'cockatoos are precocious creatures'. This would indeed go against the phenomenology of first-person authority.

But UD doesn't entail that S has to obtain a third-person perspective on her use, and thus it doesn't entail that S must infer what she means on such an observational basis.

4It seems to me that this is a good answer to Wright's concerns about dispositionalism, which I think amount to the claim that dispositionalism is wedded to behaviourism, and thus cannot solve the problem from belief-desire holism, as sketched in chapter 6§2. This shows in the way he puts the problem for dispositionalism: 'It is between the ability and its behavioural manifestation that the shadow of holism falls, if I may so put it' (Wright 1993a, p. 23). I think Wright's point can be put like this: if only dispositionalism wasn't so behaviourist that it excludes first-person epistemology, then the problem about belief-desire holism could be dealt with. I believe my argument deals with this threat.
This is because UD is not the thesis that meaning is constituted by the *manifestations* of dispositions for use of sentences. Rather, it is the thesis that meaning is constituted by S’s *possession* of meaning-constituting dispositions. So it is just a *non sequitur* to conclude that UD is behaviourist in the sense that it requires observation of manifested behaviour.\(^5\)

**d) Wherein does first-person authority consist?**

Of course this immediately gives rise to the following question: how can S have authority about which dispositions she possesses, save by observing her overt behaviour? This is an extremely difficult question, and I will not pretend to be able to give a complete answer to it.\(^6\) But I think that UD to some extent can sidestep the question (though I shall return to it later). For consider the way I presented the form that first-person authority takes (in the discussion of the platitude in Ch. 2): *viz.* that establishing first-person authority about our meanings comes to establishing our right to disquote truthfully and on an *a priori* and non-inferential basis, as competent speakers of a language. That is, that whenever I use a sentence I could in principle, on *a priori* and non-inferential grounds, truthfully disquote what I just said. Once we present what has to be explained in this way, the question shifts its complexion from the question concerning

\(^5\)Here we get an answer to the second part of Wright’s misgivings about dispositionalism, as expressed in his discussion of the dispositional response to the Kripke-Wittgenstein sceptical paradox about meaning-facts. He notes the argument from infallibility concerning the equation of performance with correctness—which I have considered in the previous chapter—, but says: ‘The reason for dissatisfaction [about the dispositional response] I have in mind, however, is not this. It is rather that [the dispositional response] threatens [...] to make a total mystery of the non-inferential, first-person knowledge of past and present meanings, rules and intentions. [...] An account of the truth conditions of such claims [about past and present intentions etc.] which [...] makes a puzzle of this aspect of their first-person epistemology should be rejected.’ (Wright 1989, p. 236). This comes across as the complaint that the dispositional responses (like UD) are committed to a crude sort of behaviourism which in turn excludes the possibility of any inner mental life. And if there is only behaviour and no mental life, then there is no possibility of first-person authority which is more privileged than the behaviouristic access we have to other speakers’ mental states. But UD isn’t committed to a crude sort of behaviourism and thus doesn’t give rise to a puzzle about the non-inferential phenomenology of first-person authority (I attempt to say more about the phenomenology in what follows).

\(^6\)A satisfactory answer to this type of question would in the end involve the big issue of the nature of consciousness, *cf. e.g.* Braddon-Mitchell and Jackson 1996, p. 34.
(1) how S can come to know mental items, e.g. meanings, in a privileged way, to the question

(2) wherein does a competent speaker’s authority about meanings consist in general?

And UD can in fact say something useful about the question in (2). If S in fact is disposed to use ‘gumtrees are highly flammable’ to indicate mental contents with the content \textit{gumtrees are highly flammable}, then there is an \textit{a priori} presumption that, if S says ‘gumtrees are highly flammable’, then she means \textit{gumtrees are highly flammable}.\footnote{Wright operates with ‘positive presumptiveness’ and makes the notion perform somewhat the same job as I do, but he doesn’t underpin the argument by adverting to dispositions; Wright 1989, §IV.} This is parallel to the case for a fragile cup: if a cup is fragile, then there is an \textit{a priori} presumption that if it is struck then it will break. It would betray a misunderstanding of the concept of a disposition, if one believed that when fragile cups are struck they don’t break. But now we can see wherein S’s first-person authority consists:

(3) if S is disposed to use $\phi$ to indicate mental states with the content $p$, then there is an \textit{a priori} presumption that S gets the meanings right when she uses $\phi$.

If there wasn’t such a presumption, then S just wouldn’t be disposed to mean $p$ by $\phi$. (3) doesn’t say anything about S observing anything, or making any inferences, so I think we are entitled to this conditional:

(4) if S is disposed to use $\phi$ to indicate mental states with the content $p$, then she would be \textit{a priori} and non-inferentially justified in believing that $\phi$ means $p$. 

And (4) is what we were after: a way of establishing, on the basis of UD, that S can, in principle, truthfully disquote what she says in an *a priori* and non-inferential manner. True, this doesn’t say much about how S in various cases can *arrive* at beliefs about what she means, but I have tried to turn the platitude about first-person authority away from this requirement, and towards a requirement concerning wherein S’s first-person authority consists.

With these remarks I have tried to show i) that UD isn’t behaviourist in any sense that may be detrimental to the prospects of first-person authority. Meaning and use do not come radically apart for UD is a thesis about the possession of dispositions, not about their manifestation. As such UD can explain wherein *a priori* and non-inferential first-person authority consists. In the following discussion about social externalism I shall return to two core aspects of my account: the notion that there is an *a priori* presumption that S is authoritative, and the question about the source of S’s knowledge of the antecedent of the conditional in (4).  

3. **UD is consistent with a plausible social externalism.**

The core premise in the problem from social use is that other speakers’ use of a sentence φ may go into constituting its meaning. In the context of UD, this premise is that other speakers’ dispositions for use of φ may go into constituting its meaning. Two questions must be answered in order to assess whether this can give rise to a threat to first-person authority: i) in what sense does other speakers’ use go into constituting meaning? and ii) does social externalism require that speakers obtain a third-person perspective on each others’ use in order to have first-person authority? My view is that there might be doctrines of social externalism that would threaten first-

---

8Ginet 1992, §2 has considerations that resemble my claims here, but he does not do it in terms of the *a priori* presumptiveness that we naturally associate with dispositions, and he settles, unnecessarily in my view, for dispositions without causal bases.
person authority, but that the account of social externalism that sits naturally with UD doesn’t work as such a threat. That is, the way UD accounts for i) ensures an easy answer to ii). In order to show this I will attempt to show how UD can generate a rather trivial but, I think, plausible account of social externalism, and then I will show that this account is no threat to first-person authority.

a) UD and social externalism.

According to UD, use is constitutive of meaning. That is, an expression gets its meaning assigned as a function of speakers’ dispositions for use of sentences. I have offered this as a spelling out of a platitude about meaning. It can help us to see what must be involved in being an understanding speaker, viz. that a speaker cannot but somehow have access to how her community and she herself would employ the sentences of her language. If this is right, then there can be no such thing as a competent language-user in isolation from such patterns of use. If we accept the platitude, then we must accept that consequence of it. This is what the idea that meaning is use-dependent captures. (And in the previous chapters I have tried to answer the question of how meaning is constituted by use).

Here is a way to describe part of what is going on as a speaker understandingly employs her language. As she uses sentences in various ways, two things happen. Firstly, her use forms part of the larger patterns of use, and thus she thereby co-determines the meanings of the expressions in question; this is the user-constitutive aspect of being a competent speaker, as encapsulated in the meaning is use platitude. Secondly, in some way she relies on her fellow speakers and, as it were, her own earlier selves because their use goes into determining what she means too; we can call this the user-subordinate aspect. This is what is captured by the platitudes about explanatory meaning, and robustness and error.
How can the user-subordinate aspect be characterised? In what ways does an individual speaker subordinate herself to other speakers and her own earlier selves? Here is a platitude-driven story. According to UD, we can say that the speaker goes along applying various sentences the best she knows, and that there is an \textit{a priori} presumption that most of the time she is doing it right (this is the presumption I discussed above). But, on the other hand, since occasional error is always possible, any particular application of a sentence may be out of accord with its meaning. That is, it may be out of accord with the robust and potentially explanatory meanings which are constituted by the dispositional use of her earlier selves and her peers. To capture the interaction between these notions, we introduce the aspect of practicality which says that subjects must be properly appraised of seeming disagreements, and react to them in some way, in order to count as competent and understanding speakers, or, as I shall call them \textit{linguistic insiders}. Let us say that as a speaker goes along she \textit{authorises}—to use a phrase suggested to me by Pettit—her peer’s and her own earlier use, and thereby opens herself to criticism in the light of this. This notion gives expression to the natural idea that, where there are conflicting uses of a sentence, speakers may disagree and try to decide which use is correct and which should be withdrawn. Such disagreements and their consequences are what make use patterns—and hence meanings—shape up.\textsuperscript{9}

This provides us with a rather trivial but plausible notion of social externalism: according to UD, there is a sense in which meaning is dependent on social use. Though every individual’s use is presumed correct, it is still \textit{possible} that any such use is incorrect. Such incorrectness shows up in the disagreements S may have with other speakers or, as it were, with her own earlier selves. In those cases, S may subordinate her use to the use in the community of which she is a linguistic insider. This is a rather

\textsuperscript{9}The exact nature of the disagreements is not my concern here. Their nature play a crucial role in Pettit’s theory which is concerned with how our contents manage to track truth and rationality. According to Pettit, normally, constancy of use is an objective of the negotiations. They may focus on normal or ideal conditions, idiosyncracies, inferential patterns, speaker’s intentions \textit{etc}. Pettit 1991, 1993, forthcoming.
trivial account of social externalism because it doesn’t say that meaning *must* be constituted socially: idiolectic use may still be possible, in which case S, as it were, forms her own linguistic community.

*b) UD-based social externalism and first-person authority.*

With this brief account of UD-based social externalism in mind, I will now show how first-person authority fits into the picture. According to the above story, a linguistic community, consisting of individuals—linguistic insiders—with each their history, will be authoritative about the meanings of their sentences. Hence, there is a sense in which a linguistic community as such has authority about meaning. This gives rise to this question: does the way in which a community of linguistic insiders has authority about meaning detract anything from the way in which an individual linguistic insider has first-person authority about meaning? This question arises because we have discerned a user-subordinate aspect with UD-based social externalism.

