THE PATTERN IN THE WEAVE: AN ACCOUNT OF WITTGENSTEIN'S REMARKS ON MEANING-AS-USE AND OF THEIR RELATION TO SOCIAL PSYCHOLOGY

Andrew McKinlay

A Thesis Submitted for the Degree of PhD at the University of St Andrews

1989

Full metadata for this item is available in Research@StAndrews:FullText at:
http://research-repository.st-andrews.ac.uk/

Please use this identifier to cite or link to this item:
http://hdl.handle.net/10023/3715

This item is protected by original copyright

This item is licensed under a Creative Commons License
THE PATTERN IN THE WEAVE: AN ACCOUNT OF WITTGENSTEIN'S REMARKS ON MEANING-AS-USE AND OF THEIR RELATION TO SOCIAL PSYCHOLOGY

ANDREW McKINLAY

Ph.D.

ABSTRACT

The thesis begins with a criticism of the 'theory of meaning' approach to the philosophy of language. It then goes on to establish an account of language understanding in terms of Wittgenstein's comments on rule-following and meaning as use. This characterization is extended to aspects of the philosophy of social science. Inferences are then drawn, on the basis of this extension, as to the overall framework within which empirical social studies should be located. A critical assessment is offered of a specific social scientific theory which is, in some ways, typical of empirical social research. This criticism is followed by a formulation of an alternative approach to empirical questions in the social sciences. The alternative approach is depicted as more sympathetic to the general perspective on social scientific theories outlined earlier.
DECLARATION

I, Andrew McKinlay, hereby certify that this thesis, which is approximately 92,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

date: 18th. October, 1988

signature of candidate:

I was admitted as a research student under Ordinance No. 12 in October, 1985 and as a candidate for the degree of Ph.D. in October, 1985; the higher study for which this is a record was carried out in the University of St. Andrews between 1985 and 1988.

date: 18th. October, 1988

signature of candidate:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D. in the University of St. Andrews and that the candidate is qualified to submit this thesis for that degree.

date: 18th. October, 1988

signature of supervisor
COPYRIGHT DECLARATION

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

date: 18th. October, 1988

signature of candidate:
## CONTENTS

### Chapter 1

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>Realism and Anti-realism</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>Anti-realism, Truth and Defeasibility</td>
</tr>
<tr>
<td>4</td>
<td>34</td>
<td>Realism, Meaning and Abilities</td>
</tr>
<tr>
<td>5</td>
<td>46</td>
<td>Conclusion</td>
</tr>
</tbody>
</table>

### Chapter 2

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>57</td>
<td>Kripke: Accentuate the Communal, Eliminate the Regulative</td>
</tr>
<tr>
<td>3</td>
<td>66</td>
<td>CT/RT Compatibility I: a 'Limited Communal Mistake' Account</td>
</tr>
<tr>
<td>4</td>
<td>82</td>
<td>CT/RT Compatibility II: An 'Internal Relations' Account</td>
</tr>
<tr>
<td>5</td>
<td>91</td>
<td>CT/RT Compatibility III: The RFA Account</td>
</tr>
<tr>
<td>6</td>
<td>108</td>
<td>Conclusion</td>
</tr>
</tbody>
</table>

### Chapter 3

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>111</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>117</td>
<td>Two Accounts of Assertion/Background Independence</td>
</tr>
<tr>
<td>3</td>
<td>131</td>
<td>Criteria and Theory</td>
</tr>
<tr>
<td>4</td>
<td>139</td>
<td>Rules and Social Science</td>
</tr>
<tr>
<td>5</td>
<td>155</td>
<td>Social Science and Theory of Meaning</td>
</tr>
<tr>
<td>6</td>
<td>160</td>
<td>Conclusion</td>
</tr>
</tbody>
</table>
Chapter 3: Appendix
Page 164 The Ethnomethodologist/Strong Programme Debate Continued

Chapter 4
§1 Page 172 Introduction
§2 Page 176 Social Representations: the Theory
§3 Page 185 Difficulties with the Theory
§4 Page 199 Social Representations Qua Representations
§5 Page 209 Representations Qua Images
§6 Page 218 Conclusion

Chapter 5
§1 Page 221 Introduction
§2 Page 227 Discourse Analysis and Topic
§3 Page 237 Topic Change
§4 Page 252 Conclusion

Chapter 6
§1 Page 257 Introduction
§2 Page 259 Scientists' Conference Talk
§3 Page 270 The Bottom-Up Proponents Speak
§4 Page 282 Conclusion

Conclusion Page 288
Notes Page 292
References Page 314
ACKNOWLEDGEMENTS

I would like to thank all those who helped in the preparation of this work. In particular I wish to express my gratitude to my two supervisors: Johnathan Potter and Roger Squires. In the past five years I have been lucky enough to benefit from Johnathan's drive, and intellectual enthusiasm and also from his unfailing kindness and generosity. I am deeply grateful for the patience and encouragement he has shown me. I have learned much about psychology from him. I hope he will regard this work as proof that his painstaking efforts in my regard have not been totally in vain. I regard myself as being equally fortunate in having been philosophically coached by Roger Squires. His kindly humour has always made the sharp edge of his penetrating insight easy to bear. Both Roger and Johnathan have been invaluable guides in the preparation of this work.

I have learnt a lot about social psychology from Margaret Wetherell, and would like to thank her for her help and support.

Earlier drafts of all six of the chapters which appear here have been given as talks, in one form or another, to a range of conferences and meetings. I would like to thank all of those who made comments at these meetings on those earlier drafts.

Finally, I would like to thank Anne, Dill and Rowan for putting up with me during the final, hectic months.
evidenced by \( C \rightarrow E \) which is how the world regularly works. In normalizing, by contrast, \( E \) is shown to be something that had to happen if the world is to continue to satisfy \( C \rightarrow E \).

The difference here is supposed to be captured by the idea that in the latter case, \( E \) is something that has to happen if the world is to continue to satisfy \( C \rightarrow E \). That is, it is because \( E \) is necessary for the satisfaction of \( C \rightarrow E \) that \( E \) occurs. Now Petit chooses, as an expression of ' \( E \) is necessary for \( C \rightarrow E \)' , the formula: \( (C \rightarrow E) \rightarrow E \).

It is noteworthy that on this formulation, the conditions which fulfil the regularising thought are the same as those which fulfil the normalizing thought. If we take the assumption which forms the background to a normalising explanation to be the assumption that if it is the case that if the principle \( C \rightarrow E \) holds the \( E \) then \( E \) occurs, then on Petit's formulation this can be written as: \( (C \rightarrow E) \rightarrow E \). This will be true except when \( C \) is true and \( E \) false.

Equally, the assumption against the background of which the regularising explanation is offered is: \( C \rightarrow E \). And this, likewise, is true except when \( C \) is true and \( E \) false.

Now the difference between a normalizing explanation and a regularizing explanation, according to Petit, comes in the assumptions against the background of which the explanation is offered. But so far, the preferred versions of the two background assumptions have the same truth conditions: the normalizing assumption is true iff the regularizing assumption is true (the antecedent of the claim ' \( E \) is necessary for the satisfaction of the law \( C \rightarrow E \)', \( (C \rightarrow E) \rightarrow E \), is true only when the antecedent of \( C \rightarrow E \),
PREFACE

Introduction

One of the most important aspects of human life - one of the aspects which separates out our existence as essentially human - is the ability people demonstrate to represent fruitfully the world to themselves and, by means of communication, inform others of the results of such representing. We can impart new knowledge to others, alerting them to features of the world we take to be noteworthy. We can formulate plans and goals, singly and jointly, in order to deal with these worldly phenomena. We can come together in discussion in assessing whether such goals are met. In short, we make sense of, and attach meaning to, the world and our activities within that world. It is this phenomenon of meaning which is the focus of the present study.

In the pages that follow, an attempt is made to provide a philosophical account of meaning and to draw conclusions about how empirical study of meaningful human activities ought to proceed. Any such project must face two competing demands. On the one hand, an author must be familiar with the detail of debate which forms the terrain of the battlefield over which he intends his arguments to range. On the other hand, the author has a responsibility to advance his forces across as wide a front as can be devised: often the key to a problem lies not in the minutiae of debate but in the general terms which are used to set up that debate. It will be clear, for example, from what is to follow that I do not accept
that the provision of a theory of meaning as currently explicated
is an appropriate goal for the philosophy of language. Chapter 1
attempts to provide some of the detailed argument which would be
required to defend such a claim. Anyone immersed in the meaning
theorists' debates will recognize, however, that such
argumentation merely scratches the surface of a complex,
exasperating yet fascinating discussion. Hopefully, however, what
is said in the following pages will demonstrate that there are
viable alternatives to this approach and at the same time show
some of the implications of adopting those alternatives.

The central aim of the thesis is to establish an account of
language understanding in terms of Wittgenstein's comments on
rule-following and meaning as use. This characterization is
extended to aspects of the philosophy of social science.
Inferences are then drawn, on the basis of this extension, as to
the overall framework within which empirical social studies should
be located. A critical assessment is offered of a specific social
scientific theory which is, in some ways, typical of empirical
social research. This criticism is followed by a formulation of an
alternative approach to empirical questions in the social
sciences. The alternative approach is depicted as more sympathetic
to the general perspective on social scientific theories outlined
earlier.
Content by Chapter

1: Several modern accounts of Wittgenstein's thought place it within a general theory of meaning framework. It is argued here that the two most popular theory of meaning approaches - semantic realism and semantic anti-realism (as characterized by Dummett) - face intractable difficulties. It is pointed out that the theory of meaning approach rests upon a general assumption: that understanding of language is explicable as a form of propositional knowledge. The conclusion drawn from the failure of the Realist and Anti-realist theories outlined is that the suitability of this fundamental assumption should be reviewed.

2: An alternative perspective on language understanding is introduced: Wittgenstein's rule-following account of meaning as use. Two central theses are identified in this picture of language use. They are seen to lead to a potential dilemma. The resolution of the dilemma involves abandoning as inappropriate the idea that language understanding can be explained in terms of propositional knowledge. Accordingly, it is concluded that a Wittgensteinian rule-following account of language understanding is feasible in which there is no attempt to provide a substantive indication of that in which grasp of meaning consists for all of language simultaneously.
3: The perspective on rule following thus developed is seen to provide a depiction of grasp of meaning which can be termed 'criterial'. This criterial account is then extended by a consideration of the philosophical background to the social sciences. It is argued that the criterial approach carries implications for the structure of social scientific explanation.

4: The general conclusion drawn from the previous chapter is that social scientific theories must offer a rules-based analysis of the meaning of social actions. It is suggested that this requirement places limitations on the nature of empirical approaches to social research. This is amplified by means of a critical study of one particular social psychological approach: Moscovici's Social Representations theory.

5: It is suggested that considerations as to the possible nature of social explanation demand that social scientists pay attention to subjects' constructive activities. In contrast to the type of theory mentioned in chapter four, a constructive account would emphasise the creative way individuals manipulate linguistic resources to achieve local or contextual goals. It is suggested that recent developments in social psychological theory, based on the analysis of discourse, represent a foundation for such constructive analyses. A demonstration is provided of what a discourse analytic approach can achieve by presenting an
analysis of a series of group discussions which take place among a number of skilled artisans.

6: The empirical approach introduced in the preceding chapter is further developed. It is argued that the social world of subjects can be viewed as a constructed achievement by means of conversational analyses of their inter-communications. A practical example is offered by means of an analysis of the constructive language use of social scientists.

The overall aim is, then, to provide a Wittgensteinian account of meaning which can be of use in the social scientist's task of empirically studying meaningful human activity. Wittgenstein's work is in this respect central to the approach adopted here. I have, however, adopted a policy of avoiding the use of direct quotation from Wittgenstein's own work. Wittgenstein's writings are notoriously difficult to interpret. For this reason it has always seemed to me that sprinkling a text with Wittgensteinian quotations seldom helps advance the argument being pursued. Nevertheless I do hope that the account of meaning offered here will be seen as sympathetic to the general approach developed by Wittgenstein in his later works.
CHAPTER 1

§1 INTRODUCTION

The goal of this chapter is to introduce the philosophical topic of meaning. The chapter begins by setting out a background for the discussion of meaning and then moves on to present (in as sympathetic a fashion as possible) one particular picture of what a philosophical account of meaning ought to look like: the theory of meaning approach. In §2, two competing theories of meaning - 'Anti-realism' and 'Realism' - are discussed and each is seen to face a particular 'touchstone' challenge. The Anti-realist's reply is outlined and criticised in §3 while the Realist's reply is similarly dealt with in §4. The conclusion drawn from these criticisms is that neither theorist can meet his respective challenge adequately. A diagnosis for this state of affairs is offered in the final section.

Setting the Background: Language, Meaning and Thought

If the twentieth century has succeeded in making a distinct contribution to the history of philosophy, then at least part of this success must be attributed to those whose concerns have been broadly characterised as analytic philosophy. Filling out such vague titles is usually a contentious affair, but at least one common strand can be identified. For the analytic philosopher, the study of human thought and of its relation to the world must be cast in terms of the way such thought is expressed. In analytic philosophy, it is taken as a given that to study thought is, to a
greater or lesser extent, to study language and meaning. Of course, as with most areas of philosophy, there is debate within this general framework. Most analytic philosophers maintain, however, not only that thought is inter-linked with language, but that the whole idea of giving an account of thought which is prior to an account of language and meaning is a mistake. The idea here is not that there could not be thoughts independent of language, but that we cannot use a theory of thought antecedently to explain meaning and language use. There are those who disagree with even this restriction – notably Gareth Evans [1] – but for most analytic philosophers, to explain our thoughts and their relation to the world is to explain how we use language to capture and express the meaning of such thoughts.

There are two broad methods of approach. The first characterises the philosopher's task as the provision of a theory of meaning in which our use of language is to be explained by appeal to a single concept around which a theory of meaning can be constructed. The second emphasises the widespread variety of human linguistic activities, with the aim of showing how such diversity of usage, when ignored by the philosopher, can create philosophical confusion. The former approach is often associated with the seminal work of Gottlob Frege [2], while the latter is usually associated with the writings of the later Wittgenstein. [3] Many philosophers attempt to draw from both perspectives, (although Wittgenstein himself regarded his own approach as excluding the idea of a theory of language). Speech act theorists such as John Searle [4], for example, explicitly emphasise the wide-ranging
nature of communicative acts while relying, at the same time, on the idea that particular assertions, commands, warnings and the like can share the same fundamental content. In such a theory, though, the idea of 'propositional content' plays a role broadly similar to that played by the theory of meaning's 'central concept'. [5] In this chapter, the focus of concern will be the school of thought which emphasises the theory of meaning approach.

Theories of Meaning
The goal of a theory of meaning approach is to provide an explanation of meaning by appeal to one unifying concept. At the heart of this form of explanation is the idea that for every utterance in the language there will be associated conditions of correctness which can be made out in terms of that central, unifying concept. By specifying the way such correctness conditions determine proper use of the language, the theory of meaning hopes to provide an adequate philosophical account of how language works. The meaning theorist begins by separating out the sense of an utterance from the force of the utterance; the force of an utterance specifies the kind of act which can be performed by using that utterance, while the sense of an utterance determines the content of that act. Thus 'The door is shut' and 'Is the door shut?' differ in force but share the same sense. [6] The meaning theorist then separates out two associated kinds of correctness condition. The first kind, associated with the idea of the force of an utterance, is the type of correctness condition established by the social conventions which govern each act of utterance: is the background context to the act of utterance
appropriate - does the individual issuing an order, for example, have the requisite authority; is the utterance produced responsibly - was, say, an individual's asserting 'p' socially acceptable? The second kind of correctness condition is used to specify the sense of the utterance in terms of the language theory's central concept.