However, it doesn’t follow from the user-subordinate aspect that a speaker can only have authority about the meanings of her expressions by way of consulting her peers, or as it were, by thinking of herself and her behaviour in the third-person. This is because the user-subordinate aspect comes with the user-constitutive aspect: in virtue of being a linguistic insider a subject is herself authoritative with respect to meaning, that is, her use of an expression co-determines its meaning. So, as in the above discussion of the problem from behaviourist use, the notion of use-dependence *a priori* gives rise to a presumption that a speaker—a linguistic insider—gets the meanings of her utterances right. And then the notion of an individual’s first-person authority comes naturally: given that *my* use co-determines the meanings of the sentence ‘wobbegongs are nocturnal’ there is no reason to hold that I have to draw on third-person knowledge of my peers’ behaviour to know what I mean when I say ‘wobbegongs are nocturnal’. 
Moreover, on the strength of this assumption about use and meaning, the subject will have available \textit{a priori} evidence that she is authoritative: it is \textit{a priori} that if she is a speaker of a given language, then there is a presumption that she counts as first-person authoritative. Notice that whatever normative aspect there is to this story is brought out by consideration, in dispositionalist terms, of the platitudes about the concept of meaning. This is not to say that learning a language is an \textit{a priori} process for of course many contingent matters have an impact on that. However, though truths about meaning have contingent ancestry, knowledge about meaning may nevertheless be supported by such \textit{a priori} evidence. Hence, this far at least, the use-dependence of meaning does not pose a threat to the apriority and privilege of first-person authority.

It is necessary, though, to note that the user-subordinate aspect does detract something from a speaker's first-person authority about what she means—after all, we are not Humpty Dumpty-ish dictators about meaning. The answer is that, though S's authority about what she means does not have to draw on any third-person perspective, it is in some sense precarious. Her authority, for which she has \textit{a priori} evidence because she is a linguistic insider, and which is exercised through the first-person perspective, is \textit{defeasible} precisely because she, as a linguistic insider, has authorised her earlier selves and her peers, and is therefore compelled to do something in the case of disagreements about meaning.

Importantly, this kind of defeasibility ensures that we are not dealing in infallible—Cartesian—authority, since S's \textit{a priori} evidence for her authority may be defeated through disagreements with her earlier selves and her peers. The idea that \textit{a priori} evidence for a judgement can be defeasible should not offend against our epistemological intuitions. The judgement in question may attempt to establish an \textit{a priori} contingent truth and hence the evidence one may have will be \textit{a priori} (here in the sense that it is based on meaning), and yet things may in fact not be as the judgement says they are.
I conclude that the kind of social externalism underpinned by UD doesn’t pose a threat to first-person authority about meaning. UD provides us with a defeasible but nevertheless *a priori* presumption that any linguistic insider will have first-person authority. This reasoning follows the reasoning in the above discussion of behaviourism in as much as it turns on the *a priori* presumption that properly disposed speakers get their meanings right, and that they can have evidence for this without having to investigate the environment or otherwise obtain a third-person perspective on their use.

4. First-person authority and linguistic insiders.

So far my defence of UD against the threat to first-person authority has been based on the strength of this conditional:

(3) if S is disposed to use φ to indicate mental states with the content p, then there is an *a priori* presumption that S gets the meanings right when she uses φ.

But now it is easy to imagine someone objecting that this just moves the issue of first-person authority to the issue of how we can know the truth of the antecedent of (3). On what basis do we know how we are disposed? In this final section of this chapter I will attempt to address this concern.

Notice that this is really a new concern. It doesn’t directly concern the issue about first-person authority, but rather the issue of how we as such know which language we speak. UD brings this out clearly because it says that S’s being disposed to use φ
to indicate the content $p$ is equivalent to $S$'s meaning $p$ by $\varphi$. So the new question just is the question of how we know the meanings of the sentences of our language.

**a) A worry about diachronic first-person authority.**

There is, however, a way of turning this new worry into a one that *does* concern first-person authority. This is an issue about diachronic first-person authority, *i.e.* about our capacity to re-identify our past thoughts with first-person authority. The problem is that the environment, or the patterns of use, may change in ways relevant to how the relations between words and meanings are determined.

To demonstrate this point, consider a non-modal version of Putnam's Twin-Earth cases according to which Earth and Twin-Earth are in the same universe (*e.g.*, on opposing sides of the sun).\(^\text{10}\) Assume that $S$ is transported overnight from Earth, where 'that's water' means *that's water*, to Twin-Earth where 'that's water' means *that's twater*. The idea behind the Twin-Earth shuffles is that the two planets may be so similar that a subject does not keep track of where she is when she changes worlds. And, by analogy to the community case, just as she may not be aware that she has changed worlds, she may not be aware that she has changed linguistic communities for normally we do not actively track such facts. And just as she may not be aware that she has changed communities, she may not be aware that her dispositions have undergone changes to such an extent, for example, that she cannot explain use by robust meanings for normally we do not actively keep track of our own earlier use. But such changes are all at least possible, and some are even fairly common.\(^\text{11}\)

Doesn't this mean that when $S$ says 'yesterday I believed that water falls from the sky', she might very well, and for entirely contingent reasons, fail to know whether

---

\(^\text{10}\) *Cf.* Ch. 2 §2.b for the various versions of Putnam's story.

\(^\text{11}\) Twin-Earth switching cases appear in Boghossian 1989b; and in Warfield 1992 who in turn is criticised by Ludlow 1995. Ludlow argues that something like Twin-Earth switching cases are prevalent, because of social externalism, and that therefore first-person authority is threatened. Here I accept that premise, but below I shall argue that this needn't be a threat to first-person authority.
by ‘water’ she means *water* or *twater* or something entirely different? That is, wouldn't she have to keep track of her way through the world(-s) to have authority about her past thoughts, and wouldn't this be an empirical matter? If this is the case, then the domain of privileged access would be strictly limited to synchronic identification of contents.\(^{12}\)

The problem is not that the subject S, has no authority at time, \(t_1\), when she says: ‘I believe water falls from the sky’, or that she has no authority at a later time \(t_2\), when she says ‘I believe water falls from the sky’. She has authority at \(t_1\) and at \(t_2\), respectively, even if she in the interval has changed interpretative contexts, say, from a *water*-context to a *twater*-context. This is what I tried to establish above in my defence of UD-based first-person authority about meaning. Rather, the problem is what happens if, at \(t_2\), she says: ‘At \(t_1\) I believed that water falls from the sky’. It will be indeterminate whether she means *water* or *twater* by ‘water’, and hence unclear what she has first-person authority about, simply because it will be unclear which of the contexts should be regarded as the relevant context of use.

The first thing to notice about this problem is that is builds on a perfectly legitimate aspect of use-dependence. Our dispositions for use come and go all the time, and there is no reason to insist that patterns of use be totally constant over time. Indeed, it is part of the story about first-person authority that subjects’ use is constitutive of meaning, so in exerting their authority meanings may change. In this way, they can be said to undermine their own authority to re-identify contents, simply by being an authoritative language-community. It is this aspect of use-dependence which the above problem seizes upon and generalises.

The crucial question is, therefore, if this natural aspect of UD can be contained so that it doesn’t generalise and become a threat to first-person authority as such. Even if use-

\(^{12}\)Similar considerations appear in Wright 1992b, p. 80; see also Boghossian, 1989b, p. 22-3.
dependence as such introduces some limits to diachronic authority, we would still like to say that those limitations are fairly innocuous. In particular, we would want to restrain the role of possible rampant Twin-Earth shuffles. It is characteristic of such shuffles that we cannot detect their occurrence, and thus we would have no cognitive hold on the defeasibility they would entail for diachronic first-person authority.

b) The threat to diachronic first-person authority is a threat to first-person authority as such.

The very idea that meaning is use-dependent is credible only on the background-assumption that patterns of use are reasonably regular and constant. If it is assumed that they can change wildly and undetectably all the time, then the very notions of robust disagreements and potentially explanatory meanings get no grip. If semantic relations change like that all the time, then subjects' claims to diachronic first-person authority would be constantly defeated: they would, as it were, be perpetual linguistic outsiders. It would no longer be determinate where they belong in the linguistic scheme of things. Since we do not constantly monitor our patterns of use, and, in addition, since Twin-Earth changes can be undetectable, we cannot rule out that this is actually the case. Such a scenario has radical consequences: the whole idea of meaning as use-dependent would not get off the ground.13

This puts the threat to diachronic authority about meaning in a new light. In being a threat to diachronic authority, it is also a threat to the very idea of the use-dependence of meaning itself. And if that notion is discredited, then the above use-based defence of (synchronic) first-person authority fails too (together with our alleged a priori explication of the concept of meaning). The situation seems to be this: either meaning is use-dependent, or constant Twin-Earth style changes in use-patterns are possible. So if we stick with use-dependence, then something has to happen with the possibility

13Notice that these considerations are orthogonal to whether first-person authority is considered privileged or not—Twin-earth shuffles can be undetectable to us even given all our present empirical knowledge. Indeed it rather seems understanding as such is at stake.
of constant change. But constant change is manifestly possible, so we cannot outlaw it.

Here we should remember that the discussion of first-person authority about meaning is essentially an epistemological debate, so perhaps the key to the issue lies in a discussion in those terms. The line of reasoning I suggest goes like this: we have, as I hope to have shown, good independent reasons to accept that meaning is use-dependent, so we have good reason not to believe that our patterns of use change all the time.\textsuperscript{14} But since such changes are possible, and also an integral part of the use-based story itself, we cannot outlaw them. Instead, we had better ignore the possibility of constant widespread Twin-Earth shuffles. Below I shall give a brief suggestion as to how this may be done.

c) Elusive authority about language.