There are at least three general models which the meaning theorist can choose in explaining how an utterance's correctness conditions determine its sense: atomism, molecularism and holism. For the atomist, specification of correctness conditions must be given at the level of individual linguistic expressions (usually single words). The molecularist accepts that the atoms of language may be individual expressions, but argues that the explanation of the meaning of those expressions can only be given at the molecular level - the level at which such atoms join together to form sentences. The molecularist insists that an understanding of any single linguistic expression can only be given in terms of an understanding of the role that expression plays within the sentences in which it appears - its role in settling what is to count as the whole sentence's correctness conditions. For the holist, a proper explanation of the correctness conditions which determine the sense of an utterance requires an account of some suitably extended fragment of the whole of natural language (or, in extreme cases, an account of the whole of natural language itself).
One currently influential variation on these accounts views language as an interconnecting network. On this view, any given sentence has correctness conditions associated with it and to understand what the sentence means is to understand the conditions under which it is correct to assert it. But the correctness conditions of each sentence are related to those of other sentences in the language. In most cases, the decision as to whether a given utterance of a sentence is correct will depend on whether a variety of other sentence utterances are (or would be taken to be) correct. This 'holistic molecularism' variant on the basic molecularist model is what separates out contemporary theory of meaning from the logical positivists. The logical positivists took the view that a sentence's correctness conditions were specifiable, uniquely for each sentence, in terms of direct observation of the presence or absence of the appropriate conditions. In contemporary theory of meaning, most assessment of correctness involves a degree of deductive inference in which the interconnections between the sentence under review and its neighbours in the linguistic network are held to be relevant in establishing whether that sentence's correctness conditions obtain. [7]

The advantage of adopting the theory of meaning approach is that it allows for a form of recursive specification of language. The theory will take the form of a set of general axioms which capture certain general linguistic abilities (the ability to determine whether such and such a sentence is well formed, for example) and a set of specific theorems which capture specific language
abilities: language abilities which can be thought of as related to discrete items of behaviour. This means that the theorist is able to explain the different patterns of use which issue from commands, warnings, assertions and the like through appeal to their common underlying feature, 'sense', as made out in terms of the central concept. Moreover, because the theory is molecular, it allows for a 'constructive' specification of the way the individual linguistic components which together form a meaningful utterance contribute to the meaning of the utterance. This in turn allows for an explanation, by appeal to semantic structure, of the way the language user is able to display understanding of novel linguistic constructions.

The disadvantage of the theory of meaning approach (from the perspective of some critics) is that it requires, of the theorist, a willingness to criticize language use. It will be quite possible for the theorist to establish that our language relies on a specific set of underlying principles which, when explicitly stated, can be seen to be in some respect less than coherent. In such a case, the theorist would be duty-bound to suggest improvements to our everyday practice. [8] The other major disadvantage of the theory of meaning approach is that it relies on the idea that such a theory is in fact known, at least implicitly, by a language user. This, to some critics, puts too great a burden on the everyday user of language.

So a semantic theory of this kind aims to give an account, for every sentence in the language, of (a) the act performed by
uttering the sentence, through an appeal to the force of the utterance (assertoric, interrogative etc.), and (b) the content of the assertion, question or whatever thereby produced by appeal to the sense of the sentence (made out in terms of the appropriate central concept).

92 REALISM AND ANTI-REALISM

Which central concept?

Once a theory of meaning approach has been adopted, the language theorist must opt for a particular notion as central concept. The range of options includes 'truth condition', 'verification condition' and 'falsification condition'. These can be thought of as different kinds of assertibility condition. That is to say they each, in different ways, represent conditions under which it would be correct to assert a particular sentence. The idea here is that particular utterances of any given sentence are but 'tokens' of one and the same sentence 'type'. By characterising the conditions under which it is appropriate to utter instances of this sentence-type, the theory is supposed to give an account of the meaning of that sentence within the language. On some accounts this means that the actual truth value ascribed to a particular utterance (sentence token) is partly a result of what the language theory says about the truth conditions attached, in virtue of its meaning, to the relevant sentence type. [9]

The debate as to which notion ought to play the role of central concept in a theory of meaning centres on the status of the term
'true' and on the consequences such status has for the explanation of meaning. [10] The Realist language theorist depicts the sense of a sentence as being determined by its truth conditions. These are the conditions under which an assertion made by means of that sentence would be true, where 'truth' is understood according to the following formula: 'each statement is to be regarded as determinately true or not true in a fashion which may transcend our abilities to establish that truth value, whose nature, in turn, is settled by an objective reality which is, likewise, independent of our thought'. The Anti-realist views the sense of a sentence as being determined by those conditions under which it would be recognisably appropriate to affirm it.

On this latter view, the language theorist's central concept is in an important way epistemically constrained. This constraint can take two forms.

(a) The central concept is a condition which is always (at least in principle) verifiable - a verification condition.

(b) The central concept is a condition which, though acting as a warrant for the relevant statement, leaves open the possibility that the statement might be false - a defeasible assertibility condition.

So in addition to Realism, the theory of meaning approach allows for at least two other positions: 'Verificationist Anti-realism' and 'Defeasible Anti-realism'.

'Verificationist Anti-realism'

The verificationist Anti-realist denies that the notion of
evidence-transcendent truth which is involved in the Realist's account ought to be the central concept in a theory of meaning. Originally, this denial was expressed by saying that the central concept in the theory of meaning must be the ability to recognize whatever is counted as verifying a statement. To understand one of the expressions which make up a statement is to understand its role in determining what is to count as a verification of the statement. More recently, this same thesis has come to be expressed in a different way. The contemporary verificationist Anti-realist accepts that the notion of truth ought to be the central concept in a theory of meaning but he retains his hostility to Realism by arguing that the notion of truth involved must be a recognizable, non-transcendent affair. On this view, a statement's truth conditions just are its verification conditions. In both cases, grasp of meaning is explained in terms which are essentially linked to our epistemic capacities for conclusively establishing the truth of a statement. [11] (It should be noted that this simple formula, however expressed, is contentious. The Anti-realist's opponent might want to know, for example, what constitutes 'conclusive' verification and whether this relies on 'in principle' verifiability — we may not actually have verified the statement here and now, but we possess an effective procedure whereby we could (with suitable idealizations) do so if pressed — or in practice verifiability.) [12]

On this 'Verificationist Anti-realism' formulation, it is important to keep clear the distinction between two quite separate capacities. The first capacity is the ability to recognize
whatever counts as verifying the relevant statement - the ability to recognize whatever conclusively establishes the statement as true. The second capacity is the ability to recognize whether the statement is true or false. Now whenever we have an effective procedure for bringing whatever it is that establishes the statement as true within the scope of our cognitive faculties (e.g. having a mechanical procedure for generating a proof), these two capacities will coincide. If I have the capacity to recognize whatever establishes a statement as true, and I possess an effective procedure for bringing that which establishes the statement as true within the scope of my cognitive faculties, then I have the capacity to recognize whether the statement is true or false. However, where I lack such an effective procedure, I will be unable to recognize the truth or falsity of the statement even when I possess the capacity to recognize that which establishes the statement as true.

For the 'Verificationist Anti-realist', there is no commitment to the idea that the individual who understands a statement will always be able actually to verify that statement - he will not always have the capacity to recognize whether the statement is true or false. All that is required is that the individual be capable of recognizing a verification of the statement should one arise and this ability is quite different from the capacity to recognize the statement's truth or falsity. It follows that, for the 'Verificationist Anti-realist', in cases where there can be no verification of a statement (all the relevant evidence may be long gone, for example) we may be unable to determine whether the
statement is true or false. But the 'Verificationist Anti-realism' formula still offers an account of what that statement means - an account which is given in terms of our practical capacities. The meaning of the statement is given in terms of what we know when we understand it. What we know is given in terms of the capacity which such knowledge provides us with. This capacity is the capacity to recognize what counts as verifying the statement - the capacity to recognize what counts as conclusively establishing its truth. What we do not have, in cases where there is no effective verification procedure, is the further capacity to recognize the truth or falsity of that statement.

**Defeasible Anti-realism**

The defeasible Anti-realist likewise denies that the Realist's notion of truth ought to be the central concept. Although he also accepts that the notion of truth is important, for the Anti-Realist, the *central* concept must be *defeasible warrant for an assertion*. The notion of defeasible assertibility condition which is thereby introduced as central concept relies on the idea of defeasible evidence - we may not be able to verify the statement here and now or in the future either, but we are able to determine whether the evidence is strong enough to warrant assertion of the statement. It seems obvious that for some types of statement, there can be no conclusive verification. In such cases, there will be no capacity for recognizing when verification conditions obtain, since there will be no evidential state which guarantees the absolute truth of the relevant sentence. It follows that as
far as recognition is concerned, the notion of defeasible evidence must take over from the idea of verification. [13]

Neither brand of Anti-realist need claim that there is a coherent global form of Anti-realism. The arguments might well have to be fought on an individual basis, within each of a number of topic areas. [14] This does not exclude the possibility, however, of forming a general conception of the nature of the Anti-realist's charge against the Realist (just as a quite general challenge to the Anti-realist, of whatever bent, may be revealing - see later). Of course the point of offering up such a challenge is both positive and negative. If the challenge succeeds, then the Theorist can conclude that the Anti-realist perspective is a necessary part of the explanation of language. If the challenge fails, then the Anti-realist has been able elicit argumentative support for a Realist trend of thought which might otherwise have been held to be mere dogma - realism au folle as much as realism au fonde. [15]

Two 'Touchstone' Challenges
The assertion conditions of both the Realist and the Anti-realist have features to recommend them as the central concept of a theory of meaning and features which weaken their claim to such a role.

The proponent of Realist truth conditions is able to claim a direct link between that which makes a statement true and what the language theory says about the meaning of that statement. Such a link seems desirable as soon as the intuitive appeal of the equivalence thesis (the thesis that for any sentence A, A is
equivalent to 'It is true that A') is noted. [16] In particular, the Realist is able to present an uncluttered view of truth - that truth is determined by the world, irrespective of what anyone thinks or believes - and link this to his account of meaning. The weakness of this approach (in the eyes of its opponents) is that the conception of meaning offered is one which irreparably severs meaning from our practical capacities to use language. In contrast, the strength of the Anti-realist position is that it directly links an account of meaning to our abilities. [19] Its weakness (in the eyes of its opponents) is that it provides too thin a notion of truth.

Clearly, if it could be shown that there exists a position which in some way possessed the strengths of both Realism and Anti-realism without their weaknesses, this would mark a step forward in the understanding of how language works. Such a position would embody a conception of truth with which the Realist could have no complaint, yet would allow for a link between the account given of meaning and human practical capacities which would prove satisfactory to the Anti-realist. Now since there is a variety of ways in which the criticisms outlined in the preceding paragraph can be levelled, and since there is a correspondingly fecund outgrowth of defensive responses, anyone who wished to establish a successful 'compromise' position would have a mammoth task. There are, however, two arguments - one critical of the Realist position and the other critical of the position of his Anti-Realist opponent - which might fairly be regarded as touchstones of the two respective positions in general. If either the Realist or the
Anti-realist was able to provide a conclusive rejoinder in these cases, then he might reasonably be regarded as on the way to formulation of just such a compromise position and would be able to claim, in consequence, an important development in philosophy of language.

The two 'touchstone' arguments around which much of the Realist/Anti-realist debate eddies and flows are: the 'Temporal Responsibility' (or 'Truth Value Links') argument and the 'Manifestation Argument'. The former is offered by the Realist in order to show that the Anti-realist perspective is unacceptable. The latter is offered by the Anti-realist as evidence that the Realist position is untenable. We can outline these arguments as follows.

The 'Temporal Responsibility Argument': a Challenge to the Anti-realist

The Realist complains that the Anti-realist's conception of meaning requires too thin a notion of truth. The latter says that the correct account of the content of a statement is one which gives the statement's *assertibility conditions* - that in virtue of which it is correct to assert that the sentence is true, with the crucial proviso that 'true' must be here understood to be in some way epistemically constrained. What is to count as a statement's being true is not something which could be severed from our abilities to determine truth values. It seems unlikely, however, that we will count as acceptable any account of meaning which abandons the intuitively appealing conception of truth as invoking
'temporal responsibilities' - the idea that if we accept that statement \( S \) is true now, we are committed to the view that somehow our present and future opinions as to the acceptability of temporal variants of \( S \) are already fixed. [18] For it is this notion of 'temporal responsibility' which lies at the heart of our everyday use of tensed statements - for any statement, \( S \), it ought to be the case that if a present-tense utterance of \( S \) is true, then versions of \( S \) cast in different tenses ought, at the appropriate times, to be equally true.

But it is at this point that the Realist mounts his attack. It is just in the nature of things that we discover new evidence and lose evidence we previously possessed. But if the correctness of an assertion is essentially linked to the availability of evidence, then since evidence is the type of thing which we can discover and lose, a statement's correctness could change with evidential circumstances. But a statement is correct when it is true. So if truth is linked to evidence, statement \( S \) made now might be true and yet some past tense version of \( S \), made in the future, may turn out not to be true. For, if a statement's truth depends on evidence, then if there is (suitably strong) evidence available now which evaporates later, then the truth of that statement would, it seems, similarly evaporate. But this is precisely the notion which is prohibited by the idea of 'temporal responsibility'. And so Anti-realism seems antithetical to the very notion which lies at the heart of our everyday usage of tensed statements.
Chapter 1

The 'Manifestation Argument': a Challenge to the Realist

One of the most important criticisms of the Realist position is the argument developed in a number of papers by Dummett [19] which has come to be known as the 'Manifestation Argument'. This begins by raising the following question. If it is correct to express someone's understanding of a sentence by saying that he knows what its truth conditions are, in what, according to the Realist theory of meaning, does such knowledge consist? There must be some answer to this question, if the theory of meaning offered by the Realist is to comply with the appealing notion that an account of our understanding of language is, somehow or other, linked with our use of language - linked, that is, with our practical linguistic capacities. So long as this link between meaning and use is acknowledged, then, the question: 'in what does such knowledge consist?' will always seem appropriate.

One answer to this question is to show how a speaker will be able to manifest to us that he possesses the requisite knowledge. On many occasions, of course, the knowledge manifestation will take the form of explicit verbal formulations of the truth conditions. But, since the theory of meaning is intended to be a general explanation, this appeal to explicit formulations cannot, on pain of circularity, be the whole story. [20] If understanding is to be cast in terms of knowledge, the knowledge possessed must be, on at least some occasions, implicit knowledge. [21] But in what, now, is such implicit knowledge to consist? Presumably the answer must be that it consists in the ability to recognize when the truth conditions obtain. But this is to include cases where, because of
the Realist's use of an evidence-transcendent notion of truth, the
truth conditions for a sentence can obtain undetectably - we
possess no effective decision procedure whereby we can reliably
ascertain the truth value of a given statement (typical examples
here are statements about the remote past, statements about other
minds and quantification over infinite domains). This, however,
rules out the idea of the language user's being able (effectively
[22]) to recognize whether the truth conditions obtain. In such
cases, then, the language user is not able to manifest his knowledge of
the appropriate truth conditions by evincing recognition that the
truth conditions obtain.

But this means that the speaker is able neither to manifest an
explicit knowledge nor to manifest an implicit knowledge of such
truth conditions. And this means that there can be nothing to the
ascription of knowledge of truth conditions to the language user.
Nothing in his actual linguistic practices will distinguish him
from those others who lack the requisite knowledge. And so the
result of opting for a Realist theory of meaning is that the
purported explanation of what it is to understand the meaning of a
sentence severs such understanding from a speaker's actual
linguistic capacities.

The Anti-realist has a range of counter-proposals to make in
attempted rebuttal of the 'Temporal Responsibility' (or 'Truth
Value Link') argument and the same can be said for the Realist in
respect of the Manifestation Argument. These proposals in turn
have attracted ripostes. What both sides agree on, however, is
that if the Anti-realist could explain how his evidentially
constrained assertibility condition could allow for a conception
of truth which met the Truth-value Link challenge, then an
important victory would have been won. Similarly, it is generally
accepted that if the Realist can retain his evidence-transcendent
truth conditions and yet meet Dummett's Manifestation challenge,
then an equally important result would have been achieved. In the
following sections, an assessment is made of the extent to which
two such claims, one made in respect of the Anti-realist and one
made in respect of the Realist, are successful.

§3 ANTI-REALISM, TRUTH AND DEFEASIBILITY

The Truth-value Link Challenge

In essence, the Realist's dismissal of anti-realism stems from the
idea that the Anti-realist's conception of meaning requires too
thin a notion of truth. As outlined in the previous section, one
form which this crucial criticism takes is the charge that the
Anti-realist's dependence on evidence-sensitive assertibility
conditions means that he must abandon the intuitively appealing
idea of 'temporal responsibility'. The Anti-realist's
assertibility condition must be associated with evidence for the
assertion of a given sentence, S. But the evidential status of
temporal variants of S will alter through time. This seems to
imply that S's correctness will alter through time. But S is
correct when it is true. So it seems that the Anti-realist must
accept that the truth value of S will alter through time.
Chapter 1

The Anti-realist either must demonstrate that the conception of truth as essentially related to evidence must be one which is, contrary to appearances, consistent with the idea of 'temporal responsibility' or accept that anti-realism implies a wholesale revision in our ordinary use of natural language. Dummett opts for the former and starts on his task by formulating the desired notion of 'temporal responsibility' in terms of 'truth value links': the truth value of each tensed version of $S$ is related or linked to the truth value of each of the others. In this 'truth value linkage' account, however, the evidence with which talk of 'true now' and 'true then' is made out is always evidence of the present moment. A statement, $S$, is true now if there is sufficient evidence now. A past-tense version of statement $S$ will be true in the future just if there is evidence now. [23] On this view, truth values are essentially 'of the moment', whether they be truth values of statements in the present, past or future tenses. Clearly, on such a view, there can be no suspicion that any of the truth value linkages will break down as a consequence of evidential changes.

But this account of truth is liable to seem implausible to the Realist, if only because the appeal to the notion of truth value links in which evidence is always taken to be evidence 'of the moment' cannot be the whole story. Take, for example, the case in which the evidence for the past tense version of statement $S$ disappears in the future. The picture offered so far allows a conception of how that statement might still be regarded, now, as being true, by appeal to the evidence that is available now. The
difficulty is, of course, that it is also apparent, even now, that in the future the verdict on S will be that it is not true - supposing that it is accepted that in the future the same 'truth is linked to evidence' view is adhered to - and so the truth value link would be lost after all, so long as the Anti-realist sticks to his account of the relationship between truth and evidence. And if the Anti-realist nevertheless insists on retaining the truth value link, this would produce the unacceptable result that, being true in the future, S would (by hypothesis, according to the Anti-realist) be verifiable - contrary to the assumption with which we started out.

Dummett's counter is to reject the idea that such an inconsistency could be drawn. The Anti-realist will claim, he says, that the meaning of '(absolutely) true' will vary across time, in which case there is no inconsistency if he accepts that S is true now and yet also accepts that, in the future, he will say, of some past tense version of S, that it is false. For a statement to be true, according to the Anti-realist, is for there to be some situation, which we can bring ourselves into cognizance of, in virtue of which that statement is true. But the claim 'there is a situation, in virtue of which...' will differ in meaning across time. The range of quantification changes with the world's changes, and this alters the meaning of the claim. So what the anti-Realist presently admits to, when he admits that in the future he will say that the past tense version of S is false, does not contradict what he says now about S. Alternatively, Dummett claims, it might be argued that there is no coherent way for us,
now, to say what will in a year's time be said by means of the
claim that the past tense version of S is false. On either view,
the threatened clash between anti-realism and truth value links is
avoided.

Defeasibility
Whatever the merits of Dummett's claim that changes in the domain
of the existential quantifier engender changes in the meaning of
what is said in claiming 'there is something in virtue of which S
is true', it cannot provide an explanation of the link between
evidence and truth in the most general form of anti-realism. Up
until now, the form of anti-realism considered has been one which
emphasised conclusive verification. In logic, this might amount to
the claim that intuitionists, for example, count nothing as true
that cannot (in principle at least) be proven. For a wide range of
statements, however, a mere extension of a 'truth is proof' view
will not be a complete account of the relationship between truth
and evidence.

There are two reasons for this. First, it is unlikely that the
Anti-realist will quarrel with the contention that in most cases
human assertions rest upon inconclusive, not conclusive grounds -
even if the meaning of such assertions is to be given in terms of
conclusive, rather than inconclusive grounds. Second, the Anti-
realist is likely to admit that in many cases, the idea of
conclusive verification is not available at all. In such cases,
the Anti-realist account would have to be one which emphasised the
connection between accepting 'S is true' and possessing evidence
of some acceptably high quality for S which falls short of conclusive proof — accepting, that is, defeasible evidence. It is this idea which underpins the 'Defeasible anti-realism' position. So it seems likely that the Anti-realist will want to allow for some notion of defeasibility.

On such an account, a statement may be held to be susceptible of evidence strong enough to warrant assertion and yet it be the case that at some point in the future, the very claim made now by utterance of that statement turns out, on the subsequent discovery of awkward evidence, to be defeated. It is, however, a difficult matter to explain just what this notion of possible defeat is supposed to be, because almost any account of possible defeat seems applicable to cases which we would normally regard as indefeasible. For example, suppose defeasibility is made out in terms of mis-perception. This would mean that checking a mathematical proof might be just as defeasible a process as making a contingent judgement. (One solution to this problem might be the claim that for some assertions the only way to raise a doubt in respect of them is to employ the most general forms of scepticism.) Whatever the precise notion of defeasibility in play, however, it is clear that it is likely to rely on precisely the kind of intuitions about truth and the passage of time which are meant to be captured by the idea of 'temporal responsibility'. We have available the idea that an assertion of 'S' can be defeasible, now, only because we accept that what we say about 'S' in the future is deemed to be relevant to what we think about S
now. It is because we may say, in the future, that $S$ is defeated, that we say, now, that $S$ is defeasible.

The Anti-realist framework now has at least three strands: a conception of meaning which conceives of statement content in terms of epistemically constrained assertibility conditions; a commitment, for some kinds of statement, to the defeasibility of such statements; an acceptance of the persistence of the 'temporal responsibilities' invoked by the claim '"$S$' is true'. It is important to the Anti-realist, then, that these three notions be consistent, if he is to continue to meet the truth-value link challenge.

The problem for the Anti-realist, however, is that his use of defeasibility turns out to be incompatible with the defence provided by Dummett in response to the truth value link challenge. On Dummett's account, the notion of 'temporal responsibility' expressed by appeal to the truth value links conception is retained only by denial of the possibility that some past tense version of $S$ uttered in the future may contradict an utterance made now of $S$ in the present tense. According to Dummett, the Anti-realist can retain truth value links because the meaning of 'is true' changes through time or because there are no linguistic resources available to say now what will be said in the future through the use of 'is true'. It follows from this, Dummett concludes, that even though evidential circumstances change, and truth is intimately linked to evidence, inconsistencies between utterances made now and in the future cannot arise. But to say
that S is defeasible is to say that the utterance of S, now, will be said to be defeated if a suitable past tense version of S is shown, in the future, to be not true. On the picture of anti-realism offered by Dummett, on the other hand, either the verdict 'not true' issued in the future is not inconsistent with the verdict 'true' issued now, or we cannot, now, even say what will be said, in the future, in issuing the 'not true' verdict. In which case, it is not clear why there should be any sense of 'defeat' attachable to the present utterance of S in virtue of what might be decided about a past tense version of S in the future.

So Dummett's manoeuvre only succeeds in retaining the desired 'temporal responsibility' at the expense of ruling out defeasibility. If the future lack of truth of the relevant past tense version of S is not inconsistent with the present truth of 'S' uttered now (because the meaning of 'true' changes) then the claim that S is true, made now, cannot be defeated by subsequent evidential discoveries against S in the future. And so the claim that S might be defeated by such future evidence is false. But the claim that S might be so defeated just is the claim that S is defeasible. And so Dummett's own account rules out the defeasibility of S. If, alternatively, we have not the linguistic resources to say, now, what will be said in the future through use of 'is true', then we have not got the resources required in order to so much as specify what 'defeasible' is supposed to mean.
Truth value links: an alternative strategy

So the Anti-realist's appeal to changes in meaning of the truth predicate across time as a form of accommodation with the truth value links is not an acceptable position, as long as there is also to be room for the idea of defeasibility. One alternative version of the truth value links, offered by Crispin Wright [24], accepts that the meaning of the truth predicate stays the same across time, but denies that the proper expression of the truth value links demands differently tensed truth predicates. Does this allow for the re-introduction of defeasibility to the Anti-realist's account of meaning?

The problem for the original Anti-realist formulation is that on the conception of truth offered, truth value seems to be as changeable as verifiability status. And this seems to conflict with the idea of 'temporal responsibility' expressed via the truth value links. This latter notion, it will be remembered, was represented by the idea of a relationship between two sentences: one expressing the current truth value of some sentence, S, and the other expressing the truth value which ought to be held in the future (or ought to have been held in the past) by some suitably re-tensed version of S. An example [25] of such a truth value link would be:

(I) (\forall t) [Fut:S is true at t, iff (\exists t_2)(S will be true at t_2)]

where 'Fut:S' represents a version of S cast in the future tense so that Fut:S represents, say, 'Andy will be at his desk at t_2' and where S represents 'Andy is at his desk at t_2'. Wright's solution to the problem of how to allow for truth value links in
Chapter 1

an Anti-realist framework is to offer re-formulations of the truth value links which are intended to avoid the contradictions which forced Dummett to appeal to his notion of changes in the meaning of the truth predicate across time.

In essence, Wright's approach is to capture the idea of passage of time by tensing a predicate other than the truth predicate. His replacement for (I), for example, would be:

(I*) $\forall t_1 \forall t_2$ [What $\text{Fut}:S$ expresses at $t_1$ is true iff $(\exists t_2)\text{ (what } S \text{ will express at } t_2 \text{ is true)}$]

The importance of this change can be seen by considering the type of case mentioned earlier - that in which evidence for $S$ vanishes with the passage of time.

Suppose, for example, that $S$ is 'Andy is at his desk at $t_1$', and $\text{Past}:S$ is 'Andy was at his desk at $t_1$'. Suppose $S$ is asserted. According to the re-worked truth value link formulation, the Anti-realist is able to affirm the truth value link so long as he can also assert the following:

What $\text{Past}:S$ will express at $t_2$ is true (at $t_1$).

But because the truth predicate is in the present tense, the truth of this statement will, anti-realistically, depend on the evidence which is available at $t_1$, not $t_2$. So even though the Anti-realist views truth as essentially related to verification, he is able, on this formulation, to retain the truth value links. Of course, by hypothesis, the relevant evidence for $S$ will still have disappeared by $t_2$. And so, anti-realistically, what $\text{Past}:S$ will express at $t_2$ will not be true (at $t_2$). But this can be the case even when what $\text{Past}:S$ will express at $t_2$ is true at $t_1$. And it is
only the latter, couched solely in terms of present-tense truth predicates, which is required for expressing truth value linkage. In other words, Wright hopes to retain the notion of 'temporal responsibility' which lies at the centre of our everyday use of tensed expressions by a suitable re-expression of the truth value links. What cannot be retained, of course, is the notion that truth is somehow essentially timeless - the idea that if S is ever true, it is always true. But this is a casualty the Anti-realist is willing to accept.

What is less likely, however, is that the Anti-realist is also willing to accept a further casualty: loss of the notion of defeasibility. This brings matters back to the question raised several paragraphs ago as to whether the attenuated sense of 'temporal responsibility' offered by means of the re-formulated truth value links is strong enough to allow for an adequate account of defeasibility. For the Anti-realist, the claim that S is defeated must be understood constructively: we must be in a position to assert that S is wrong if S is said to be defeated. So if S is defeasible, then the Anti-realist will say that even though S is warranted, there may be evidence to suggest that S is not true. According to the notion of defeasibility outlined earlier, we make out this idea by saying that S may be warranted at t, and yet, subsequently, further evidence may be gathered at t₂ which undermines it. But it seems certain that this notion of defeasibility is not one which is open to the Anti-realist who opts for Wright's re-formulation of the truth value links.
We can separate out two claims, on Wright's account:

(i) 'what S expresses at t₁ is true (now, at t₁)'

(ii) 'what Past:S will express at t₂ will be true (in the future, at t₂)'.

Wright's central assumption is that claim (i) is truth value linked to the claim that what Past:S will express at t₂ is true at t, but not to claim (ii). It follows that claim (i) is consistent with ~(ii), so long as the timelessness of truth is abandoned. In respect of defeasibility, it is clear, since (i) is a mere re-formulation of S, that if the claim made by means of S is defeasible then claim (i) is defeasible. Now on the account of defeasibility offered earlier a statement, S, is defeasible if it may be warranted and yet open to defeat by the subsequent turning up of awkward evidence. But if ~(ii) is true then (anti-realistically) such awkward evidence will subsequently turn up. However, (i) is consistent with ~(ii), so (i) must also be consistent with the subsequent turning up of such awkward evidence. But if (i) is consistent with the subsequent turning up of such awkward evidence, then S is consistent with the subsequent turning up of such awkward evidence. But if S is consistent with the subsequent turning up of such awkward evidence then S cannot be open to defeat by the turning up of such evidence, so S cannot be defeasible. So as a result of Wright's central assumption, S cannot be defeasible.

The upshot, then, is that the adaptation of truth value links offered by Wright as a means of accommodating them to the anti-realism programme would be inconsistent with the claim that the
Anti-realist can make use of the idea of defeasibility if 'defeasibility' is understood in terms of the account offered earlier. The conclusion which must be drawn, therefore, is that the formulation of defeasibility outlined earlier must be rejected by the Anti-realists. But for at least some S's the Anti-realist must leave scope for some such notion, it was suggested. So he must, it seems, offer a revised account of defeasibility.

Revising 'Defeasibility'

The foregoing relies on an understanding of defeasibility which emphasises the role of subsequent evidence. This leaves open the possibility of adopting a conception of 'defeasibility' which relies solely on present evidence. Now it must be admitted that much of the textual evidence runs against such a re-formulation. Dummett, for example describes conclusive verification thus:

"Verification is conclusive if, in its presence, we could not conceive of subsequently encountering stronger counter-evidence." [26]

Wright, in a similar vein, describes verification in this way:

"S is capable of actual verification if and only if there is some investigative procedure, I, such that... subsequent grounds sufficient to call S into question would, if we are rational, have to be allowed to call into question simultaneously the truth of at least one member of the original I class..." [27, my underlining]

What both of these formulations imply is that a statement which, though warranted, falls short of verification, is defeasible with respect to subsequent evidence. However, if the preceding arguments carry weight, then this appeal to subsequent evidence in making out 'defeasibility' is inconsistent with other aspects of the Anti-realist programme. This suggests that it is important
that the Anti-realist attempt to re-work the defeasibility notion in terms of present evidence. Unfortunately, however, it is not clear that there is scope for such a treatment.

If defeasibility is made out solely by appeal to present evidence, then when we say of some S which is currently supported by all available evidence that although 'S' is warranted, S might not be true, we must nevertheless mean 'not true' with respect to present evidence. Now so long as 'not true' is, in general, construed in terms of evidence, there will be nothing about such a truth-transcending-warrant claim to which the Anti-realist will object. For the Anti-realist nowhere denies that with many statements we may turn out to be fallible in our recognition of when they are correct. The subsequent gathering of defeating evidence is just such a case. But in the present instance, in which appeal to such subsequent evidence is ruled out, it is less clear that the Anti-realist can explain how it is that truth may transcend present warrant.