We need a way of showing that the possibility of Twin-Earth shuffles doesn’t impede our knowledge about the antecedent in the conditional:

\begin{equation}
(3) \text{if } S \text{ is disposed to use } \varphi \text{ to indicate mental states with the content } p, \text{ then there is an } a\ priori \text{ presumption that } S \text{ gets the meanings right when she uses } \varphi.
\end{equation}

or, by UD, equivalently,

\begin{equation}
(4) \text{if } S \text{ means } p \text{ by } \varphi, \text{ then there is an } a\ priori \text{ presumption that } S \text{ gets the meanings right when she uses } \varphi.
\end{equation}

Without an account of our knowledge of the antecedent, my defence of the general notion of defeasible \textit{a priori} evidence for first-person authority will come unstuck. My suggestion is to counter scepticism about our knowledge of which language we speak.

\textsuperscript{14}This is what I showed in Ch. 6 §2.
by thinking of such knowledge in terms of elusiveness, somewhat in the sense of Lewis’ (1996) idea of ‘Elusive Knowledge’. According to this idea, rampant Twin-Earth shuffles are not deemed impossible, but, rather, properly ignorable. If we mostly allowed ourselves to consider them relevant possibilities, then knowledge of our language, and thereby first-person authority, would disappear in front of our very eyes (much in the way knowledge that p, for any p, would disappear, if no possibility in which not-p could properly be ignored). In the tradition of Lewis, I will set up a definition and formulate some rules of ignorability. Here is the definition:

*Elusive knowledge of language:*

A subject S knows her language L if S has no reason to doubt she is a linguistic insider (i.e. a competent understanding speaker of L).

The definition has the following proviso concerning which reasons for doubt are relevant:

*Proviso:*

Some possible reasons for doubt are properly ignored.

And here we must say something about what is properly ignored, or else the danger is that ordinary disagreements about meaning (which are surely relevant reasons for doubt) are deemed ignorable. There are only two rules of ignorability, one positive, and one negative. The first is the:

*Rule of linguistic reliability:*

Subjects may properly, but defeasibly, ignore the possibility that they are not reliable and normal language users.

The second is the:
Rule of Disagreement:
Subjects may not properly ignore any actual disagreement which seems to be about meaning.

Here is how the rules work and interact. By the rule of linguistic reliability, subjects may without further ado presume that they are competent linguistic insiders. The default case, as it were, is that we do know our language, whichever it is. But, by the rule of disagreement, this presumption may be overruled by any actual disagreements that seem to be about meaning. In such cases, S’s status as a linguistic insider is being questioned and she must respond accordingly, either by retracting or standing by what she said. This should contain the threat from the possibility of rampant Twin-Earth shuffles because, given this definition, her knowledge of her language, and thus her first-person authority, is only threatened when their possibility is actually raised. And this seems right: how could we as much as begin to raise the spectre of rampant Twin-Earth shuffles unless we ordinarily think that we understand each other?

In general the rules are tailored so that, firstly, linguistic insiders count as such without having to all the time judge that they are so, hence there should be no further issue about our knowledge of the antecedent in (3) and (4); and secondly, so that the aspect of practicality can be respected in the case of disagreements about meaning (such that, e.g., if S is disposed to retract what she said, then her a priori evidence for her first-person authority about that sentence has been defeated).

On this account, knowing our language is elusive knowing. The default case is that we have the knowledge, but as soon as someone raises the possibility that we don’t, then it may elude us on that occasion. This kind of knowledge provides the basis for our first-person authority about meaning, and therefore this latter notion cannot be any stronger. It follows that our a priori evidence for first-person authority is defeasible, in the sense given by the definition of elusive knowledge of our language.
d) Comments on the idea of elusive a priori evidence for first-person authority.

Let me end with a more speculative reason for why elusiveness is especially suited for first-person authority about meaning. It has to do with the fact that over time meanings change as a function of use-changes. Say a group of interlocutors detect disagreements about the meanings of some of their sentences, and then discuss who should withdraw their use and who shouldn’t. During such discussions, it may happen that the meaning of the term undergoes more or less subtle changes—because discussions are themselves part of proper use. Thus the fact of the matter—the meanings—may change as we, so to speak, look at it. And that is an important reason why elusiveness seems appropriate for first-person authority about meaning. As soon as we enter such discussions, and have a closer look at what we supposedly have authority about, we may change the very object of investigation. A dispositionalist theory like UD is especially apt to capture this phenomenon because it fits the notion of dispositional ‘altering’ very well. It is a possibility that, even if people are properly disposed as linguistic insiders, the way they are disposed may change when the time comes for those dispositions to manifest themselves (cf. Ch. 6 §4 for more detail).

5. Conclusion.

In this chapter I have discussed the apparent tension between the meaning is use platitude and the platitude about first-person authority. My argument is that, given UD, we have no reason to think that there is such a tension. This is because UD isn’t behaviourist in any sense that is detrimental to the prospect of first-person authority, and because UD is consistent with a form of social externalism which, if anything, makes us understand better wherein first-person authority consists.
My arguments rest on the assumption that if we know how we are disposed to use sentences, then we have defeasible *a priori* evidence for our first-person authority. I went on to argue that the antecedent of this assumption should not be too controversial. We can devise an epistemology for our knowledge of our language such that, eventually, first-person authority is underpinned by elusive *a priori* evidence.
MINIMALISM, USE-BASED MEANING, AND DEFLATIONIST TRUTH

1. Introduction - minimalism and metaphysically infected concepts.

I want to end this essay by considering a possible alternative to my use-dependent conception of meaning. This alternative refutes primitivism, but refuses to adopt use-dependence in its stead. Minimalism, as it is called, puts itself forward as the sensible alternative to hopelessly reductive use-based theories and metaphysically infected primitivism. According to minimalism, purging a concept of a realist or objectivist metaphysical picture should not automatically lead us to reconstruct the concept in anti-realist or subjectivist terms. There will often be an objectivist, but metaphysically austere, minimal doctrine to be had instead. Such doctrines reflect the objectivist commitments of a discourse, but do so without appealing to substantial reductive or platonist facts. If there really are such positions available, then that is surely sound advice. We should not give up our objectivist or factualist intuitions about a concept just because some philosophical pictures turn out to be untenable.

I don’t think this kind of position is available in the debate about the concept of meaning. It is not clear in what sense the available minimalisms about meaning are

1This statement is inspired by Johnston 1993.
minimal. They seem either to be inadequate in their own right or, at best, disguised use-dependent accounts of meaning. Moreover, adopting use-dependence, as I have argued, is fully consistent with sustaining our objectivist intuitions about meaning, as those intuitions are captured in the platitudes about meaning.

Paul Horwich’s recent use-based theory of meaning is similar to and justified by deflationism about truth. This would be the best sense in which meaning is minimal, so I consider it in detail below. I shall argue that its deflationist element can be factored out.

a) An example of minimalism from the debate about truth.

The debate about the concept of truth provides a core example of the dialectic leading to minimalism. On the one hand, there is a commitment to a substantial analysis of the concept of truth, for example, in the form of a substantial correspondence relation between sentences or propositions and states of affairs. On the other hand, there is the anti-realist recoil away from such ‘metaphysical realism’: according to this position truth should be conceived in terms of justification conditions, verifiability at the ideal limit of investigation, etc. Neither position is satisfactory: substantial notions of truth are usually circular or regressive, and anti-realist concepts normally force us to give up too much of the idea that truth is independent of our responses and what we do. It is here that minimalism comes in. It says that truth is not a substantial property of sentences or propositions, merely a handy device for making certain types of generalisations.² But, according to minimalism, the fact that truth is not substantial doesn’t mean that we should cease to regard it as an objective notion. Minimalism points out that we do not in practice care much about aiming at establishing substantial relations between sentences and states of affairs, and that in its minimalist guise the concept can do everything else that we care about it doing.

²Cf. e.g. Horwich 1991.
Minimalists about truth claim this feat on surprisingly sparse grounds. Though there are various versions of minimalism, their common core is that the concept of truth can be all but captured by an infinite conjunction of equivalences like this:

(i) the proposition that grass is green is true iff grass is green, and
(ii) the proposition that snow is white is true iff snow is white, and...so on.\(^3\)

The strategy for defending minimal truth, then, is to go through all our core beliefs about truth, and show how none of them rely on more than the minimal conception of truth.

\(b\) Minimalism as a general strategy.

Johnston argues that this dialectic may be quite general such that, whenever an anti-metaphysical revision of a concept is suggested, we should first look for a suitable minimalism to take the place of the old inflated concept, rather than resorting directly to subjectivism and anti-realism. He calls the anti-realist, who argues a certain concept is response-dependent, a 'Revisionary Protagorean', and says:

I do not say that Minimalism is the last word for all the concepts of philosophical interest, but it is the position which the Revisionary Protagorean must first overcome. And there are enough metaphysically austere response-independent notions to severely limit the successes of even the Revisionary Protagorean.\(^4\)

---

\(^3\)This is Horwich's version (1991). Wright 1992 has a slightly more inflated version, but the differences need not detain us here.

\(^4\)Johnston 1993, p. 111. I discussed the notion of 'response-dependence' in more detail in Ch. 4.
With regard to my account of use-dependent meaning, the following question then naturally arises: we know that the 'metaphysically infected' primitivism about meaning is a non-starter, but isn’t there a minimalist notion of meaning available which we should adhere to, rather than recoiling all the way to a use-dependent conception of meaning?

My answer to this question will be ‘no’. There are no adequate meaning-minimalisms which are use-independent. That is, there is no gap between the use-independence associated with primitivism, and use-dependence. I shall show this by discussing three degrees of minimalism about meaning: (i) very minimal minimalism, which is a non-starter; (ii) versions of minimalism, which are not non-starters, are not minimal because rely on use; and (iii) a use-based account of meaning, which is similar to and justified by deflationism (as e.g. Horwich *Ms.*) can have its deflationist content factored out.

2. ‘Minimal’ minimalism.

One can imagine a ‘minimal’ minimalism which accounts for facts about meaning simply by giving a list of words or sentences and their meanings like this:

(i) ‘dog’ means *dog*,
(ii) ‘cat’ means *cat*,
and so on.

Or perhaps, with a focus on reference, like this:

(a) ‘Fred’ refers to x iff x = Fred,
(b) ‘The Eiffel Tower’ refers to x iff x = The Eiffel Tower,
and so on.
Or, with an emphasis on satisfaction:

(I) $x$ satisfies ‘...is green’ iff $x$ is green,
(II) $x$ satisfies ‘...is square’ iff $x$ is square,
and so on.

These lists should then be shown to provide everything that is required to enable the concept of meaning to do its job for us. I don’t think anyone actually holds any of these accounts, but that is not important for my present purpose which is simply to point out that on their own such accounts are non-starters.

The point I want to make is exceedingly simple. How do we explain which things are apt for meaning some thing, referring to something, or being satisfied by something? That is, how do we know which things go into the quotation-marks in clauses in the three kinds of lists? We know that not everything can have a meaning, or refers, or is satisfied; or rather, we know that there is something special about the things that obtain such status. So neither of the three accounts can be complete as I have presented them. They must come with an account of what it is to be meaning-apt, or reference-apt, or satisfaction-apt.\textsuperscript{5}

This is not yet to say anything terribly incriminating about minimalism, it is only to say that some such story must be forthcoming. My argument will be that the only probable story that can be told, short of reverting to primitivism, is really inflationary and certainly use-dependent.

\textsuperscript{5}This point is similar to what leads Wright 1992, Ch. 1-2 to talk about truth-aptness. Daniel Nolan has used this point in relation to Horwich’s account of meaning.
3. How minimalism relies on use.

Here is a way to focus the problem for 'minimal' minimalism. One could begin to give an account of meaning-aptness by saying what it is for an expression to belong to a language. This is in fact how Johnston himself begins to set out meaning-minimalism. But it would not be sufficient merely to say what it is to belong to a language, because it matters very much for meaning which language a given word belongs to. We can see this by considering the notion of error. For it to be determinate whether an application of 'dog' is an error or not, it must be determinate whether 'dog' belongs to a language where 'dog' means dog, or where 'dog' means cat, or something else. Disquotation only makes sense within a language. What counts as an error in one language possibly doesn't count as an error in another language. So formulae of this type:

\[(1) \varphi \text{ means } p, \]

should really be recast in this form (and this is how Johnston himself does it):

\[\text{M-platitude (i): } \varphi \text{ considered as a sentence of } L \text{ means that } p \text{ iff in uttering } \varphi \text{ in the assertoric mode competent } L\text{-speakers would be asserting that } p.\]

Here the content of an utterance, say, 'Fido is a dog', is explicitly made relative to a language L. That would help with the problem of error because we would have to acknowledge beforehand that it is the case that S speaks the language L where, say, 'dog' means dog. The alternative, and damaging, story (where 'dog' means cat) could not be told in this case.

---

7This is adapted from Johnston 1988, p. 36.
The issue now is whether M-platitudes are consistent with minimalism. The crucial point is this: what would determine who is right if we disagree about whether S speaks L or L*? Here is a revealing quote from Johnston: 'Talk of meaning in the sense of the M-platitudes is talk of reified use.'\textsuperscript{8} This seems to indicate that we can formulate such platitudes in terms of the use of expressions. So, S’s dispositions to use certain expressions in certain conditions determine whether she speaks L or L*. If she normally uses ‘dog’ in certain circumstances C, then she is, say, an L-speaker, not an L*-speaker. Consequently, when we substitute ‘Fido is a dog’ and *Fido is a dog* for ‘q’ and ‘p’ in our M-platitude, it is already determined, in virtue of use, that ‘Fido is a dog’ means *Fido is a dog*.

Where does this leave the minimal theory of meaning? It does set out some minimal-looking platitudes about meaning, but only because talk of the meaning-stating platitudes require talk of patterns of use.\textsuperscript{9} This is the level at which error is made possible: by allowing the identity-conditions for the expression in question to invoke use. It is still debatable whether a minimalism combined with a use-based theory of meaning should count as minimal but, crucially, it certainly isn’t use-independent. This sustains my claim that there is no space for a minimal, but use-independent, theory of meaning between primitivism and use-dependence.

We have taken the first step in overcoming minimalism: minimalism about meaning is not a position which offers itself as a use-independent, but metaphysically austere, account of meaning. I will now turn to a theory which is minimalist, or deflationist, but which nevertheless allows a substantial use-based account of meaning. My main aim is to show that its minimal element can be factored out, to leave us with a substantial use-based theory.

\textsuperscript{8}Johnston 1988, p. 37.

\textsuperscript{9}Cf. Johnston 1988, p. 38.

To begin with, I shall set out some core notions of Horwich's account of meaning. In the next section, I first point out a potential problem about the diversity of mental states underlying the use of words. I propose a functionalist reading of Horwich to accommodate this, and then raise a problem for this proposed reading. Finally, in section 6 I will indicate an alternative way of accounting for word type meaning, and thereby try to assess what makes Horwich's explanatory strategy work.

a) Horwich's project.

The question Horwich is concerned with is: "What is meaning?" That is, how come some inert sounds and ink-marks are imbued with meaning and others aren't; how come they can reach out and mean some determinate portion of the world?10

These questions have received enormous attention in the last 100 years or more. But despite all the effort, there is no consensus on what meaning is. There is hardly any consensus on what we should expect a theory of meaning to look like.

Horwich's perspective promises not only to give an account of what constitutes meaning, but also to show how to solve all the traditional problems in the area—or at least to show why they are not real problems.

The slogan that "meaning is use" can seem trivial, obscure or vague, steeped as it is in Wittgensteinian aphorisms. Horwich nevertheless defends a use-based theory of meaning. If, moreover, use can be specified in non-semantic terms, then this promises

to be a *reductive* account of meaning. To achieve this, Horwich must show how some aspect of our use of a word or sign, described in non-semantic terms, could suffice to assign a determinate meaning to it. The meaning-constituting properties Horwich identifies are dispositionally circumscribed *basic regularities* of use.

It is quite important to notice that Horwich does not promise us an analysis of what meanings *are*. The slogan "meaning is use" does not amount to the claim that meanings somehow *are* uses. Rather, the claim is that use constitutes the relation between word and meaning. Thus, for example, it is in virtue of the use of the word 'wombat' that it comes to mean *wombat* in English. Meanings themselves are the concepts typically expressed by tokens of the word. For example, when I say "there’s a wombat" that typically indicates the presence in me of a thought with the propositional character *there's a wombat*. The meaning of the word 'wombat' is then the concept *wombat* which is an abstract component of such mental states.

*b) A central problem in the traditional approach.*

But what makes Horwich think he can succeed where so many others have failed? To see this, we must discuss what he discerns as one of the core problems with the traditional approach. The problem concerns the manner in which meaning-constituting properties are, or are not, *relational*.

We saw that there is a straightforward sense in which meaning is relational. A word’s meaning-constituting property is that which ensures that use of the word *indicates* the presence, within the mind of the speaker, of a certain concept. But this aspect of relationality should not make us look for relational meaning-constituting properties. Horwich argues that relational properties can be constituted by non-relational

---

12Ms. p. 23, 43, 89f.
properties. The argument itself rests on a general point about what constitution is.\textsuperscript{13} I will not go through it here. Instead I note, with Horwich, that it seems very plausible that meaning-constituting properties are not all relational. There are many different types of words. We might be tempted to give relational accounts of the meanings of some: \textit{e.g.} ‘wombat’ gets its meaning because it is commonly used for wombats, ‘red’ gets its meaning because it is what we commonly say in front of things that are clearly red. But other words need not have relational meaning-constituting properties: there is no intuitive relationality in what constitutes the meanings of such words as ‘and’, ‘plus’, and other words, such as ‘democracy’ do not immediately call for relationality. In general, different words have different functions and these functions can be explained in various ways. Surely Horwich is right that the meaning-constituting properties for some words can be monadic.\textsuperscript{14}

But why should we insist on relational properties at all? Why is this an issue? The answer is that many of the traditional approaches to meaning have rested on the notion that the only way to achieve a reductive account of meaning is by way of extensional relationality. That is, these accounts have assumed that, in order to analyse what it is for a word type \(\alpha\) to mean something, say, \(F\), there must be some uniform relation between \(\alpha\) and the extension of the predicate ‘\(F\)’, the Fs. For example, for the word ‘echidna’ to have the meaning \textit{echidna}, there must be some relation between the word and echidnas. According to such a view, any word has its meaning assigned in virtue of such a favoured extensional relation. This can be, for example, some uniform causal relation. Of course, this strategy is based on the intuition that if we are going to explain why ‘dog’ is \textit{true of} dogs, then there had better be some relation between tokens of the word and the dogs.

\textsuperscript{13}Ms. p. 22-23.

\textsuperscript{14}To repeat, this is not the point that no meaning-constituting properties are relational, it is the claim that relationality is not a necessary condition for a property to be meaning-constituting.
If Horwich is right, then this demand in unjustified. Nothing in the *analysandum*, "α means \( F \)", indicates that we should insist on a uniform extensional relation such as causality. This is an important result because it is exactly when trying to force an account of meaning into a uniform relational mould that we come up against one of the seemingly insurmountable problems in the theory of meaning. The problem is that it seems impossible to adequately specify a relation between a word and an extension such that the meaning of the word can be *read off* from this relation. One cannot help but beg the question when specifying the relation. How for instance do we exclude dog-like sheep from the extension of the word ‘dog’ without presupposing that ‘dog’ is true of *all and only* dogs?

This is one of the core problems in the theory of meaning. It is, for example, closely related to Kripke-Wittgenstein meaning-scepticism about dispositional accounts of rule-following.\(^{15}\) Horwich’s point is that if we avoid the relationality constraint, then we do not get the problem.

c) *Deflationism and meaning.*

Horwich insists that the meaning-constituting properties should not be forced into a uniform relational mould. There is a clear sense in which this is similar to one of the tenets of deflationism about truth. Everyone agrees that, for example, the sentence ‘snow is white’ is true iff snow is white; likewise, the sentence ‘grass is green’ is true iff grass is green. The deflationist about truth points out that snow’s being white and grass’s being green have nothing in common. Different facts ensure the truth of different sentences. We should not expect there to be anything substantial in common between the truths of the two sentences.