By hypothesis, all available evidence supports S. So the claim that S is defeasible must be the claim that there may exist unavailable present evidence that S is defeated. But this now looks uncomfortably Realist in tone. In some cases, naturally, we can make out the idea of currently unavailable evidence by contrasting, say, my cognitive position with that of some other individual who is, in the relevant epistemic respects, in a superior position. But this still leaves those cases in which, in general, all of the evidence currently available to anyone
whatssoever supports S and yet, intuitively, S is defeasible. (This is, presumably, the status of many scientific hypotheses, for example.) The difficulty is that in such completely general cases, the Anti-realist's use of unavailable present evidence in making out the defeasibility of S seems to rely on precisely the kind of appeal to matters as they are, independent of our cognitive capacities, which he criticizes in the Realist. The question is, then, whether the Anti-realist can give an account of the appeal to unavailable present evidence which would not lapse into such realistic thinking.

Of course, the Anti-realist's solution might be that, on Wright's re-formulation of the truth conditions, we can differentiate two senses of 'defeasibility' by differentiating two senses of 'not true'. Just as there are two different ways to express the truth of S, he may say, so there are two different ways to express S's lack of truth:

(i)' 'what S expresses at t, is not true (now, at t,)' 
(ii)' 'what Past:S will express at t2 will be not true (in the future, at t2)'.

To say that S, is currently defeasible is, on this view, to say that S is warranted but that (i)', rather than (ii)', cannot be ruled out. But now the original difficulty merely re-asserts itself. It is true that there will be a range of cases of S being warranted in which there will be evidence at t, which supports S and evidence at t, which supports (i)'. In such cases, the claim that S is warranted though possibly not true will be correct if the balance of evidence favours S, and the verdict as to when such decisions are correct or incorrect will be a matter settled by
some form of convention. However there will also be a wide range of cases in which S is warranted and there is no evidence available at t, to suggest that (i)' is true. But the claim that (i)' might be true just is, for the Anti-realist, the claim that there might be evidence at t, which demonstrates that (i)' is true. So the claim that (i)' may be true must be the claim that there may be present evidence, presently unavailable, which demonstrates that (i)' is true. The problem is, though, that this is precisely the kind of appeal to unavailable evidence which, when used to motivate the claim that S might possibly not be true, raised the worry that the Anti-realist was relying on an implicitly Realist conception.

In such cases, it seems as though the Anti-realist must admit that there is no sense in which (i)' can be true. But this would be an uncomfortable position for two reasons. First, it seems to rule out the intuitively appealing idea that for a wide range of statements, there might be currently unchallenged, strong evidence for S without it being the case that S must be regarded as indefeasible. Second, it has already been admitted that, where evidence for S cannot be obtained, there is no sense in which (i) can be true. So on this view, whenever there can be no evidence for S, S must be regarded as both not true and not not true - a breach of *tertium non datur*. [28]

So the Anti-realist's appeal to present evidence in explicating the notion of defeasibility is unlikely to appear convincing if the account offered is restricted to present 'negative' evidence
which directly supports (i)''. Of course the temptation is to respond to this by saying that the appeal to extra, negative evidence is unnecessary. It might be pointed out, for example, that there is no need for present evidence in support of the claim that (i)' might be true. All that is required is that there might be evidence that (i)' might be true. And this will be guaranteed, it could be said, if the present evidence which warrants S is, in some recognizable respect, inconclusive. The difficulty is, however, that there seems to be little room left for the idea of 'recognizably inconclusive warrant'. If future evidence is ruled out of consideration as evidence that S may be defeated, there is, for the Anti-realist, nothing left to explain what such inconclusiveness consists in. The Realist can always maintain that, all questions of evidence aside, statements might be defeated no matter what the warrant and, a fortiori, might thus turn out to be defeasible. But this, of course, is not an option for the Anti-realist. Even if we could make out such a notion of inconclusiveness, it is not certain that the resultant concept would meet our intuitions about defeasibility. When we say that S is defeasible, we mean to leave open the possibility that S might be defeated - the possibility, expressed in present terms through (i)', that not-S. But it is not clear that the recognition that evidence for S is inconclusive just is the recognition of evidence for possibly not-S, and it is, therefore, likewise uncertain that such inconclusiveness warrants the claim that S is defeasible.

Anti-realism and truth value links

So long, then, as the notion of defeasibility is made out in terms
of subsequent evidence, it seems that there will always be a tension, in the Anti-realist account, between the notion of defeasibility and the notion that truth is related to evidence in a manner which respects the 'temporal responsibility' which lies at the heart of our everyday use of tensed expressions. But the only other alternative formulation seems to be one made out in terms of present evidence. But this raises difficulties of its own. The unavoidable conclusion seems to be, then that the Anti-realist can make no space for the notion of defeasibility.

What this suggests is that the Anti-realist does face genuine difficulties in allowing for the truth value links within his theory of meaning. To this extent the sought after compromise position has not yet been achieved. There is, however, at least one other competitor, as will be seen in the next section.

§4 REALISM, MEANING AND ABILITIES

The Manifestation Challenge Accepted

At least one philosopher with Realist leanings, Christopher Peacocke, claims that the Realist can make direct reply to Dummett's Manifestation Challenge. Peacocke's aim is to demonstrate that there can be a substantive theory of content meeting this challenge which nevertheless relies on the Realist's notion of truth conditions. [29] His central tactic is to demonstrate that individuals can manifest a grasp of certain assertion conditions which in turn determine truth conditions of contents. He offers specific versions of his analysis for a
variety of different types of content, depending on the nature of the concepts involved. In the rest of this section, though, attention is restricted to the account he provides for those cases in which observational concepts are in play, since it is the observational solution which Peacocke takes as a model in more general cases.

Peacocke's central claim is that there are many cases in which it is true that one could not verify the judgement under consideration - there would be no recognising of the obtaining of the truth conditions - but one could nevertheless manifest a grasp of those truth conditions. His procedure is the following:

(i) Formulate an account of a particular kind of acceptance condition for a Thought or content involving an observational concept.
(ii) Demonstrate that some set of such acceptance conditions determines the truth condition of that content.
(iii) Show that a thinker can manifest grasp of any one of the acceptance conditions.
(iv) Draw the conclusion that since the truth conditions of the content are determined by conditions of which a thinker can, for any one, manifest a grasp, it follows that the thinker can manifest a grasp of matters which settle the issue of the relevant truth value. Further conclude that, depending on the nature of the content, this manifestation can arise in matters which transcend verification, thus avoiding Dummett's challenge.

This procedure is to be understood in accordance with the following suggestions. First, the conditions of acceptance Peacocke has in mind here are those which specify what must be accepted if a given content is judged. Specifically, they are those commitments which an individual incurs in a particular judging of a content. Because they are intended to be responsibilities which we must ascribe to anyone who judges the
content, in that we cannot imagine someone judging that content but not having the commitment, Peacocke calls these commitments canonical. In the case of observational concepts, the commitments involved would be captured by the following formulation, \(C\):

"The spectrum of canonical commitments of one who judges a content at \(t\) 'That block is cubic' is that: for any position from which he were to perceive the block at \(t\) in normal external conditions when his perceptual mechanisms are minimally functioning, he would experience the block from that relative position as cubic, or as a cubic object would be perceived from that relative position." [30]

Next, the idea of a link with truth conditions is captured by the following conjecture:

"There is a class of contents whose truth conditions are directly determined by certain of their acceptance conditions; the truth conditions of contents outside this class are determined ultimately by their relations to contents inside this initial class." [31]

Note that the 'certain' here does not imply, as might be suggested, 'certain of the canonical commitments'. The notion of 'assertion condition' is wider than that of 'canonical commitment'. The 'certain' is intended to restrict the conjecture solely to those assertion conditions which are canonical commitments.

The idea here is that observational contents are such that their acceptance conditions determine their truth conditions. This follows from the following two claims about the commitments associated with observational contents:

\((\text{S.Obs})\)

"For any such content, if all its canonical commitments are met, then it is true." (a sufficiency condition)

\((\text{N.Obs})\)

"For any such content, if it is true, then all its canonical commitments are met." (a necessity condition)
Chapter 1

If S.Obs and N.Obs are true, then these canonical commitments determine the relevant truth conditions.

An individual can manifest a grasp of any one of the acceptance conditions involved in a particular observational judgement, according to Peacocke, by confirming its truth (or by dis-confirming it) and reacting accordingly. (This will not involve an infallible knowledge claim, given the caveats over normality conditions and normal functioning of the thinker's perceptual apparatus.) For that instance, then, the thinker can manifest that it is one of his commitments in judging the content. The idea here is that such manifestation leads to a manifestation of a grasp of truth conditions. Peacocke's suggestion is that:

"grasp of a truth condition is manifested by a thinker's manifesting his conformity to certain normative acceptance conditions in making his judgements..." [32]

Whether such conformity is ascribed to the speaker depends on the circumstances which characterise the pattern of his actual and counterfactual acceptances and rejections of 'That block is cubic'. If they are of the appropriate sort:

"...then the pattern of his actual and counterfactual acceptances and rejections of 'That block is cubic' in rational response to his perceptual experiences can make it reasonable to ascribe to him the spectrum of commitments in (C)." [33]

This means that the individual involved can manifest, through his pattern of actual and counterfactual acceptances and rejections of 'That block is cubic' that just the commitments in (C) are his commitments. (Of course 'counterfactual acceptances' needs some spelling out.) And this, in virtue of S.Obs and N.Obs is, according to Peacocke, a manifestation of grasp of a truth
Chapter 1

condition. Since there is no suggestion that the individual has been in a position to recognize the obtaining of every one of the members of the set of canonical commitments, this, according to Peacocke, represents an instance of manifestation without verification.

It is noteworthy that nothing said so far involves a reductionist account of the observational predicates under inspection. There is no suggestion that anything other than a non-trivial reply is available in answer to the question 'what makes this statement true'. In parallel cases dealing with theory of meaning, rather than theory of content, there is nothing amiss in stating the meaning of sentences in terms of trivial truth condition formulations, (complaints have only been raised when attempts are made to rely on such formulations in explaining what a grasp of the meanings so specified would consist in.) The same point ought to apply to a theory of content. This does mean, however, that for anyone impressed by Dummett's manifestation challenge, a direct statement of truth conditions cannot be used substantively to explain the content of observational judgements. A substantive account of what the content of such thoughts is must therefore find some other way of characterising these judgements. The obvious recourse, in the observational case at any rate, is to the way thinkers react to what they experience. Peacocke's aim is to characterise such reaction by giving a specification of reactions to a variety of different experiences. These reactions are, specifically, those which anyone who makes the observational judgement ought to take himself to be committed to exhibiting. It
is in the nature of a consequent discovery that the commitments so outlined are precisely those which delineate the conditions which settle the truth value of the judgement involved.

The Canonical Constraints on 'That Block is Cubic'

If the thinker perceives an object and makes the judgment 'that block is cubic', then C demands that he either:

(a) perceives the object as a cube, or:
(b) perceives it as a cube would be experienced.

By contraposition, it follows that when the thinker perceives the object, he will be expected to withdraw his judgement if both (a) and (b) are false - if he both perceives the object as non-cubic and he perceives the object other than as a cubic object would be perceived. A thinker manifests that the commitments described by C are his commitments by his pattern of judgement withdrawals.

This leaves an important distinction to be drawn where the thinker accepts the commitments in C. When a commitment fails to be met, the thinker ought to withdraw his judgement - such required withdrawals will be determined by C and by the character of the thinker's experience. The set of occasions on which the thinker ought to withdraw his judgement may, however, be distinct from the set of occasions on which he will withdraw his judgement. This latter set will be determined by C and by what the thinker takes the character of his experience to be. There will be cases of experiences of cubes in which the thinker's perspective is such that it is not obvious to him that the object is a cube - he perceives the object as non-cubic. In such cases, the thinker's
acceptance of the commitments described by C will cause him to withdraw the judgement 'that block is cubic'. But clearly, C demands the withdrawal of judgement only if the thinker perceives the object as non-cubic and he perceives the object as other than a cube would be experienced. And however the cube looks to the thinker, no matter how odd the perspective, just is how a cube would be experienced. In these cases then, C does not demand that the thinker withdraw his judgement. Nevertheless, the thinker may (incorrectly) withdraw his judgement. So there is a distinction, when a thinker accepts the commitments in C, between cases when he ought to withdraw his judgement and cases when he will withdraw his judgement.

Of course the thinker does think that even his mistaken withdrawal of the judgement is demanded by the commitments given at C. But this only shows that according to C the thinker can have a commitment which is such that he can be mistaken as to what action is demanded when the commitment is met. The thinker's acceptance of C's commitments as his commitments will be constituted by his accepting that if he judges that some object is a cube and if he perceives it, then either (a) or (b) will be true. And the difficulty is that the thinker is unable to identify all of the occasions on which either (a) or (b) is true - there will be occasions on which (b) is true, and so the commitment met, and yet the thinker is unable to recognize this. So the importance of the distinction between the set of cases in which the thinker ought to withdraw his judgement and the set of cases in which the thinker will withdraw his judgement is that the thinker's being in a
situation which fulfils one of the commitments described by C will be a matter which on occasions transcends his recognition capacities.

Clearly, there are some commitments which a thinker can have which are such that they can be met without the thinker's recognizing that they are met. The question is, however, whether any commitments of the kind described by C ought to be such that they could be met without the thinker's recognizing that they are met. The answer has to be: no. If there are some commitments which the thinker cannot recognize as having been met then for those commitments he will alter his judgement, but not in accordance with the dictates of the commitment he in fact holds. So in these cases, his withdrawal of judgements will not be a reflection of his commitment-acceptance. But the whole point of the appeal to commitments is that, for any given commitment, the thinker can manifest that it is his commitment by recognizing through experience whether it is met and altering his judgement accordingly. So for these cases, the thinker will be unable reliably to manifest that the relevant commitment is his commitment. Now the theory states that the assertibility conditions represented by such commitments only constitute truth conditions because in total they are jointly necessary and sufficient for the truth of the judgement. So if it is not the case that each of the commitments is manifestable, then it is not, contra the claims of the theory, the case that (even were such commitments to constitute truth conditions) the related truth conditions are manifestable. And since the central claim was that
the theory provides an account of manifestable truth conditions, the theory falls.

Of course, this problem can be readily resolved by limiting C to those cases in which the nature of the cube is recognizable. This would mean doing away with the (a)/(b) distinction. This, in effect, is what is suggested by Peacocke himself, who is aware of the difficulties presented by odd perspectives. Peacocke's suggestion is that the perspectives mentioned in the characterization of commitments offered by C should be restricted to 'obvious' perspectives. [34] The relevant withdrawal of judgement would then be limited to cases where the thinker's view of the object is 'obviously not' one which would be afforded by looking at an object which did fall under the concept 'cubic'.

A Problematic Restriction

Peacocke's claim is that the removal from C of the commitments associated with such 'non-obvious' views is justified on the grounds that such commitments could not be canonical. The thinking here is presumably something like the following.

(1) Suppose that the object is a cube.
(2) If the thinker takes any of the commitments associated with 'non-obvious' perspectives to be his commitment, he will incorrectly withdraw the judgement 'that is a cube'.
(3) No withdrawals which are incorrect should determine the canonical commitments of the content judged.
(4) Therefore no commitment determined by withdrawals consequent upon experiences from 'non-obvious' perspectives can be canonical.

But this cannot be correct. In making a judgement 'that block is cubic', one commits oneself to the truth of the content judged. But even with 'non-obvious' perspectives, (where (a) is false even
though the content judged is true) if the content judged is true, (b) is true. So even with 'non-obvious' perspectives, if one makes a judgement 'that block is cubic', one is committed to the truth of (b). However, if some specific commitment, C', is non-canonical, then by definition it makes sense to conceive of someone judging the content 'that block is cubic' yet not incurring that commitment. The thinker, that is, can make the judgement without being committed, by C', to the idea that were he to experience the block from the 'non-obvious' perspective associated with the commitment, either (a) or (b) would be true. So he is not committed to the idea that (b) would be true. But as just seen, the thinker must be committed to the truth of (b). So Peacocke's claim, that the commitment determined by a withdrawal based upon experience from a 'non-obvious' perspective is not a canonical commitment, must be false.

So Peacocke's theory offers no reason for excluding non-manifestable commitments from C. But the claim that grasp of the truth conditions for the relevant contents is a manifestable affair depended on the truth of the claim that all of the commitments in C are manifestable. Consequently, Peacocke offers no support for the claim that such contents are manifestable.

What is needed is a restriction upon C which limits relevant commitments to recognizable affairs. One alternative argument to this end might be the claim that a demonstration of competence in observational judgements will centre on recognition capacities. The suggestion would then be that the only commitments which are
of interest in specifying the content of observational judgements are those which stipulate responses to experiences whose nature the thinker recognizes. On this account, the commitments which are said to be jointly and necessary and sufficient for the truth of the relevant judgement are a sub-set of some wider set of canonical commitments. The difficulty with this is that the other, excluded, canonical commitments, by definition, must be true if the judgement is true. This means that the commitments in the restricted set cannot, after all, be sufficient for the truth of the judgement.

Of course the Anti-realist is willing to accept that for any decidable judgement, grasp of the truth conditions of the content judged will be a recognizable affair. Certainly, as noted above, it seems fair to suppose that we would not count anyone as competent in the use of the concept 'cubic' if he or she could not demonstrate an ability to recognize cubic objects. So although Peacocke is mistaken in his claim that 'non-obvious' views produce non-canonical commitments, the existence of such perspectives ought not to be regarded as demonstrating that the truth conditions for the contents of judgements such as 'that block is cubic' are non-recognizable. What the existence of such perspective does demonstrate, however, is that the content of judgements cannot be adequately characterised by appeal to a spectrum of commitments which is as wide as the spectrum of 'canonical' commitments.
Chapter 1

The Anti-realist, then, is quite willing to accept that such non-obvious cases will not feature in whatever is to count as a measure of competence in dealing with cube recognitions. But this only shows that the appropriate measure of competence cannot be captured by appeal to canonical commitments. It might be protested that, since the Anti-realist account is willing to admit that the problematic 'non-obvious' cases ought to be excluded, Peacocke's spectrum of commitments should be allowed the same powers of exclusion. But note that the implications of the Anti-realist position and those of Peacocke's positions are not the same. In the Anti-realist's case, the essential recognizability of truth conditions requires that 'non-obvious' cases be excluded from whatever notion it is that correctly captures the idea of competence in wielding 'cube'-talk. On Peacocke's theory, on the other hand, the canonical nature of the commitments associated with cube talk demands that such 'non-obvious' cases be included in the specification of 'cube-talk' competence.

Apart from the observational cases, this is bound to have an effect on any attempt to extend the theory to more controversial cases. Of particular interest is Peacocke's account of universal quantification. Peacocke states:

"When we reflect on the observational case, it appears that there is no need to appeal to a thinker's recognition of a truth condition as obtaining for there to be a manifestation of grasp of a classical truth condition...On the neo-Ramsayan account, universal quantification is just another illustration of this general point." [35]

Since it is no longer true that the observational case makes no use of the notion of recognition of a truth condition, this will also hold for the case of universal quantification, since this is
said to be just 'another illustration' of the same point. But whereas the idea of recognizing the fulfilment of a truth condition was a mutually acceptable notion in the observational case, it is precisely the notion which is at dispute in the case of universal quantification. In consequence, Peacocke's claim to have provided an account of universal quantification which both relies on the classical notion of truth and yet meets the anti-realist's challenge must be rejected.

§5 CONCLUSION

This chapter began by outlining a variety of ways in which those interested in the theory of meaning approach to language might develop their ideas. What most of these developments have in common, it was said, was an emphasis on the adoption of a central explanatory concept such as the Realist truth condition or the Anti-realist verification condition. The suggestion was that although the choice of any one of these concepts had virtues, it also involved difficulties. The chapter then outlined two 'touchstone' arguments: the 'Temporal Responsibility' argument and the 'Manifestation Challenge' argument. It was suggested that if the Anti-realist or his Realist opponent could meet the relevant challenge, a valuable 'compromise' position, containing the virtues of both positions but involving the difficulties of neither, might have been reached. An examination of recent attempts to provide such direct replies to the respective arguments was made, but the conclusion drawn was that neither
Realist nor Anti-realist had yet satisfactorily met the initial problem.

It should be emphasised that this discovery does not, of itself, demonstrate either that there is something inherently wrong in the theory of meaning approach or that one or other of the two theory of meaning opponents - Realist and Anti-realist - is in a relatively worse position than heretofore. What can be said, though, is that two initially attractive responses, to what are crucial difficulties for the two meaning theoretical positions, have been ruled out. This provides an indication that it may be fruitful to explore further the preconceptions which lie behind the theory of meaning approach.

Perhaps the most instructive outcome of the preceding sections is that an account of the philosophy of language has need of a more developed epistemological stance. The difficulty is that we expect our philosophy of language to combine two seemingly opposed positions: we want it to respect the way we use 'true', such as in expressing the idea of temporal responsibility, while at the same time we want it to account for the abilities which normal people seem to display in their everyday language use. It is implicit in many Realist and Anti-realist approaches to the theory of meaning that the way to tackle these tasks is to provide an explicit account of that which people implicitly know. What the preceding sections indicate, however, is that the idea of 'knowing that' may be inadequate.
Chapter 1

In the case of Peacocke's response to the Manifestation Argument, for example, it became clear that the position on offer was required to analyse understanding of meaning in terms of knowing that a given statement has a specified set of canonical commitments. The difficulty with this account was that it proved unable to isolate as 'canonical' the appropriate commitments. This stemmed from the fact that the statement of commitments on offer had to be viewed against a background context of normality which excluded odd viewpoints and the like. Now what the failure of this account ought to suggest is, not so much that there could not be complete accounts of this type in which these background contexts are explicitly stated for fragments of our language, but that, just because this explicit account of background is necessary in such a model, it cannot act as a completely general model of language explanation.

One might equally well suggest the same moral for the case of the Anti-realist. On his account, language understanding is viewed in terms of the link between statement and evidence which the speaker must know. But just because, in many cases, this evidence is not conclusive for the truth of the statement, this way of stating that which is known by one who understands the statement's meaning must present a misleading view of when statements are true and false. For the only alternative would be to accept that individuals might usually be in a position of asserting statements while knowing that they might be false. Once again, this implies that the difficulty stems from an insistence that the account of
understanding in play must be one centred on that which the individual knows.

What this suggests, then, is that there is at least scope for the suspicion that the meaning-theorist's difficulties, whether Realist or Anti-realist, stem from his conception of language understanding as 'knowledge that'. The next chapter attempts to explore a re-working of the problem by looking in closer detail at the very idea of understanding a language.
CHAPTER 2

§1 INTRODUCTION

In Chapter 2, we begin by re-assessing the meaning theorist's 'knowledge that' characterization of language understanding. This re-assessment leads to an alternative account of understanding based on Ludwig Wittgenstein's remarks on meaning as use and rule-following. This alternative account is seen to have two central features which raise the possibility of a dilemma for the Wittgensteinian. §2, §3 and §4 chart a range of unsuccessful attempts to resolve this dilemma, while §5 provides a more satisfactory resolution.

Understanding as 'knowledge that'

The Realist and the Anti-realist arguments presented in the preceding chapter rely on the assumption that the fundamental goal of the analytic philosopher is to construct a theory of meaning. Now in general, theory of meaning accounts fall into two main groups: those theories which are substantive and those others which are non-substantive. [1] A theory of meaning can be seen to be substantive or non-substantive by examining the interpretation offered for the individual theorems involved. If the meaning theory on offer is content to produce an account of meaning which will, for any arbitrary sentence, merely state by means of a specific theorem what that sentence's meaning is, then the theory can be termed non-substantive. If, on the other hand, the theory is more ambitious, and aims to state, for an arbitrary sentence,
what that sentence's meaning is and explain that in which an understanding of the sentence consists, then the theory can be termed substantive. Both the Realist and the Anti-realist of the previous chapter attempt to present substantive theories. They not only claim that the theorist can state the meaning of a sentence by stating its assertibility conditions, they also claim that the theorist can state what an understanding of that sentence consists in by stipulating what would count as 'manifestation of grasp of assertibility conditions' for that sentence.

So for the meaning theorists presently under consideration, analytic philosophy has two goals: to provide an account of that which the competent language speaker knows and to offer an explanation of that in which such knowledge consists. Within this theory of meaning framework much of the discussion centres on whether, for a language speaker who understands the meaning of some item of language, it is best to represent 'that which is understood' in Realist or Anti-realist terms. This leaves untouched (if not independent) the issue of how best to represent that in which such understanding consists.

It is implicit in the perspectives discussed in the preceding chapter that such understanding consists in knowledge of a 'knowledge that' form which issues in (manifesting) behaviour. Now the difficulties faced by both the Realist and the Anti-realist suggested the possibility that there might be a more appropriate format for the representation of language understanding. The difficulty with this proposal, however, is that initially it is
hard to see what could be done to augment or replace the account of language understanding as an account of that which a speaker knows. The basic assumption in the philosophy of language is that a sentence's meaning depends on the elements from which it is composed; the whole point of viewing language as a rational activity is that one thereby presupposes that there is purposeful selection of phonemic or graphemic signs in communication. So the idea of structure seems to be inextricably linked with the idea of such rational linguistic action. The aim of a substantive theory which employs a 'knowledge that' format is that language understanding be explained in terms of an expression of speaker's knowledge which is structure-reflecting. [2] So if we claim that the standard of correct use for an expression cannot be represented by means of meaning-theoretic theorems we seemingly reject the possibility of explaining language use in terms of structure, and thereby seem to undermine the idea of rational language use.

So there is a sense in which it ought always to seem appropriate to say of an individual that he knows the meaning of a sentence involving 'green', say, only if he knows that this expression applies solely to green things. [3] And this makes it look as though some form of meaning theory which specifies, by means of its theorems, specific items of 'knowledge that' which the competent speaker must possess, should always be open to the analyst. But this specification of the 'knowledge that' required for an understanding of 'green' is not yet explanatory in the substantive sense which was discussed earlier. It is, then, a
second, separate claim to maintain that the 'substantive' models of language understanding under review must cast language understanding in terms of meaning-theoretical 'knowledge that'.

[4] What this means is that there is scope for developing an alternative account of language understanding in which the first claim is retained (thereby respecting structural constraints) but the second claim abandoned.

The argument of the rest of this chapter is that there is at least one model for such an alternative account already to hand: Ludwig Wittgenstein's remarks on meaning as use. [5] To make out this claim, we begin by examining the rule-following notion which lies at the heart of Wittgenstein's remarks on meaning as use.

**A Wittgensteinian account of rule-following**

It has been claimed that an understanding of Wittgenstein's later work on meaning must be predicated upon a clear grasp of the thoughts outlined in sections 139 - 242 of *Philosophical Investigations*. In these sections, Wittgenstein attempts to develop and clarify his ideas on concepts such as understanding and explanation through the exploration of what it is to follow a rule. His aim is to explain language abilities in terms of the ability to employ a linguistic expression in accordance with the rules for its use. Unfortunately, there is little contemporary consensus in the interpretation of these sections and, consequently, a variety of competing 'rule-following' accounts of meaning has developed. However what several of these accounts have
in common is an acceptance of a set of ideas which can be
encapsulated in the following two general theses.

(1) Regulation Thesis (RT): A rule following account of
meaning must treat an individual's uses of an expression as
being correct/incorrect relative to standards set by the
rules for the use of that expression.

(2) Constitution Thesis (CT): communal practice over a
suitably specified temporal period constitutes, in part at
least, the standard of correctness against which uses of an
expression are to be gauged.

In general, a regulation component seems to be a necessary element
in a rule following account because any such account which lacked
a regulatory element would not fit our everyday understanding of
what it is for a certain usage to be constrained by rules. If the
standard of correct use of an expression is in no way regulated by
the rule, then any talk of rule-guidedness, over and above
whatever explanation is given of correctness, may well seem
superfluous. [6] The general idea behind the CT thesis is that any
account of meaning which lacked a constitution element would
present an undesirable consequence. If the meaning of an
expression does not somehow depend on our use of it, then it makes
sense to conceive of our communal use diverging once and for all
from the 'correct' usage - even if communal usage represented a
coherent practice which fulfilled for us whichever communicative
goals were associated with use of the relevant concept. But the
meaning of an expression is normally taken to be directly
associated with such communicative function. It follows that an
account which lacked a constitution element would unsatisfactorily
Chapter 2

separate out the question of what an expression means from the question of how it is actually used.

There is, however, a difficulty attached to combining CT and RT in this way. Although CT and RT both deal with the notion of correctness, they differ in terms of the concepts used to make sense of this notion. Indeed it might even seem that the two theses represent conflicting accounts of what it is for someone to use an expression correctly. This potential conflict could be mapped out in the following way.

RT Argument

(RT₁) According to RT, any individual’s use of a linguistic expression must be explained by showing how the rule which that individual is following marks out correct and incorrect usage. This presupposes that it makes sense to talk of ways in which the individual could go wrong in his use of that expression. (RT₂) But sometimes if it makes sense to think of any one individual going wrong in his use of the expression, it makes sense to think of all individuals in the community going wrong in the same way. If the individual’s error is caused by drug-induced misperception, for example, there is no reason to suppose that such a drug could not influence all the other individuals in the community. (RT₃) If it makes sense to think of all the individuals in a community going wrong in their use of an expression, it makes sense to think of the community itself going wrong. But to think of the community itself going wrong in its use of an expression is, in some respect, to think of communal uses of that expression as going
wrong relative to some rule which guides communal usage. So RT implies that the community's usage of some expression can be thought of as rule-governed in such a way that the whole community could be mistaken in that usage.

CT Argument

(CT), on the other hand, maintains that: (CT,) communal usage is constitutive of correct usage. It follows from this that the community cannot be considered to be governed by a further, external rule. (CT$_2$) For such an external rule would either have the potential to deliver different verdicts on correct usage from those delivered by communal use or it would not. If it did not, then nothing would be gained, in giving an account of the correctness of such an expression in terms of communal usage, by appeal to such a rule. If it did, then there would be some possible situation in which (a) there is communal agreement that a particular usage of some expression is correct, but (b) the community regards this usage as incorrect relative to the external rule. But there is no sense in which the community can be thought of as looking behind its own decision (as to when an expression is used correctly) to judge whether that decision is correct. So there can be no sense in which communal usage can be thought of as responsible to such a rule. (CT$_3$) It follows from this that the community's usage of some expression cannot be thought of as rule-governed in such a way that the whole community could be mistaken in that usage.
If these arguments hold up, then the Wittgensteinian seems to face a dilemma. On the one hand, a Wittgensteinian account which lacks a version of CT or a version of RT will not meet our intuitive notions about the rule-guided aspects of language. On the other hand, RT and CT seem, on the face of things, to be incompatible. By the RT Argument, RT seems to imply that the community as a whole could be mistaken in its use of an expression. By the CT argument, CT seems to imply that this cannot be the case. If the Wittgensteinian is to avoid this dilemma, it follows that there must be something awry in the arguments developed above. There are three general possibilities:
(a) rejection of CT
(b) rejection of RT
(c) demonstration that CT and RT are compatible.
The first of these seems to be clearly ruled out. Conceiving of the community as a standard of correctness as captured by CT is a central aspect of the whole Wittgensteinian 'meaning-as-use' approach. So in order to provide an account of language understanding which uses the Wittgensteinian notion of rule-following, we must begin by examining (b): ways in which RT can be rejected, and (c): ways in which the seeming conflict between CT and RT is to be avoided.