---

\(^{15}\) *Cf.* Horwich, Ms. p. 21 and ch. 10.
It is similar in the case of meaning. The meaning of 'wombat' is constituted by a property $F$ possessed by the word 'wombat'; the meaning of 'democracy' is constituted by a property $G$ possessed by the word 'democracy', and so on. There are substantial properties that constitute meaning, but the properties $F$ and $G$ need have nothing in common. Different properties constitute the meanings of different words.\(^{16}\)

The use-based account of meaning is not only similar to deflationism about truth, it is also justified by it.\(^{17}\) I will discuss this aspect of the theory in the last sections of the chapter, where the focus is on how Horwich manages to explain facts about meaning without relationality. For now we need to get a better grasp of what the meaning-constituting properties are, and how they are supposed to work.

\textit{d) What are the meaning-constituting properties and how do we identify them?}

The property of a word that constitutes its having this or that meaning is the basic regularity that best explains the overall use of that word. Here are some of Horwich's own examples:

The basic regularity that best explains a speaker's overall use of the word 'and' is (roughly) his tendency to accept 'p and q' iff he accepts both 'p' and 'q'.

The basic regularity underlying our overall use of the word 'red' is (roughly) the disposition to apply 'red to an observed surface when and only when it is clearly red.

The basic regularity governing our total use of the word 'true' is the inclination to accept instances of the schema 'the proposition that p is true iff p'.\(^{18}\)

\(^{16}\)Of course it is also part of deflationism about truth that the property of truth is not substantial at all. When you ascribe truth to a single sentence you do not ascribe to it a substantial property. It is unclear, as yet, what role this aspect of deflationism about truth could play in a theory of meaning. Though, of course, if the deflationary theory of meaning is too much like deflationism about truth in this sense, then it becomes hard to see how it could be reductive in any interesting way (however, see §6 below).


\(^{18}\)Ms. p. 44, \textit{cf.} also p. 123.
Straightoff this look fairly plausible, and since the talk is of tendencies, dispositions and inclinations of use it also looks wholly in line with a *reductive* use theory of meaning. In general, different dispositions for use may be in play for different words, constituting different meanings.\textsuperscript{19}

How do we go about identifying these basic regularities when we cannot rely on the ‘reading off’ method? Here is what Horwich says:

\begin{quote}
[W]e must look for the simplest way of explaining the overall use of a given phonological type. Such a type has a single meaning when there is a single, simple acceptance property, such that all instances of the type may be explained in terms of it.\textsuperscript{20}
\end{quote}

Hence, the *explananda* consist in the entire set of tokens of the word in question. The linguistic theorist can then, using his or her home language, employ ordinary scientific criteria (unification, simplicity, predictive power etc.) in arriving at the *explanans*: some basic regularity which best explains the overall use.\textsuperscript{21}

I think it is a real question to what extent this method presupposes some semantic knowledge. One might think that *holding true* is a semantic notion. Quine, of course, was met with the same kind of objection and resorted to a purely behaviouristic notion of ‘surface assent’.\textsuperscript{22} However, this is an old complaint, and I propose to grant

\textsuperscript{19}Horwich affiliates himself with the dispositional analysis at Ms. p. 215. It complicates matters somewhat that he argues that understanding should be explicated in terms of implicit knowledge (cf. Ms. p. 16-18, 82). How does dispositionalism fit with implicit knowledge? Here is what Horwich says (Ms. p. 81): ‘[W]e can [...] be said, simply in virtue of implementing [the dispositional basic] regularities, to implicitly know what they are’. This looks like an analysis of implicit knowledge in terms of capacities.

\textsuperscript{20}Ms. p. 74.

\textsuperscript{21}For example, the basic regularity in the use of the word ‘one’ is the holding true of the Peano axioms of arithmetic, Ms. p. 123.

\textsuperscript{22}Cf. *e.g.* Quine 1975, p. 91.
Horwich that we can, on purely psychological grounds, get a feasible idea of the holding true or not of native speakers.

There are two main parts in what follows. Firstly, in section 5, I concentrate on the extent to which Horwich’s theory is similar to deflationism about truth. In section 6 I concentrate on the extent to which the use-based theory is justified by deflationism about truth.

5. Basic regularities, functionalism, and permutation.

a) Diverse explananda.

The first worry I have is that the explananda, the tokens of the word in question, may display so much diversity that explanation of them on scientific grounds, in terms of Horwich’s notion of dispositional basic regularities, is put under pressure. Horwich argues that different word types have such diverse meanings that we shouldn’t expect the same type of property to explain tokens of these respective types. My worry is that the tokens of one word type are so diverse that, prima facie at least, we shouldn’t expect the same dispositional property to explain all of them. I’ll suggest a reply to the worry—on behalf of the use-theorist—, pause to observe an obstinate doubt about the redicive ambition of the use-theory, and then use the section as a springboard to a discussion in functionalist terms.

Let us assume, along with Horwich, that we have no knowledge of the language to which the word in question belongs. Instead the task is set like this: we observe all the utterances of tokens of the word, and then, drawing on all the resources of the home
language and the scientific criteria of explanation, we identify the basic regularity of the word.23

Consider the set of all the diverse tokens of the word ‘poisonous’. Say I believe that a certain substance is poisonous. I may manifest this belief in ever so many ways: I may avoid eating it, I may eat small portions of it because I also believe that that will immunise me to it, I may ensure my family doesn’t eat it, but I may also ensure that they do eat it, if my beliefs and desires were somewhat different, and so on. Likewise, I may manifest different beliefs by similar means. Thus, for example, I may token ‘poisonous’ when I believe that something is poisonous and I desire to warn someone; or I may be in the pub telling a lame joke, so I token ‘poisonous’ believing that some substance isn’t poisonous, but desiring to pull someone’s leg; or I may token ‘poisonous’ believing that some substance is poisonous, believing that my interlocutor is a daredevil and desiring him or her to take the poison, and so on. There are also many more wayward examples: I token ‘poisonous’ believing that someone wrote me a nasty letter and desiring to let someone know.

The result is that the set of my tokens of ‘poisonous’ may have very little in common. Different beliefs and desires underlie different tokenings of the word. In the normal run of things my entire belief state and set of desires change all the time. I constantly acquire and lose different beliefs and desires. Due to the holistic nature of belief-desire explanation, any such change could influence the state underlying a particular token of ‘poisonous’. Hence, there is potentially endless diversity in the states underlying my tokens.24

23 What follows is a rehearsal, for tokens of words rather than of sentences, of the ‘problem about unity’ which I discussed in Ch. 6§2. It is a development of an argument in Evans 1985, p. 336f.

24 There is a version of the worry which doesn’t afflict a non-relational theory of meaning. It goes like this: Speakers can correctly token ‘poisonous’ in all sorts of circumstances—in front of cyanide, on the beach, on the moon, and so on. The set of circumstances in which ‘poisonous’ is tokened need have nothing in common. So, the initial question is, if they have nothing in common, how could we hope to find some unifying basic regularity that best explains all the tokens? The deflationary use-theorist is entitled to respond that this is only a problem if we must search for the ill-fated relational
The point is that from such a bottom-up point of view there is no unified set of dispositions underlying all uses of a word. Different tokens may indicate ever so many different concepts. This presents a *prima facie* problem for Horwich’s account of meaning: does the diversity in the *explananda*, e.g., the tokens of ‘poisonous’, preclude true explanation of them in terms of a dispositional basic regularity, *e.g.* a dispositional basic regularity for ‘poisonous’? If we know that there is diversity in the underlying states, then surely we must adjust our explanations accordingly.

A little digression before I suggest a reply to this. I think this worry is close to what Kripke-Wittgenstein meaning-scepticism comes to in the case of someone, like Horwich, who has renounced the ‘reading off’ requirement. Consideration of the example of the token of ‘poisonous’ in the pub is a bit like consideration of the famous ‘quus’ case where 57 quus 68 equals 5.25 Only, the problem according to the scenario I have presented isn’t that no fact about the speaker is such that we could ‘read off’ it whether, on a certain occasion, a token of ‘poisonous’ meant *poisonous* or *joke-poisonous*. Rather, the problem is that the facts about the speaker are so diverse that it is hard to see how any non-semantically identified disposition could unify them. If this is right, then it is all the more incumbent on Horwich to say something about this problem.

Here is how I envisage a reply from Horwich’s scientific point of view. There *must* be some regularity for a meaningful word just isn’t legitimately tokened completely at random. So one should attempt to make divisions among the diverse tokens: assign some to a set of core uses, and others to the periphery, maybe even assign some tokens to a basic regularity for a different word type with a different meaning. Among meaning-constituting properties. A core idea in a use-theory of meaning, which is similar to deflationism about truth, is that it frees us from that very demand. The basic regularity need not advert to relations between ‘poisonous’ and poisons.

25Kripke 1982, Ch. 2.
the core uses, one could presumably formulate a regularity such that, if speakers were disposed in general to use 'poisonous' according to that regularity, then most of all the other tokens of 'poisonous' could be explained.

Such a scientific process is no doubt workable (we can generalise over so many kinds of phenomena), and so there is an answer to the 

**prima facie**

problem about diversity in the explananda. I think that a lot of substantial assumptions, extrapolations and abstractions must be made in order to do this. One would have to draw on all the resources of the home language in order to formulate basic regularities that could predict linguistic behaviour with reasonable success. Think for instance of the fact that the mere frequency of one sort of token (e.g. the preponderance of tokenings of 'poisonous' in front of cyanide) shouldn't always warrant decisions about which tokens belong to the core and which do not; in fact, perhaps the metaphorical uses of 'poisonous' are more frequent than the non-metaphorical uses.26 Once again this puts pressure on the extent to which Horwich's use-based account of meaning is really reductive. Given the diversity in the states underlying the tokens it is hard to see how a translational approach can get off the ground without invoking some semantic knowledge of the language in question.27 Notice that this complaint is not based on the suspicion, voiced against Quine, that assent and dissent are semantic notions.

*b) Functionalism?*

I want to sideline this old complaint for now, and assume, with Horwich, that it is in principle possible—on scientific grounds—to come up with a basic regularity for each of the words of a language.