§3, §4 and §5 explore different attempts at the latter option. We begin, however, in §2 with an attempt to demonstrate that the former option - rejection of RT - is the appropriate response.
Chapter 2

§2 Kripke: Accentuate the Communal, Eliminate the Regulative

The Kripke account

In his book *Wittgenstein on Rules and Private Language*, Saul Kripke argues that a Wittgensteinian account *must* abandon RT. On his view, the proper way to take Wittgenstein’s remarks on rule-following is to see those remarks as formulating a challenge which can be cast in a sceptical light. They are taken to show that having a grasp of the rule for the use of an expression cannot be regarded as involving some kind of commitment as to one’s future use of that expression. Any use can be made out, on some interpretation, to accord with the rule, and so one can never be said to know what the future correct use of an expression will be. Consequently, there is no content to the idea that one must do so—and-so if one is to employ the expression correctly, where 'correctly' is taken to mean 'according to a standard of correctness represented by the rule'. The best that one can do, by way of prediction, is to assert that the question of which use will be correct on some given date in the future is an issue which will be settled, somehow or other, by appeal to the subjective inclinations of those in the community. [7]

The key to this scepticism is the idea that claims about meaning suffer from a chronic lack of justification. For example, claims about what I meant in the past (as recovered from some past explanation, say,) are seen as inadequate justification for regarding some current use of an expression as correct. Even if a variety of introspective facts are admitted as evidence together
Chapter 2

with whatever shows up in the behavioural record, the Kripke-sceptic's claim is that there is nothing about the totality of such facts to establish that some present application of a term is determined as correct by what I meant in the past, in a way that some other is not. In this sense, there is just no fact about what I meant. And since the strength of the evidence about what I meant in the past has been approximated to that of contemporary evidence, the Kripke-sceptic's conclusion is that the justificatory shortfall also extends to current claims about what I mean now by some expression. In effect, then, there is no fact of the matter as to what we mean by any expression. However we go along in our usage just is what is meant. In this respect, RT must be completely abandoned.

Parry: riposte

Now a short reply to this kind of scepticism would be that it ignores Wittgenstein's idea of grammatical linkage between rule and explanation. [8] Kripke tries to raise a doubt about what is to count as the correct present application of a rule-guided term, even given a full range of explanation. There is, he says, no evidence strong enough to prove that such and such an explanation really meant that we should follow one rule rather than another. But if explanations of an expression are grammatically linked to the rules for the use of that expression, there could not be, nor is there any need for, any form of final support or justification for the claim that 'if he is to follow the rule for the use of this expression, he will do so-and-so' which stands external to such applications. And so it makes no sense to talk here of a
shortfall in justification. For conditionals such as this, the notion of external evidence is completely misplaced. Of course there is still doubt as to whether any particular individual or group of individuals will stick to this correct usage. But this is not the same thing as saying that there is no such standard of correctness. [9]

But it will be replied that this answer misses the point. It might fairly be said that the worry is not: 'even given that we know what explanation we were following in the past, how can we be sure about what the appropriate present usage is', but rather: 'it makes sense to wonder whether some present usage is in accord with the rule we were following in the past, as 'grammatically' determined by some explanation, so we can never be certain which rule we were following then'. On this latter formulation, even if it is accepted that the correct way to take talk of rule-following is to see it as the claim that rule and explanation are 'grammatically' related, a sceptical worry remains. For it might be said that nothing in the past or in our current usage settles the issue as to whether, in the past, we intended to follow the rule/explanation which forbids our present applications or another rule/explanation which allows them - nothing in the present or past guarantees that we were following one rule rather than another. So we can never be sure that our current applications are correct with respect to the rule which, historically, we might be said to have been following.
The bite in this version of the argument is that it undermines the whole idea that our use of an expression can be understood in terms of our being in conformity with a rule in precisely the way Kripke suggests. If it ever makes sense to doubt whether present use is in conformity with a given rule said to be previously grasped, then there can be no evidence strong enough to prove that we were, in the past, following one rule rather than another. The only alternative is to claim that no such doubt about current usage will arise. But now there seems no difference in content between the claim that we feel confident about applying the relevant expression in such-and-such circumstances and the claim that we are confident that we are following the rule. If any occasion arises in which we are not confident that so-and-so is the correct way to go on in the employment of some expression, nothing about the appeal to the rule for the use of that expression will help resolve matters. But this leaves matters in the same position as before: normativity seems to have been abandoned.

The problem is, then, that we seem to face a dilemma: either it makes sense to question whether present usage is in accord with past usage or it does not. If the question does make sense, then this opens up a sceptical doubt: no evidence is strong enough to settle the question. If it does not make sense then the kind of guarantee required for the normativity associated with RT, to the effect that communal usage is somehow responsible to a standard of correctness, cannot be made out. In either case, it seems that a
proper account must, as Kripke suggested, do without an RT-style standard of correctness.

A problem with Kripke's account
Kripke's sceptical conclusion is that there can be no normative facts. What this means is that his positive account - the 'sceptical solution' - offers a view of language use which relies solely on CT. [10] Individuals are ascribed language understanding by others solely in virtue of the extent to which they agree in their linguistic responses with those others. To this extent, the individual is responsible to the community, but the usage of this linguistic community as a whole cannot be factually described at all. There is nothing to our communal use of language other than the contingent fact that we happen to agree with one another in the way we are inclined to use our linguistic expressions.

Independent of the status of his argument towards this sceptical conclusion (see §5 for criticism), it is clear from what was said in §1 that there ought to be some difficulty in generating a positive Wittgensteinian account in which RT is absent. And, indeed, there is a flaw in the positive picture represented by the 'sceptical solution'. It lies in Kripke's explanation of the point of talk about meaning. Kripke describes his sceptical solution as one in which the search for the 'truth conditions' of statements about meaning is replaced by the search for 'assertibility conditions'. [11] This search for assertibility conditions has, according to Kripke, two aspects.

(A) First, it identifies the 'justification conditions' for
attributing grasp of meaning of an expression to someone. Roughly, these cash out as feelings of satisfaction: I will correctly attribute grasp of meaning to you if you apply the expression in the way that I am inclined to apply it. This is, of course, a contingent matter. [12] (B) Second it identifies the point or utility of the statements which ascribe such grasp of meaning. This utility can be captured by inverting a popular form of conditional. Usually, we would ascribe meaning by saying something like: 'If Roger means addition by "+", then if he is asked for "8+8" he will reply "16".' (And for Kripke, this means 'If Roger means addition by "+", then if he is asked for "8+8" he will reply what we are inclined to reply') But on the Kripke account, we should invert this conditional. We should read it as emphasising that if Roger does not come out with "16", we cannot assert that he means addition by "+". Although truth-conditionally the same as the original conditional, the inversion emphasises that it is a restriction upon us that we can only ascribe grasp of meaning to those who conform with what we would do. This, then, is Kripke's 'sceptical solution'. It makes out room for a version of CT by appeal to the idea of 'communal' inclinations - inclinations which we might, non-collusively, happen in fact to share - while ruling out the idea of RT. All there is to meaning, on this account, is a set of socio-psychological contingencies held together by brute fact.
Chapter 2

But the abandonment of RT in this positive account leads to difficulties. According to Kripke himself, we are to regard the point of ascribing grasp of meaning to others as reflecting a 'restriction' on the extent to which we include people in our language community. Now the importance of this inclusion is that we thereby place 'reliance upon his responses'. [13] So there is a point in saying (in the contraposited form advocated by Kripke) 'If Roger does not reply as we are inclined to reply, then Roger does not mean addition by "+"'. Such statements separate those upon whose responses we can rely from those upon whose responses we cannot rely. But the difficulty here is to discover what the sense of restriction involved in 'cannot' is supposed to amount to.

According to Kripke, it is merely a contingent fact about us that we have the inclinations that we do. [14] But there is no justification for our having such inclinations. [15] So that I have a particular inclination is an unjustifiable, contingent fact about myself. In which case, it seems reasonable to inquire as to the sense of restriction involved in the discovery that Roger does not conform to one of my inclinations. If I withhold ascription of grasp of meaning from Roger, I cast doubt on the reliability of his responses. The grounds for this doubt are supposed to be a disparity between the way I am inclined to go on and the way Roger is inclined to go on. But suppose Roger does not reply as I am inclined to rely in the case of "8+8". If inclination is an unjustified, contingent fact, there is no reason for me to suppose that this disparity between myself and Roger will continue. His disagreement with me in replies might suddenly change. Of course,
I have no evidence for this, but then I have no evidence for the contrary either.

There is, in other words, no sense in which I ought to feel restricted in ascribing grasp of the meaning of "+" to Roger even though our inclinations differ in any particular case. What we want is to say that our difference in replies in any particular case is indicative of a continued disagreement in inclination, such that his future responses are unreliable. But there is no evidence for such a conclusion. Of course we might appeal to an inductive principle: 'his inclination has not changed in the past, so it will not change in the future'. But even supposing there was no problem with such an inductive manoeuvre (and if there is not, it is hard to see why Kripke's dispositionalist opponent cannot make good use of it), a further difficulty remains. There is nothing to convince me, even given as much evidence about past usage as could be imagined, that Roger does not have a notion in mind, similar to 'plus' (call it 'quus') which is such that, from now on his responses will accord with mine. Inductive support might convince me that he will go on in the same way. But it does not convince me that his idea of how to go on in the same way will continue to differ from my idea of how to go on in the same way. Alternatively, suppose Roger and I agree in our use of "+". There is nothing to convince me that he does not have some concept (call it 'muus') which is such that his usage will, after tomorrow, diverge from mine.
Chapter 2

So the difficulty which Kripke's positive account faces is that there is, contrary to Kripke's claim, no point to the ascription of grasp of meaning on such an account. Similarity or difference in inclination now tells us nothing about disparity in inclination in the future, and yet it is future responses whose reliability we are supposed to be characterizing. This difficulty stems from the fact that on Kripke's view there is nothing normative about agreement or disagreement in inclination. What this implies is that, as argued in §1, on any rule-following account there will always be some role for a normativity element of the kind captured by RT.

This leaves undiscussed the question of where it was that Kripke's negative, sceptical argument went wrong. In §5, we will discover what this mistake consisted in. First, though, we must produce a more convincing version of Wittgensteinian rule-following. To do this, as we saw at the end of §1, requires that we either (a) abandon CT or (b) abandon RT or (c) demonstrate that, contrary to the arguments outlined in §1, CT and RT are compatible. Since both (a) and (b) have now been ruled out, we must now pursue task (c).

§3 CT/RT COMPATIBILITY I: A 'LIMITED COMMUNAL MISTAKE' ACCOUNT

The preceding section concluded that to avoid the CT/RT dilemma, the Wittgensteinian must challenge the claim that CT and RT are incompatible. One way to do this is to demonstrate that even when RT and CT hold the community can, contrary to the conclusion of the CT Argument, go wrong, but that the extent of this mistaken-
ness cannot, contrary to the conclusion of the RT Argument, be as
great as the misguidedness which can afflict individuals when they
attempt to follow the rule of use for a linguistic expression.
What such a viewpoint would offer is a weakened version of
'communal mistake' which is strong enough to demonstrate that the
community in some respects is in the same boat as the individual
in respect of RT, but is weak enough to allow for an etiolated
version of CT.

To see what such a weakened version of 'communal mistake' might
look like, we can begin by considering a version of the CT
Argument offered by Steven Holzman and Christopher Leich [16].
Leich and Holzman claim that:

"...if asserting $S$ is constitutive of meaning
whatever the majority means, it cannot be both
wrong for the majority to claim that $S$ holds
and yet right for them to mean what they do. We
might try to sum this up by saying that
Wittgenstein is committed to holding that, for
a large class of their judgements, the majority
of speakers of a given language cannot be
(simply) mistaken..."

[17]

If the Wittgensteinian can demonstrate that this instance of the
CT Argument is problematic while retaining some version of CT, he
will be in a position to retain both CT and RT, thereby avoiding
the rule-following dilemma sketched in the previous section.

You can fool all of the people some of the time...

To retain both CT and RT, the Wittgensteinian must demonstrate
that CT is compatible with the communal claim that $S$ holds and
with it being the case that $S$ does not hold. In order to supply
such a demonstration, the Wittgensteinian must show that the
standard of correctness represented by communal use does not rule out occasions in which the community claims that $S$ holds and yet $S$ does not hold. The communal standard must be one which allows for exceptions. Clearly, what matters here is the extent to which exceptions to the communal standard are allowed. If the claim were made that the community could *always* be in the wrong in its use of some expression, for example, this would have to be regarded as directly ruling out the idea that communal use might, nevertheless, be taken as setting the standard of correctness for communal use of that expression - CT would be completely undermined. If the communal usage is to set a standard at all, there must be *some* limitation on the extent to which the community can ever go wrong. So the Wittgensteinian must demonstrate that there can be some conception of the community's use of an expression which represents that use as a standard of correctness which (a) allows for occasions on which the communal use of that expression goes astray, and yet (b) stipulates that those occasions must form at worst a limited part of such communal usage.

The difficulty now, however, is finding a principled cut-off point for degree of communal error. If the Wittgensteinian accepts that the community may go wrong on any particular occasion, he then accepts the responsibility of demonstrating just how to rule out the idea of the community going wrong on all such occasions. One suggestion, offered by Crispin Wright [18], is the drawing of a distinction between two separate claims. The first is the claim that we can conceive of the community being possibly wrong in all
of the occasions of its use of some expression. The second is the claim that we can conceive, for all of the community's occasions of use of that expression, of the community being possibly wrong. The first claim, which is inconsistent with CT, is clearly a stronger claim than the second. If Wright can demonstrate that the second claim does not imply the first, then the second claim will represent a possible formulation of the kind of principled cut-of-point for degree of communal error which was required. This would then allow the Wittgensteinian to retain CT while admitting that the community can sometimes go wrong, such error being formulated in terms of the second, weaker claim.

Wright argues that we need not suppose that the second claim entails the first provided we accept that communal use of an expression stands to the standard of correctness represented by such use as do the episodes of behaviour produced by an individual to his personality or character. When we describe an individual's character, we intend to speak of a feature of that individual which will in general explain further behavioural episodes without faultlessly predicting such episodes. Applied to the sphere of language, this idea of general characterization suggests that just because each episode of communal use counts toward the totality of communal use, it makes sense to suppose that any particular occasion of use might be erroneous, without supposing that such error might occur generally. If I think about a room-full of people of varying but unknown heights, it makes sense for me to wonder, of each person, whether he or she is the tallest person in the room. It does not, however, make sense for me to wonder
Chapter 2

whether every person in the room is the tallest. This state of affairs results from the fact that the current 'value' of the tallest-person-in-the-room 'formula' is one which is determined by the heights of each individual person in the room. In a similar vein, we can wonder whether a particular communal usage of some term is correct without being thereby entitled to wonder whether all communal uses of that term are correct.

The original difficulty was that the Wittgensteinian faces a dilemma if he accepts both RT and CT, so long as RT and CT continue to prove incompatible. To alter this state of affairs, the Wittgensteinian had to demonstrate that the possibility of communal error (guaranteed by the RT Argument) is somehow consistent with the claim that communal use is constitutive of correctness (guaranteed by the CT Argument). This required a proof that there could be some principled cut-off point for such communal error which, by allowing for a weakened version of 'possible communal error', would leave room for correspondingly weaker versions of CT and RT. The principled cut-off point, it is now suggested, is provided by following Wright's distinction between the two cases outlined above and opting for the weaker version of 'possible error'.