---

26 Also, the majority of users may in general be wrong about the meanings of some terms, e.g. scientific, terms.
27 Notice that in my solution to the problem, as it arose for UD (in Ch. 6, §2), I do not rely on scientific methods, but merely on what the dispositional facts may be.
Now an interesting question arises: how does this kind of 'top-down' story combine with the above point that from a 'bottom-up' perspective we must say that diverse sets of dispositions explain different tokens? It seems there is a surplus of explanations: each token is explained in the standard way by different sets of beliefs and desires, but all the tokens are also explained by some basic regularity.

I think that the best answer to this question is found by recalling how functionalism in general is especially well-suited to handle such diversity problems. Perhaps, that is, we should conceive of basic regularities as functional roles that take concepts as input and produce tokens of words as output. The problem of diversity in the underlying states can then be dealt with if we allow different sets of beliefs and desires to realise this role in each case. Adopting such a functionalist strategy would, after all, allow a familiar general conception of the use properties that constitute what words mean. Meaning-constituting properties are functional roles identified on the basis of scientific criteria of predictive power, unification and simplicity. This is still in accordance with an account of meaning that is similar to deflationism about truth in as much as the functional roles for different word types need have nothing much in common.

However, this suggestion exerts some pressure on the extent to which the theory is truly deflational. Perhaps Horwich's insistence that there is no reason to expect the basic regularities for different word types to have anything in common is too pessimistic. Maybe we should expect different discourses to be internally homogenous, but still quite different from other discourses? This comment is not meant as a direct criticism of Horwich's position, but rather it indicates that once we have adopted functionalism, we should acknowledge that it comes in various degrees of inflation.28

---

c) Are the explanantia too similar?

I'll end this section by motivating my functionalist reading of Horwich, and then proceed raising a problem for this kind of functionalism. This problem, however, has an independent link to an anti-relational theory of meaning.

I don't think a functionalist reading of Horwich's theory is entirely unreasonable, especially in the light of what he says about the role of *implicit definition* in determining the meanings of theoretical terms. Horwich has argued that a plausible notion of implicit definition of theoretical terms is available, if that notion is based on the use theory of meaning. Thus, a set of postulates involving the word $\alpha$, call it $\#\alpha$, can be the basic regularity for the word. Then the meaning that is being implicitly defined is: the meaning constituted by regarding $\#\alpha$ as true. Think for example of the axioms in Euclidean geometry implicitly defining the meanings of the words 'point' and 'line'. Speakers do not have to be explicitly committed to $\#\alpha$: it is enough if they actually use the word according to the regularity given by $\#\alpha$. Along with Horwich, call this *indirect implicit definition* of $\alpha$.\(^{29}\)

Once the use theory is given this application there is little reason to restrict the notion of indirect implicit definition to *theoretical* terms. It looks more like a general model of meaning determination.\(^{30}\) Different types of words have their meanings constituted in virtue of being used in accordance with different sets of platitudes (or axioms, postulates, commonplaces, commitments etc.). Any type of word could be indirectly implicitly defined by use in this way. This sustains my functionalist reading of Horwich because identifying a core set of platitudes as governing the use of a term is one way of identifying a functional role giving the meaning of the term.

---

\(^{29}\) *Cf.* Ms, Ch. 6; and Horwich 1997b, p. 423-440. The force of Horwich's account of implicit definition is firstly, of course, that it sits neatly with the use-theory of meaning, and secondly, that it accounts for how *a priori* determination of meaning is possible without having to claim that the postulates must also be *a priori* true.\(^{30}\)*Indirect* probably means something like 'implicit', which removes the story somewhat from its scientific connotations.
Going along with this functionalist interpretation of Horwich’s theory, then, there is an issue worth noticing which relates directly to this type of platitude-driven functionalism: platitudes we could regard as true for each of certain tight-knit families of terms are not sufficiently distinguishable, even though the terms have distinct meanings. If this is the case, then, contrary to how the use-theory would have it, regarding them as true would not instigate different basic regularities.

Consider the case of colour terms. Say we set out the platitudes we indirectly regard as true for ‘red’, ‘green’ and ‘blue’. If we Ramsify over these platitudes, by writing all terms in property name form and substituting them for variables, we get indistinguishable platitudes for each of the colour terms. Therefore, regarding #red as true is no different from regarding #green, or #blue as true, even though, in fact, ‘red’, ‘green’ and ‘blue’ do have different meanings. This kind of problem has been called the ‘permutation’ problem for this style of analysis.\(^{31}\) It goes like this:

The meaning of ‘red’ could be determined like this:
S indirectly regards as true: the property red is the x such that \(\exists y \exists z\) ‘red’ applies to all and only surfaces that seem clearly x, and x is more similar to y than it is to z, and so on.

The meaning of ‘blue’ could be determined like this:
S indirectly regards as true: the property blue is the x such that \(\exists y \exists z\) ‘blue’ applies to all and only surfaces that seem clearly x, and x is more similar to y than it is to z, and so on.

And similarly for ‘green’ and the other colours. Clearly, if this is how the platitudes are set out, a speaker would end up regarding the same things as true for different

\(^{31}\)By Smith 1994, 2.11.
colour-terms, thus determining indiscernible meanings for terms we know have different meanings.\footnote{In outline, the story goes this. The meaning of a term could be defined by the platitudes (etc.) which govern behaviour given different inputs of beliefs and desires. The process goes like this: list all the platitudes, then rewrite them so that the term in question is in property name form. Now all the platitudes are conjoined to form a relational predicate 'T' true of all the colour properties 'r', 'g', 'b' (T [r g b...]). Then the property names are removed and replaced by free variables (T[x y z...]). Then, if there actually are colours, it is correct to regard as true that $\exists x \exists y \exists z...T[x y z...]$ & $(x)(y)(z)...$ iff (x=x*, y=y*, z=z*...). The particular colour red is then defined like this: the property of being red is the x such that $\exists y \exists z...T[x y z...]$ & $(x)(y)(z)...$ iff (x=x*, y=y*, z=z*...). Here I have followed Smith 1994 and Miller 1997. I think Horwich's claim would be that, even if T is false, regarding the platitudes as true would suffice to constitute meaning for the terms.} 32

What drives this effect is that colour terms are determined by relations to terms within the same discourse, as well as the notion that there are paradigm cases. Other families of terms that have similar characteristics may display the same kind of 'permutation' problem. These families include, possibly, moral terms and, interestingly, semantic terms themselves.\footnote{Smith 1994, p. 55; A. Miller 1997. Perhaps even the truth-theoretic terms 'true', 'satisfies' and 'refers' display similar characteristics: an (infinite) domain of 'paradigms' and interdefinability. When the platitudes are put together and then stripped of their truth-theoretical content is there then enough left to characterise the defined terms as different truth-theoretical terms? Cf. also my discussion of the permutation problem in Ch. 6 §4.} 33

Assuming Horwich would accept this functionalist reading in the first place, he might reply to the permutation problem like this. It is not necessary to strip away use of $\alpha$ (e.g. use of 'blue' and 'red') and put in variables in the platitudes. That is only necessary if we want to avoid circularity. But circularity is not a concern once we allow scientific identification of the basic regularities. When identifying the regularities we can rely on scientific criteria and the conceptual capacities of the translator, and that may, of course, involve the concepts red and blue.

Accepting such a reply would render it unclear who is doing the 'regarding as true' in indirect implicit definitions: is it the speaker or the scientific linguist? Assume we do not strip away use of the term to be defined. If we then say the speaker is doing the
regarding-as-true, then the regularity cannot be basic: regarding it as true would presuppose grasp of the terms’ correct use. If it is the translator who is doing the regarding-as-true, then it seems unwarranted to claim that the speaker’s use of the terms is governed by the regularities set out in the platitudes.

\textit{d) Permutation and relationality: some general remarks.}

This is quite a serious problem for a deflationary theory of meaning.\textsuperscript{34} This is more so because there seems to be an \textit{independent} link between a deflationary theory of meaning and the permutation problem. Consider what causes permutation problems. Colour terms are implicitly defined in terms of there being paradigm cases and relations to other terms within the same family (as, perhaps, are moral and semantic terms). So perhaps we, in general, get permutation problems because of what we can call the ‘discursive closure’ of some families of implicitly defined terms. The postulates or platitudes governing those families of terms have not got enough in the way of ‘external hooks’ or relations to other discourses or sets of extensions.\textsuperscript{35} Presumably, therefore, the problems could be remedied by providing the platitudes with \textit{relations} to other discourses and sets of extensions. But recall that the starting-point for the deflationary theory of meaning is precisely the \textit{abandonment} of any relationality constraint. This gives rise to the following worry: is the permutation problem the price that must be paid once we abandon relationality?

How can the deflationist respond, if this worry is on the right track? There are various options, but I shall briefly suggest that none are entirely satisfactory. The first option is to object to the way I lump discourse about colour in with the notion of indirect

\textsuperscript{34}In general, counterexamples to a use-based theory of meaning can take this form: $\alpha$ and $\beta$ have similar uses, but different meanings. Thus Twin-Earth cases (with from-the-skin-in similar subjects) could provide counterexamples. Horwich has something to say about these well-worn cases, which looks right (remember that he does allow that some words may have relational use properties). The permutation problems have the same form (same use, different meanings), but they operate within the same language and do not require Twin-Earth scenarios and narrow content.

\textsuperscript{35}\textit{Cf.} Smith, 1994, p. 55f.
implicit definition. Colour discourse is relational (it concerns the disposition to accept 'red' in front of red surfaces), and is therefore different from discourses comprising theoretical terms.36 I have two comments on this objection. Firstly, from the perspective of a platitude-driven functionalism, there does not seem to be any difference between colour discourse and geometry discourse. We list the core platitudes and thereby identify the functional role for the terms in question. Moreover, there is not much difference between Horwich's definition of indirect implicit definition and the definition of the kind of implicit knowledge which underlies understanding of every term.37 Secondly, the objection trades on the admission that the permutation problem would have been serious if colour discourse had not been relational. But this implies that, according to the deflationist theory of meaning, some discourses must be relational. Thus, the objection reinstates a relationality constraint for some discourses. And now it seems we are in familiar empiricist territory. Some discourses must be relational, in the sense that they rely on our discriminatory capacities, and other discourses are either purely theoretical, or indirectly relational.