If this suggestion is acceptable, then the dilemma outlined in the preceding section can be avoided. CT is retained but the CT Argument is amended: the communal standard of correctness picked out by CT becomes a general standard of correctness of the sort described above. So although the community cannot be thought of as
going wrong generally, on any particular occasion the community can be possibly mistaken. RT is likewise retained although the RT Argument also survives only in an amended form. RT still implies that the community can be mistaken. But such mistakenness is limited to a restricted range of occasions. This implies that the community cannot, unlike the individual, ever be wholly wrong. In other words, just because the standard of correctness the individual faces is now constituted by community usage, there is still a distinction between the individual case and the communal case. It still makes sense to think of any particular individual as being possibly always wrong, relative to the community. But the community itself is only ever, even in the worst case, wrong on a limited number of occasions.

A fresh difficulty: no 'point of comparison'

What this position is supposed to yield, then, is a notion of the community being able to go wrong which can be combined with the community-constitutes-correctness notion. The solution is provided by restricting the extent to which the community can, in terms of RT, be thought of as going wrong while limiting the range of occasions of communal usage which, in the fashion outlined by CT, are determinative of correct use. This position, if correct, might then represent a starting point for an explanation of language understanding in terms of rule following.

There is, unfortunately, a danger in placing too much emphasis on the distinction drawn by Wright. The picture is one in which the community can be seen to guide itself. But the difficulty with
this picture is that it has, at best, only the appearance of accommodating to RT. In fact, it is hard to see how the community's position, on such an account, differs from that of the speaker of a 'private language' - a position which rules out any possibility of the kind of regulation of usage demanded by RT. [19] On the account just offered, what is supposed to separate out the position of the 'private language' speaker from that of the community is that the latter can be thought of as going right or wrong on one occasion in comparison with how it goes on other occasions, whereas in the case of the former (according to the Wittgensteinian, at any rate) there is no such comparative distinction to be drawn. But it is, in fact, hard to see just why the community should be thought of as faring better than the 'private language' user.

The problem is that we must conceive of some perspective from which it can be said that relative, say, to past communal usage, present communal use of a given expression is mistaken. What is required is some point of comparison from which it can be determined whether present use is (incorrectly) different from that which constitutes correct use (e.g. instances of previous use). But where can this comparative perspective be located? If we are to conceive of it as external to the community's language practices, then such a perspective would, of itself, directly provide a standard of correctness, independent of communal language use (and contrary to CT). Independently of the way the community may suppose correct usage to run in the present instance, an external point of comparison might reveal that such
usage is incorrect relative to previous usage. There would be no sense in which past communal usage could, of itself, be seen as solely representing the standard of correctness. For the question of whether present usage is the same or differs from previous usage is one which would be settled by such external comparison.

External comparison would rely, on other words, on an authoritative statement of that which constitutes correct communal usage in the past which can be appealed to in drawing such comparisons. But if this were the proper way to take the idea of the community guiding itself, it is hard to see how this would represent the requisite weakening, in the sense discussed above, of the RT thesis. For now the comparison between different communal usages seems to rely upon a statement of what constitutes correct usage. And such a statement, once formulated, would serve as a standard by which, in terms of the original, strong version of RT, one might express the possibility of the community going completely and irrevocably wrong in its use of the relevant expression.

This idea seems to run counter to almost all of Wittgenstein's thoughts on the matter. And yet, on the other hand, if the perspective is to be located within communal language use, it is difficult to see how it provides the requisite function. The perspective, it will be remembered, is to provide a viewpoint from which present communal use can be compared with some informative statement of relevant previous usage. It is supposed to allow for cases in which present practice is, relative to previous usage,
deemed to be incorrect. But if this perspective just is one of the actual occasions of communal language use, there is no sense in which the view from such a perspective offers a check on variation in language use. On the relevant occasion, present usage will be deemed correct and views on the community's uses of the relevant expression in the past adjusted accordingly.

The matter can be put another way. We intend to augment the communal account by introducing a regulation element. This element is based on a distinction between those communal uses which constitute the standard of correctness and those other communal uses which are mistaken. (The thesis that communal usage is 'weakly' constitutive of correctness is supposed to guarantee that there is such a distinction, thereby guaranteeing an element of regulation within the account.) The picture offered is a view of communal usage throughout time which is regulatory inasmuch as certain uses are seen to be incorrect relative to certain other uses. But the difficulty is that the distinction between occasions of communal usage thus drawn seems inadequate to support the idea of regulation. The aimed-for contrast between correct and incorrect usage must be drawn by the community itself (otherwise it is hard to see what the point of the communal thesis is). But, if drawn by the community, the 'contrast' now looks nugatory: whenever the community thinks it is correctly continuing its previous usage, it is correctly continuing its previous usage.

The difficulty analysed: 'personality' as a problematic metaphor
This is where the analogy with individual personality breaks down.
Chapter 2

The example of an individual's personality relies upon the idea of someone (either the individual concerned or some other) being in a position to make a comparative judgement about a particular item of the individual's behaviour, to the effect that it is, or is not, characteristic of that individual's personality. Such a judgement relies on an explicit or implicit comparison between, say, the behaviour produced now and behaviour produced in the past, or between behaviour produced now and a current expectation of behaviour (an expectation formulated, perhaps, by past experience). Whatever the actual nature of the comparison, it will be grounded in some independent set of principles or guidelines which establish the way such comparison works - what counts as more unpleasant behaviour, what counts as another instance of vicious behaviour, and so on - which is such that one could, at least in principle, formulate a statement of why a specific behaviour is, or is not, characteristic of a general type of behaviour and is, thereby, characteristic of a given individual.

Now the difficulty here is that there seems to be no such set of principles or guide-lines which allow for an equivalent form of comparative judgement in the case of the community as a whole. There are no independent resources through which one could state what is to count as the community's correctly remaining 'in tune' with its own, communally-set standard - if there were, this would breach CT. But to the community itself, its own usage is bound to seem the correct extension of previous usage, in which case there is no sense in which the community can be thought of as gauging its own use against previous use, independently of how it feels it ought to go on.
Of course it is tempting to point out that the community is still in a position to issue judgements about previous communal usage. The community, it might be admitted, cannot question its own current usage, but this does not prevent it from criticizing its own earlier usage. And, it might be urged, the possibility of such criticism suggests that there could still be some distinction to be drawn between different sets of communal uses - the community will regard its present use as correct, but may regard certain previous uses as incorrect. In this respect, it might be pointed out, the community can be thought of as gauging past use against present use and this guarantees that there will some kind of comparative process analogous to that undertaken in the case of the individual and his or her personality. There will, it might be concluded, still be some distinction to be drawn between right and wrong. But the question now is whether this is a distinction which makes a difference.

Note, to begin with, that there is a curious ambivalence in both the individual's personality case and in the case of communal usage of a linguistic expression. In both cases, we can separate out what might be termed a 'de dicto' reading of matters from a 'de re' reading. On a 'de dicto' reading of the individual's personality, we say that a current item of behaviour is out of character if it does not accord with the individual's character as captured by some description. If we take the individual's character to be that of a kind person, for example, then should we see the individual kick the cat up and down the hall, we would probably be forced to conclude that this is an act which is out of
character. A similar 'de dicto' conception could be developed for the notion of communal usage over time. We might say, for example, that when the community uses 'red', it means red, as individuated by some descriptive process. But on such a 'de dicto' reading, though, we already have a ready-made account of how the community can be right or wrong in its usage. The community must be viewed as going wrong on any occasion in which its usage does not accord with the supplied description. And the difficulty with this is that it presupposes the very notion we seek to explain – the notion of usage being regulated in some way – by stipulating that there will always be some authoritative description under which communal usage is to be subsumed. This suggests that the 'de dicto' reading does not offer an appropriate account of the required distinction.

On a 'de re' reading of individual personality, if the individual kicks the cat up and down the hall, there is still scope for raising the question as to whether this behaviour is in character. On the one hand, we might decide that this behaviour is sufficiently unlike previous behaviour that it is 'out of character'. Alternatively, we might decide that this behaviour is so revealing that it is determinative of character (even if the character thereby determined is not that suggested by analysis of previous behaviour). Since there is no suggestion here that what counts as 'in character' is already decided by some authoritative description, this might imply that the 'de re' reading is apt for the communal linguistic usage case. But this is where the dis-analogy between the individual personality case and the communal
linguistic usage case bites. In the individual personality case, we have a clear conception of what the 'de re' reading does: it is set within a background against which it already makes sense to talk of 'deciding whether current behaviour is in character', 'making judgements about character in the light of current and previous behaviour' and so on. This background centres on an overview or summary of past and present behaviour which, although it jointly constitutes 'the personality' of the individual, can be specified or informatively described independently of that personality description and which, thereby, acts as a standard of comparison. What, in the communal linguistic usage case, does the analogous 'de re' reading do?

The temptation is to think that the 'de re' reading similarly sets out a background against which it make sense to talk of deciding whether current communal usage is in accord with, or differs from, previous usage - a background which centres on an overview or summary of current and previous usage constituting 'the meaning' of the linguistic expression under discussion. But the difficulty is that, in the communal case, there is no sense in which this overview or summary of usage could act as a comparison. Just because there is no independent description of communal usage which does not already imply that one rule or another is being followed, when the community comes to make its comparative judgements these will be acts of 'judgement' without a content. In the 'de re' reading, nothing about 'the meaning' settles what counts as the similarity or difference in usage, and so the community's decision as to whether previous usage is correct or
incorrect relative to present usage is not guided: whatever will seem the correct decision to the community is the correct decision.

Back to the drawing board

The goal in this section was to weaken the CT and RT theses in such a way that they could be viewed as compatible. To this end, it was suggested that the two theses could be viewed as analogous to an individual's personality: setting out a general standard which allowed for exceptions. This meant that there should be some perspective which allowed for comparison to take place. But if this perspective were external to communal usage (the perspective favoured on a 'de dicto' reading), RT would be over-emphasised at the expense of CT. If the perspective were internal to communal usage (the perspective favoured on a 'de re' reading), CT would be over-emphasised at the expense of RT.

But note that this argument supposed that either

(i) the community possesses a reflective knowledge that its own standards (externally) imply such and such

or

(ii) there is some (internal) perspective from whose vantage point one can know that such and such, being the community's most frequent (or whatever) relevant usage, constitutes the relevant standard of correctness required.

In other words, both arms of this dilemma view the idea of a standard of correctness in terms of substantive knowledge. This shows that the 'Wittgensteinian' model used in the preceding
argument shares the very 'knowledge that' bias for which the accounts in Chapter 1 were criticised. In our attempt to provide a model of understanding which challenges the meaning theory approach, we have smuggled in the very concept which lies at the heart of that approach.

In fact it can be shown that the difficulties which arose from our attempt to combine CT and RT arise automatically for any rule-following account which relies on a 'knowledge that' framework. Consider CT first. Suppose the account tries to express the CT standard of correctness in terms of 'knowledge that'. Such an account, it will be remembered, must describe that which the competent speaker knows and state that in which possession of such knowledge consists. In a rule-following account, the description of communal usage will involve stating the rule involved. So, loosely borrowing from the meaning theoretic framework, we might provide the requisite description by stating, for example, that the community's usage follows the rule: apply 'green' only to green things. To make the theory substantive, however, we must also state that in which knowledge of how to follow the relevant communal rule consists. [20] We require an explanation of how the rule is understood and acted upon by the community.

Such an explanation requires an unambiguous statement of the way the community interprets the rule. And the difficulty with this is that, according to CT, there can be no final, authoritative interpretation of the rule. If there were, then this interpretation could be used to state that in which following the
rule must consist - *even if the community persists in doing otherwise*. But the inclusion of CT is supposed to rule out the possibility of the community, in toto, deviating from correct usage. What this implies is that, so long as CT is retained, there can be, in the required sense, *no* such authoritative interpretation. But if there can be no such interpretation, there will be no way in which substantively to account for that which a competent speaker knows in knowing how to follow the relevant rule. [21] So long as CT is retained, the 'knowledge that' format must be rejected.

Moreover, the RT thesis states that rule-following depends on the idea of regulation in terms of some form of standard of correctness. Even communal usage of an expression must be thought of as somehow responsible to a standard of correctness. Now according to CT, the community's use is itself involved in setting this standard. If the 'knowledge that' framework is adopted, this means that in order to understand an expression, one must know that it is to be used in the way the community uses it. But if this is all that is to be known about the correct use of the expression, then whenever one knows what the community would do (e.g. by knowing what the majority would do) one knows how to apply the expression. In such a case, whatever the community does is correct. But this implies, contrary to RT, that the community has no standard of correctness which its own usage must respect. The conclusion required is that an account which contains RT cannot utilize the 'knowledge that' framework.
Chapter 2

So both CT and RT seem to be inconsistent with a 'knowledge that' format. And yet the model offered above for combining CT and RT relied upon just such a format. This implies that the difficulty raised - that either CT or RT comes to pre-dominate in any such model - came, not from the compromise attempt to combine CT with RT, but from setting this compromise position within a 'knowledge that' framework. What is demanded is an alternative conception of CT, RT and their interrelationship.

§4 CT/RT COMPATIBILITY II: AN 'INTERNAL RELATIONS' ACCOUNT

The overall goal of this chapter, it will be remembered, is to explore the extent to which Wittgenstein's account of rule-following offers an alternative to the knowledge-centred perspectives of Realism and Anti-realism outlined in Chapter 1. The difficulty with Wittgenstein's account, however, was that it seemed to lead to a dilemma: a Wittgensteinian ought (pace Kripke) to retain some version of both CT and RT; and yet these two theses seem to lead to inconsistent conclusions. Rejection of RT does not work, and rejection of CT underruts the whole 'meaning-as-use' programme. Attempts to weaken these theses led to a position which was not much better than outright abandonment of RT or CT. What was needed, it was decided, was an alternative account of rule-following in which RT and CT can be seen to be genuinely compatible. One such perspective is offered by the work of G.P. Baker and P.M.S. Hacker.
In a number of works, Baker and Hacker have themselves attempted to build regulation and constitution elements into a Wittgensteinian rule-following framework. [22] According to these authors, far from being potentially contradictory accounts of 'correctness', the contents of the RT thesis and the CT thesis deal with matters which are somehow part of a unitary whole:

'We talk correctly... of the rule as determining what steps to take. But we are inclined to misconstrue this idiom as if the rule in some mysterious way already contained all its applications independently of us. The truth is that we fix these steps as what we count as being in accord with the rule.' [23]

The Baker and Hacker view is that for there to be a state of affairs describable in terms of the RT just is for there to be a state of affairs describable in terms of CT. When it is correct to talk of a rule determining future steps, it is, in virtue of that fact, equally correct to talk of our fixing what is to be regarded as the correct step.

To understand how such seemingly disparate accounts of correctness come together in this way, say Baker and Hacker, it must be realized that the essence of a rule-following account is its use of the notion of internal relations. Such relations have a number of features: (i) if an internal relation holds at all, it holds necessarily; (ii) if A and B are internally related, it is inconceivable that they should not be thus related; (iii) this relation is not open to analysis, in that it is not possible to consider breaking down the internal relation between two entities into a pair of relations each of which holds between one of those
entities and a common third. Much of what Baker and Hacker say about rule-following can be understood by concentrating on this third aspect of internal relations.