In any case, this whole objection seems misguided to me. It is true that colour discourse, as envisaged by the platitude functionalist, in fact is relational. But it is not the right kind of relationality. A defender of the permutation argument is happy to acknowledge that there is some kind of relationality involved in discourses afflicted by permutation: viz. the notion that there are paradigm cases. But the permutation problem arises notwithstanding.38 The reason is, I think, that such cases are identified from within the contested discourse (either by the speaker in question or a colour competent observer) and so the paradigm-based relationality is very easily obtained. Therefore, this kind of relationality is not enough to make the permutation argument go away.

36This response would be good if the problem was a Twin-Earth generated counterexample, but the permutation problems do not trade on such devices.
37Compare Ms. p. 82 with p. 136-7.
38Notice that once this is acknowledged I do not need my functionalist reading of Horwich's theory as a crutch any more: the permutation problem can be generated directly on the basis of how Horwich explains the basic regularities for colour discourse.
What is needed to make the problem go away is a much stronger conception of what it is for a discourse to be relational. It is not sufficient for such a constraint to be a platitudinous part of the use-theory itself. One such conception could be the idea that the use-theory must conform to a relationality constraint.\textsuperscript{39} It is extensional relationality, as conceived from a somehow external perspective, that will do the trick, not the 'thin' kind of relationality we have access to from within theory.

Here the deflationist might spot an opening. Perhaps permutation problems can be generated because the paradigm cases are all of this 'thin' disquotational form: 'red' applies to red surfaces, 'blue' applies to blue surfaces etc. If we inflate the functional description of colour discourse, then perhaps there will be enough substance to withstand permutation, without having to insist that the theory conforms to an extensional relationality constraint. Thus the defining platitudes would comprise richly described 'platitudes', not only from colour discourse, but from various other discourses (from aesthetics to science). But of course the deflationist would be ill-advised to go this way: such inflated functional roles can hardly count as basic regularities which the speakers have implicit knowledge of. Furthermore, the whole idea of working with basic regularities, or core platitudes, was to bridle the diversity of beliefs and desires underlying tokens of words. Now it seems that this strategy will only work if there is an equally rich diversity of regularities in play.

In sum, the deflationist seems caught in a three-way battle. Firstly, if the deflationist insists that there is no danger of permutation because permutation-prone discourses are relational, then he or she admits that some meaning-facts must be relational. Secondly, it must be strong extensional relationality: the permutation problem embraces and builds on 'thin' paradigmatic relationality. Thirdly, trying to avoid extensional relationality by inflating the functional roles is self-defeating for the deflationist: it will

\textsuperscript{39}Recall here Lewis' reply to Putnam's "just more theory" argument in his version of a permutation argument: Lewis 1984.
force the deflationist to relinquish any hope of positing implicitly known basic regularities.

6. Explanations of meaning and explanations of content.

a) An account of meaning justified by deflationism about truth.

In §4 I mentioned that Horwich’s use-based account is not only similar to, but also justified by, the deflationary theory of truth. I shall outline how I think this story goes, then I’ll suggest an alternative way to cash out part of Horwich’s explanatory strategy, which doesn’t have to be justified by deflationism about truth.

The argument, in short, is that once the meaning of a word has been specified it should be possible to specify what it is true of. For example, once the meaning of ‘wombat’ is specified, it is trivial that ‘wombat’ is true of all and only wombats. If we are inflationists about the notion ‘true of’, then we must insist on analysing that notion in relational terms. This is because inflationists hold that the fundamental principle of being true of is something like:

\[ \text{Inflated: } (x) (y) (x \text{ is true of } y \text{ iff } Rxy) \]

Where ‘R’ is some substantial relation. According to Horwich, this relationality is inherited in the meaning-constituting property, which then itself must be analysed in relational terms, creating all the problems associated with the ‘reading off’ requirement.

---

All this can be avoided if we accept deflationism about the notion ‘true of’. That simply amounts to accepting schemes like this:

\[
\text{Deflated: } (y) \ (F \text{ is true of } y \text{ iff } Fy)
\]

In this case there is simply no relationality constraint on the ‘true of’-notion, and the meaning-constituting property remains untainted.\(^{41}\)

Since we, according to Horwich’s overall position, have good independent reason to accept deflationism about the truth-theoretical notions, we have good independent reason to believe that meaning-constituting properties are not relational.

\(b\) Explaining aboutness.

But, one might ask, isn’t the relationality constraint imposed for a good reason? Isn’t that the only way we can hope to explain how words are about things? How can we get aboutness without being able to read off what the word is about from its meaning-constituting property? And how can we begin to read off in this way if the property isn’t relational?

Horwich argues that satisfying the relationality constraint is not the only way to achieve aboutness. If we adopt deflationism, and discard the notion of ‘reading off’, then we can motivate another explanatory strategy:

\[\text{[T]here will be no way to read off which meaning is constituted by a given use property. The best we can do, in order to get from one to the other, is to appreciate that some word (say, ‘glub’) has the use property—i.e. to actually}\]

\(^{41}\text{Cf. Ms. p. 65-6, 104-5.}\)
use it in that way; in which case we can deploy that very word to characterize
the constituted meaning (as “x means GLUB”).\(^{42}\)

In other words, deflationists can avoid meaning-theoretical relationality and yet explain
aboutness because they in effect \textit{order} their explanation in two steps:

\textit{Step 1.} the non-relational basic regularity \(U\) constitutes the meaning of the
word type \(\alpha\).

\textit{Step 2.} Once we know that \(\alpha\) has the property \(U\), \textit{then} \(\alpha\)’s extension (what it is
true of or refers to) can be specified by putting oneself in a position to use the
term.

As I see it, one of Horwich’s core claims is that segmenting the explanation of
meaning like this frees it of extensional relationality and thereby obviates the Kripke-
Wittgenstein related ‘reading off’ problems. The idea is that this move is only available
to deflationists about truth.

This is an interesting and attractive idea. The basic notion is that from an external or
objective perspective (embodying extensional relationality) on linguistic practices, it is
easy to see how indeterminacies and multiple interpretations are possible, but from an
ordinary insider perspective questions of indeterminacy and interpretability hardly
arise. As such linguistic insiders we are entitled to disquote in a carefree manner.\(^{43}\)
Horwich rightly sees that insisting on extensional relationality is insisting on an
external or objective perspective, and therefore tries to block that route by drawing
heavily on the services of deflationism about truth. This shows that what makes
Horwich’s explanatory strategy work is step 2 just as much as it is step 1.

\(^{42}\)Ms, p. 61. This not unlike Johnston’s move: ‘a theory of meaning could be at most a statement of
propositions knowledge of which would enable us to come to acquire the practical ability [...] to use
the expressions to assert, command, ask about, etc., various things’, 1988, p. 38.
\(^{43}\)This general idea is canvassed by Blackburn 1998, p. 178-9. Note that this is akin to how I
conducted by defence against the threat to first-person authority in Ch. 7.
c) Meaning and content.

I want to briefly direct attention to an alternative use-based account of expression type meaning which is like Horwich's in one important respect, viz. that the use-properties it posits need not be constrained by extensional relationality between expressions and the objects the expressions are about. Nevertheless, this theory could be relational in another respect. The aim of this is to suggest that a use-based theory of meaning need not be justified by deflationism about truth, even though it relies on an explanatory strategy somewhat similar to Horwich's.

The kind of theory I have in mind is a two-stage theory which has it that the truth-theoretical properties of linguistic items are derived from the more basic semantic properties of mental states.44 This is a fairly common functionalist idea, expounded for instance by Harman:

1. The meanings of linguistic expressions are determined by the contents of the concepts and thoughts they can be used to express.
2. The contents of concepts and thoughts are determined by their functional role in a person's psychology.45

But we also find it in a Lewis-style theory, which has whole thoughts, not concepts, as basic:

A language L is a function that assigns truth conditions to certain verbal expressions, called the sentences of L. [...] We [...] use L iff, by convention, we are truthful and trusting in L. That [...] is how use determines meaning.46

---

44For an overview see Loewer 1997.
According to such theories, the explanation of expression type meaning has as its explanatory starting point already contentful mental states. The explanation itself goes like this: the expression type $\alpha$ means $F$ because it is used to express (or indicate) concepts and thoughts with the content $F$.

The property of use alluded to in this explanation of word type meaning need not be relational, at least it need not relate expressions directly with their extensions. Thus the account can cash out one core aspect of Horwich's strategy: it renounces the extensional relationality constraint for the meaning-constituting property. Thus it could be in keeping with step 1 in Horwich's explanatory strategy.

But how would such a theory cash out step 2 in Horwich's explanatory strategy? Simply like this: the extension of the word can be specified by using the word according to the use-property, *given how mental content is determined*.

But notice that this kind of use-based theory could very well co-exist with a relational account of mental content, e.g. in the form of a causally based functionalism. The combined account of meaning and content would therefore, by Horwich's lights, be inflationist about truth: we would have to explain what words are true of in relational terms. But this is in spite of the fact that the meaning-constituting use property itself could be non-relational. The offending relationality resides in the content-constituting property, not in the distinct meaning-constituting property.\(^{47}\) So, a non-relational use-theory of meaning need not be justified by non-relationism and deflationism about truth.\(^{48}\)

\(^{47}\)Why should the properties be considered 'distinct'? i) because linguistic behaviour is explained in terms of meaning and mental states; ii) because relationality in one property does not have to be inherited in the other property. Compare with this case: according to the deflationist about meaning, the truth-predicate 'true' doesn't inflate just because the following holds: 'p' is true iff p and 'p' means p, where "'p' means p" is substantial.

\(^{48}\)Deflationism about truth isn't sufficient for a deflationary theory of meaning either, as Ramsey has shown (cf. Blackburn 1998, p. 165ff).
By directing attention to two-stage theories I am not suggesting that this is how Horwich should have done it and nor am I suggesting that this is really how Horwich’s theory is (though see below). The point is rather to argue that the properties constituting meaning and content can be distinct, and that therefore relationality in one need not be inherited in the other. One can deal with meaning without taking a stance on truth and deflationism. This suggests that there is yet another way to be a ‘minimalist’ about meaning: one can argue that there is really only a rather minimal, but distinct, functional role to play for the concept of meaning, and that an adequate account of the concept must reflect this role while leaving room for other concepts—such as mental content, truth and reference—to play their correlated functional roles alongside it. Call this the doctrine of ‘functional minimalism’.49

d) Two objections and a final worry.

Here are two immediate objections to this kind of two-stage approach. One, why should we be interested in a meaning-constituting property which derives all its truth-theoretic characteristics from some other more basic content-property? And, two, what has this to do with Horwich’s account of meaning?

I think that it is a very interesting philosophical question how words come to possess meanings, and what role mental content plays in this process. Even if we do presuppose mental content, we might still wish for substantial accounts of first-person authority about meaning, understanding, and transmission of knowledge in communication. In this I agree with Paul Boghossian, even though he is a non-reductionist Platonist about mental content:

49Of course, this view is what motivated all my previous chapters where I set out my analysis of the concept of meaning and gave my account of use-dependence.
Anti-reductionism [about mental content] is not only consistent with, but positively invites, a theory of the relation between thought and language. How do public language symbols come to acquire meaning and what role does thought play in that process?50

So there is reason to believe that the linguistic component of two-stage theories is interesting in its own right. And there are philosophers, apart from Harman and Lewis, who in quite different ways try to give two-stage accounts. Think of Grice and Fodor, but also of Evans and other proponents of thought-based accounts of meaning.51

Is there any reason why Horwich’s account of meaning should be associated with a two-stage account? I think so, in spite of Horwich’s insistence to the contrary.52 Recall I suggested that the basic regularities are best thought of as functional roles that take concepts as input and tokens of words as output. I suggested that these roles can be realised by different sets of beliefs and desires on each occasion. If so, there seems to be a clear sense in which the notion of basic regularities itself *presupposes* the availability of already contentful mental states. Functional roles let us identify realisers by description, but do not have anything to say about how the realisers are constituted.

Assuming this reading is adequate I think we can begin to appreciate what makes Horwich’s explanatory strategy work: in step 2, above, we are asked to *put ourselves in a position to use the term in question* to specify its extension. But, as I have just suggested, putting the basic regularity to use presupposes the availability of contentful mental states. Therefore, the locution ‘putting oneself in a position to use the term’ could very well conceal a vital—and potentially relational—distinct component of the theory. And, of course, if it doesn’t have such a component, then the speaker only

50Boghossian 1989, p. 549.
51See *e.g.* Evans 1982, p. 106.
52Cf. Ms. p. 50, Ch. 4.
relies on the basic regularities in question in which case the permutation problem is brought into sharp focus. What makes it the case that, when a speaker puts him- or herself in a position to use—to disquote—the term, he or she employs the basic regularity for, for example, 'red' and not for 'green'? If the permutation problem sticks, then nothing ensures this in the afflicted cases, and the speaker will disquote to no avail.

I think this points to a worry about accounts, such as Horwich's, which try to overcome the Kripke-Wittgenstein problems by allowing the agent to disquote from within a first-person perspective of the language. How much must this capacity presuppose in order to ensure that the same meaning-constituting property is associated with utterances of the same type of word? My own answer is that we have to presuppose a lot, viz. a notion of mental content.\(^{53}\) My contention, then, is this: it is possible, and probably desirable (in the light of the permutation problem), to factor out the deflational element in Horwich's use-based account of meaning.

7. Summary remarks.

In this chapter I have discussed minimalism, or deflationism, about meaning as an alternative to adopting use-dependence in the face of the failure of the 'metaphysically infected' primitivism. It quickly became apparent that minimalism needs a substantial notion of use to get off the ground, and that this of course implies that it doesn't really constitute a genuine challenge to use-dependence, it is therefore questionable in what sense such a theory could be 'minimal'.

\(^{53}\text{Cf. Ch. 6§4.b.iv.}\)
I therefore turned my attention to Horwich's use-based account of meaning which is similar to, and justified by, deflationism about truth. I raised the following questions: i) how does it handle the diverse mental states underlying tokens of the same type of word, without presupposing semantic knowledge? ii) if it adopts functionalism, how deflationary should we then expect it to be, and how does it then handle the so-called permutation problem? iii) is the permutation problem produced by renouncing the relationality constraint on the meaning-constituting properties? iv) could there be a non-relational account of meaning which is not justified by a deflationary theory of truth? v) how much must be presupposed in order to make an explanatory strategy like Horwich's work?

In my answers to these questions, on Horwich's behalf, it turns out that use-based deflationism could well adopt a general type of functionalism which will allow it to be similar to deflationism in the sense that different words may have input-output profiles, but which need not be justified by deflationism about truth. In my own account of the concept of meaning, this is more or less what I have done: my functionalism is given in purely dispositional terms, and it builds on, and presupposes, a notion of contentful mental states.
We can now see that meaning is use-dependent. That is, that the concept of the relation expressed by ‘φ means p’, where ‘φ’ is a declarative sentence type and ‘p’ is a proposition, is the concept of a use-dependent property.

This result is arrived at in three main steps. Firstly, I offered an analysis of the concept of meaning. The *analysans* consists in five platitudes, or core beliefs, about the concept. The five platitudes concern the possibility of communication and expression of thought, the notion that meaning is use, the notion that sometimes meanings can explain uses of sentences, the possibility of error and the robustness of disagreement about meaning, and, lastly, the notion of first-person authority about meaning. This result should be fairly uncontroversial, since I motivated the platitudes such that they come out as neutral in various philosophical debates.

Nevertheless there are three tensions among the platitudes. The platitude that meaning is use appears in all three tensions, *viz.* with the platitudes that meaning can explain use, that error and robustness is possible, and that we have privileged first-person authority about meaning. These tensions are quite serious. If they cannot be resolved then we have little reason to regard our ordinary concept of meaning as coherent at all.
That brought me to the second step of my argument. What should we do in response to the tensions? Should we reject one or more of the platitudes, or search for a property which could underwrite all five platitudes such that the tensions are resolved?

The fact that the meaning is use platitude appears in all three tensions explains, I believe, why so many people are suspicious about the slogan “meaning is use”. But we should not respond to the tensions by claiming that the meaning is use platitude is the culprit. I argued that rejection of that platitude leads to trouble with satisfying the remaining four platitudes. The price of endorsing use-independent meaning-facts is too high.

Instead, I suggested that the five platitudes are jointly underwritten by a property of use-dependence. My assumption was that the meaning is use platitude is involved in all three tensions because it is the central platitude, and that we would be able to resolve the tensions by adjusting the other platitudes according to its central place in the analysis.

The property of use I arrived at is use-dependence. This property belongs with the larger family of response-dependent properties. I discussed the broader notion in some detail in order to find a version of it that would suit the intuition behind the meaning is use platitude. I argued that the order of determination is pivotal to an interesting notion of response-dependence, and that an order of determination can be established by satisfying four constraints originally put forward by Wright. However, I suggested an alternative version of one of the constraints. Furthermore, I argued that use-dependence must be dispositional. We know that there are dispositional properties, so this shouldn’t be too uncontroversial, even though there are counterexamples to more naive analyses of the notion of dispositions.
Having set out the idea that the concept of meaning is the concept of a use-dependent property, I moved to the third step of my argument. Here I showed how the notion of use-dependence allows us to resolve the three tensions while respecting the phenomenological motivations for the platitudes. The dispositional character of use-dependence is what drives the solutions in all three cases.

The first tension is resolved by noting how operating with dispositions allows us to distinguish possession of dispositions from manifestation of dispositions. I argued that this fits the phenomenology behind the platitude that meaning can explain use, and demonstrated that this sits well with a response to Johnston’s so-called Missing Explanation Argument.

The second tension is resolved by taking full advantage of the sophisticated notion of a disposition. There is a forceful battery of arguments against dispositionalism in an account of meaning, but I find none of these arguments convincing. They either i) do not acknowledge the dispositional character of dispositionalism at all; or ii) work only against naive analyses of the notion; or iii) amount to an across the board refutation of the very notion of a disposition, and thus rob us of normal dispositional concepts such as fragility or water-solubility. I finally showed what it special about meaning-constituting dispositions, as opposed to other dispositions, such as fragility.

The third tension is also resolved by relying on the dispositional character of use-dependence. I showed that we have no reason to think that dispositionalism entails behaviourism in any sense that may be detrimental to first-person authority about meaning. And, if meanings are dispositional, then we have to hand a good notion of wherein such authority consists, viz. in an a priori presumption that speakers get their meanings right. Use-dependence is also consistent with a fairly unsurprising type of social externalism, which, moreover, incorporates the notion of a priori presumed
authority. By employing Lewis’ notion of elusive knowledge I finally mooted a more
general epistemology for knowledge of languages.

Lastly, I discussed a major rival to use-dependence, viz. Horwich’s notion of
deflationary use-based meaning. I argued that we can, and perhaps should, factor out
the deflationary element of Horwich’s theory.

It is easy to see why the “meaning is use” slogan is viewed with suspicion in the
debate about meaning. It seems either obscure or platitudinous, and yet it is involved
in some of the major problems in the debate. My suggestion is that we should embrace
its platitudinous character, while acknowledging how it is networked with other
platitudes about meaning. Once we have done this, we can begin to see how the
problems in the debates, as reflected in the tensions among these platitudes, may be
resolved. There is only a small price to be paid for this, viz. that some of the platitudes
must be cleansed of philosophical misconceptions (e.g. about semantic primitivism,
the causal element in meaning, and the nature of first-person authority). There is a
bigger sacrifice, though, because the concept of mental content is left un-analysed.
The doctrine of use-dependent meaning has nothing to say about what constitutes
content as such. Still, the doctrine allows us to give a reductive analysis of where, in
the landscape of semantic notions, sentence meaning should be located.


—— 1993: ‘Does an Inferential Role Semantics Rest Upon a Mistake?’, *Mind and Language*, vol. 8, no. 1.


— 1987: Ursprünge der analytischen Philosophie, Suhrkamp, Frankfurt/M.


—— 1984a: *Bemerkungen über die Grundlagen der Mathematik*, Suhrkamp, Frankfurt/M.

—— 1984b: *Letzte Schrifte über die Philosophie der Psychologie*, Suhrkamp, Frankfurt/M.