On their view, it is correct to say that a rule determines what is to be the correct future use of an expression - but 'determines' here must not be misconstrued. All that is implied is that the rule determines its applications in the same way that the presentation of the obverse of a coin determines that the other side of the coin is the reverse. [24] When one describes a set of activities by saying that the participants are following such and such a rule, the description does not work through the identifying of some entity (the rule) which is in principle separable from those activities. In particular, Baker and Hacker urge, it must not be thought that the rule and its applications are separable to the extent that they require to be related by a third entity such as an interpretation or inductive step, for this would result in a kind of fracturing of the internal relation which might lead to an infinite regress. [25]. Of course, it does not follow from this that a rule and an application of that rule are the same thing (any more than it follows, from saying that obverse and reverse are two sides of the same coin, that the two sides of a coin are the same thing). A rule is a standard of correctness while an application of a rule is the use of the rule-governed term in actual discourse.

So according to Baker and Hacker, it is true that something akin to RT is involved in the rule-following account although the talk
of 'determines' in that thesis must be taken in the correct fashion. [26] Baker and Hacker also employ a variant of CT, which centres on their claim that it is we who 'fix' what is to count as accord with a given rule. The rule may determine correct use of an expression - what is to count as a correct application of the rule for that expression - but this is not an issue separable from an account of those of our activities which involve such rules (activities which include teaching, explaining, correcting and so on, together with practices of actually applying a rule).

This appeal to practice is not, however, a reference to foundational matters upon which talk of rules and applications rests. The difference between a practice and a regularity of occurrences is that the actions which constitute a practice exhibit normative regularity. The only actions which possess this normative regularity are those which are carried out with the intention, on the part of the agent, of sticking to the pattern which is discernible in those actions. [27] If an individual is to satisfy us that his actions constitute a practice, those actions must demonstrate that he understands the technique involved in applying a rule - that he recognises the criteria of correctness associated with that technique. (It is for this reason that we are unwilling, typically, to describe a set of behaviours as rule-following unless they are embedded within a relatively complex context of explanations, exemplifications and so on.) The practices of using a particular rule, of correcting others in its use and so on are, then, themselves internally related to that rule. And the chain of internal relations which holds between
rule, application, practice and so on forms a circle of normative concepts which cannot be broken. It would be a mistake, therefore, to think that when we say 'we fix what is to count as accord with the rule' we are grounding what is to count as an explanation of the correct application of a rule by appealing to the 'majority verdict' or whatever. [28] Rather, all that CT tells us is that one can re-cast talk of rule-following in terms of talk of following a practice. In this sense, the Baker and Hacker account is an explicit rejection of the 'knowledge that' formulation attacked in chapter 1 and illicitly relied upon in the previous section.

Thus when Baker and Hacker say that we fix what is to count as accord with the rule - that our practices produce the rule/application internal relation - what they mean is that rules themselves are not separable entities. [29] In terms of the RT thesis, it is acceptable to say that a rule determines what is to count as correct applications - but only in virtue of the fact that rule and application are internally related. On the other hand it is equally correct to say, in terms of the CT thesis, that human practices determine correct application - but only by virtue of the fact that rule and practice are internally related in a similar way.

Baker and Hacker's solution to the problem of how to combine RT and CT should now be clear. It is true that one can talk, in terms of RT, of the rule determining what is to count as correct use of an expression - but only in virtue of the fact that rule and
application are internally related. It is also true that one can
talk, in terms of CT, of human practice fixing what is to count as
correct use of a term. But, once again, this only holds in virtue
of the fact that our practice is not a separable 'third entity',
independent of rule and application. To talk of a practice is
merely another way of talking about actions which form an
established pattern, such that they have become responsible to a
particular criterion of correctness. So it is possible to
consistently maintain both RT and CT; they are merely different
ways of describing the same set of internally related concepts.
Indeed, one could not be said to have a correct grasp of the truth
of one thesis until one had grasped the truth of the other. Taken
together, they represent a closed, internally related circle of
normative concepts.

The Problem With Internal Relations

Of course, the difficulty with filling out CT and RT in terms of a
'normative circle' of concepts each internally related to the
other is that the account comes to seem uninformative. There is a
temptation, for example, to construe the Baker and Hacker argument
as: one explains what it is for someone to use an expression
correctly by stating that the expression must be employed
according to the rule for its use, while one explains what it is
for an expression to be employed according to the rule for its use
by stating that to follow the rule is to use the expression
correctly. Formulated in this way, the Baker and Hacker account
will come to seem explanatory only to those who already regard
talk of 'grasping a rule' as perspicuous.
In the case of language learning, for example, if one explains this phenomenon by stating that the language learner acquires a grasp of the rule for the use of an expression, it does not seem to be a further explanation to add that 'acquiring a grasp' is the same thing as learning to pick out occasions where the term can be correctly used – for, in virtue of their being internally related, one would only have understood what correct use is when one had understood what grasp of the rule is. [30] So the difficulty with the Baker and Hacker account is that their explanations of language-learning and so on seem bound to turn in on themselves in an unsatisfactorily trivial manner.

The essence of a rule-following account of language is that it explains meaning in terms of rules for the use of expressions. So the question of how a language learner learns the meaning of an expression becomes: how does he or she acquire a grasp of the rule? Now the worry here is that, if the Baker and Hacker account of rule-following is the correct one, rule-following accounts might seem to offer no explanation of this phenomenon, and thereby fail to offer an account of the essential difference between the competent language speaker and the novice. To see why this is so, consider the following relatively straightforward explanation.

A language learner gains an ability to use a certain expression correctly; but this ability is restricted to a limited set of basic circumstances. He or she then develops a grasp of the rule for the use of that expression by some further process such as the
formulation of an inductive hypothesis as to what the rule would have to say in a range of different, as yet unmet, circumstances. Once this further stage has been safely negotiated, the language learner can be said to have learnt the meaning of the expression. Now, however this might stand as an explanation in its own right, it has the merit of giving a clear-cut interpretation of what might be meant by grasping a rule. But it is clear that this approach is denied to anyone who emphasises the internality of the rule/application relation. If what counts as a correct application of a rule is internally related to that rule, then it seems clear that one cannot grasp what it is for something to be a correct application without, thereby, grasping the rule. And this excludes the idea of the language learner having the ability to employ an expression correctly, in a limited set of cases, prior to acquiring an understanding or grasp of the rule for its use.

So what Baker and Hacker deny to themselves, then, is the possibility of making out the concept of grasp of a rule in terms of the idea of a prior ability to employ the rule-guided expression correctly in a limited set of circumstances. If rule and correct application are internally related, then there can be no concept of a prior grasp of the one with which to make sense out of the idea of grasp of the other.

The story so far: Baker and Hacker claim that the key concept in a rule-following account is the idea of an internal relation. In particular, it must be understood that rules are internally related to their applications. Because of this internal
relationship, it is correct to talk of rules determining their applications. But it is equally correct to talk of our normative practices such as explaining, correcting and so on determining the correct applications. Our practices determine what is to count as correct application of a rule because it is through our normative activities that the internal relation between rule and application is created. These features of our rule-guided activities come together to form a 'normative circle' of concepts. The difficulty with this account, however, is that it seems to restrict the analyst's conceptual resources in offering explanations of the difference between the presence and absence of rule-guidedness.

The response by Baker and Hacker to this worry is to stress the importance of rule formulations. As has already been made clear, an application of a rule for the use of an expression can be thought of as the occurrence of a correct employment of that expression. A rule formulation, on the other hand, can best be thought of as an item of behaviour, either verbal or nonverbal, aimed at the expressing of a rule. Now in virtue of this fact, say Baker and Hacker, it follows that a rule formulation cannot also be a rule application. The two kinds of activity are mutually exclusive. [31] However, since one and the same form of words, or the same nonverbal gestures, can be used as either formulation or application of a rule, it is clear that the difference between a formulation and an application must be explained by reference to the context in which the form of words or gestures are employed. A rule formulation, say Baker and Hacker, is only a formulation in virtue of the use to which we put it. [32] And the contexts of use
in which a form of words becomes a formulation of a rule are
normative activities - occasions on which the rule is being
explained, or on which a novice is being trained in the use of the
expression guided by the rule, and so on. [33] Indeed, a vital
step in gaining an understanding of what it means to say that we
fix what is to count as accord with a rule is grasp of the fact
that we employ rule formulations in the context of normative
activities.

So Baker and Hacker claim that the language learner learns his
language (in part at least) by means of formulations of the
language rules. Someone can acquire the grasp of a rule if a
competent speaker formulates the rule for him. But this merely
resurrects the original difficulty. For the difficulty with this
claim, of course, is that it introduces the task of explaining how
such rule formulations represent substantive training techniques
for the language novice. Rule formulations, like rule
applications, are internally related to the relevant rule. So what
is needed now is an explanation of how the language novice
correctly apprehends a rule formulation which is an activity
neither separable (as a third entity) from, nor more basic than
the following and applying of rules itself. But such an
explanation will not, on the Baker and Hacker model, be easy to
supply. For, just because rule formulations are internally related
to the other rule-following activities, anyone who finds those
activities difficult to understand ought to have the same reaction
to their internally related cousin, the rule formulation.
Another look at the CT/RT dilemma

We began this chapter by questioning the framework of explanation used in Chapter 1. The suspicion was raised that the 'knowledge that' model might prove too restrictive for a proper account of language use. It was decided to attempt the provision of an alternative account of language understanding, based on Wittgenstein's notion of rule-following. The difficulty with this suggestion, however, was that this notion seemed to rely on two theses - CT and RT - which are inconsistent. §2, §3 and §4 attempted to surmount this obstacle but without success. In particular, in §3 it seemed that if either thesis is weakened, the other comes to dominate whatever account of rule-following is left. But this, it was discovered, stemmed from an illicit use of the very 'knowledge that' concept which was supposed to be supplanted. This concept was explicitly rejected in the 'internal relations' account offered in §4. But this account relied upon the idea of a 'normative circle' of concepts which allowed no way in which to explain language learning. So far, then, no advance has been made in the production of a rule-following account.

To remedy this, we must start afresh. We begin the attempt to construct a Rule-Following Account (RFA) by restating what constitutes following a rule for the use of a particular expression in terms of CT and RT. The standard of correctness represented by the rule is constituted by communal usage in the following way. We gauge a language-learner's ability to employ a
term correctly by the extent to which he matches our own applications. We expect him to follow our explanations properly, to respond appropriately to our corrections, to be able to apply the term in novel situations in the way we feel we ourselves would, to offer explanations of the right sort in response to our queries and so on. Explanations and the like are not, though, a complete formulation of the expression's assertibility conditions of the kind aimed at by meaning theorist's. It follows that an explanation, for example, does not settle the issue of how the rule is to be applied by stating necessary and sufficient truth conditions (or necessary and sufficient assertibility conditions). In one context, the explanation will be seen to determine one course of action, in another context, it will be seen to recommend a quite different course. In order to work properly, then, the explanation must be taken in the right way. But nothing in the explanation itself guarantees this.

Individual and communal usage is still regulated, however. If the community uses a specific explanation in teaching novices a rule, it regards that explanation as settling the issue of how the rule is to be applied. For it is a natural fact about human beings that, faced with a given explanation in a particular context, they typically agree in seeing that explanation in the same way. It is similarly a natural fact that people agree in what they see as 'going on in the same way in following the rule', and so on. It is this natural fact which gives explanations their regulatory force. From the perspective of this agreement in judgement, we cannot but
see particular explanations as meaning that, in a given context, the relevant rule should be applied in one specific way.

But if explanations do not offer necessary and sufficient conditions for the application of the rule, neither does the combination of an explanation together with a specification of the agreement in judgement in operation. For it is always possible to provide different interpretations for such specifications. This raises the question of how a natural fact such as the fact that we agree in judgements can ensure that we take an explanation in the same way. How, in other words, can a natural fact such as 'agreement in judgement' have a normative consequence? (34)

**Agreement in judgements: a new dilemma?**

As an example of agreement in judgement, consider the following:

"People agree in seeing a certain range of responses produced by a language learner as showing that he understands some specific explanation."

Such an assertion is supposed to be a contingent statement which has normative consequences. For an assertion to be contingent and yet have such a normative outcome it must have two features:

(i) it must be capable of denial without contradiction

(ii) it must help distinguish between correct and incorrect usage so that if the language learner does such and such, people have no option but to accept that he understands the explanation.

The implication of (i) is that it makes sense to imagine people disagreeing in seeing the 'such and such' range of responses as
correctly following the relevant explanation. (ii) implies that people cannot regard the 'such and such' range of responses as other than proof that the language learner understands the explanation. (i) and (ii) taken jointly therefore seem to generate an inconsistency. On the one hand, if the assertion is contingent it should be possible to deny, without contradiction, that if the individual involved does such and such, he understands the explanation. On the other hand, we expect the relation between explanation and application to be somehow necessary - to deny that he will do such and such just is to deny that he understands the explanation.

It seems, then, that even in RFA we are still on the horns of a dilemma. (i) seems to be incompatible with (ii), and yet both are required for the idea of agreement in judgement. If we reject either the contingency or the normativity, we lose the idea of agreement in judgements and so lose the alternative account of rule-following. If we retain both contingency and normativity, we end up with an inconsistent version of agreement in judgements and so once again have to abandon RFA. Such is the tension which seems to be created by the appeal in RFA to the idea of a natural fact - agreement in judgements - which has normative consequences.

Defusing the dilemma

But, to say that our agreement in judgement is a contingent matter is not to say that, for us, the relationship between an explanation, say, and applications of the rule thereby explained is a contingent relation. It is a contingent matter whether we
have the agreement in judgement that we do. So it is a contingent matter that we apply the expression in the way that we do. It is, therefore, a contingent matter that we follow the rules that we do. But this does not imply that, following the rule that we do, whether we apply the expression in such and such a way is a contingent matter. For, given our own agreement in judgement, to follow a certain rule just is to apply the relevant expression in such and such a way. From the perspective of our agreement in judgements, we see a particular explanation as forcing a certain set of expectations upon the language learner. If he does not meet these, then this, to us, constitutes a lack of understanding of the associated explanation. It is in this sense that the link between rule and application can be thought of as, from the perspective of our agreement in judgement, 'grammatical'.

We can see this by supposing that agreement in judgement is a property attributable to me. As such, it is only contingently true that I have such a property, just as it is only contingently true that I weigh 154 pounds. But the former property, unlike the latter, conditions certain of my reactions to the world. In virtue of having this latter property, I view the world as having normative constraints: certain actions seem, to me, the correct way to react. It is in this sense that the natural property attributed to me has a normative outcome. It is, in turn, this normativity which underpins rule-following. Given such an agreement in judgement, I am constrained to view certain ways of continuing to apply a rule as the correct way to go on.
Of course this normativity is 'perspectival', or relativistic, in that it is only from my perspective that going on in such and such a way is seen as determined by the rule. There is no statement of the rule which, independent of such a perspective, is such that the rule necessitates going on in such and such a way. But necessity is only required if we claim that a formulation of a rule should be a substantive statement of all that is known in understanding the rule. Since we have already abandoned this claim we have no need of such necessity.

It is because we already share with one another a range of natural responses prior to language - for example reacting in some respect in the same way to medium sized objects under normal conditions which are of the same colour - that the process of explanation comes to a halt. If there were not such natural agreement in judgement, then every explanation would require further explanation in order to work. Of course we could, if necessary, explain what would count as following any particular explanation properly. For example, when the language novice learns the use of 'green', it is likely that he will have gained this ability as the result of training offered by his language teacher. This training will have involved formulations intended to demonstrate how to use 'green'. Now any such formulation might be misunderstood by the language learner - it might take further explanation for the language learner to understand that formulation correctly. But this does not imply that these formulations are 'imperfect' relative to some other formulation of the meaning of 'green' offered in terms of the 'knowledge that' format. For we have
already seen that such fully substantive formulations would be incompatible with CT and RT.

So, although the process of explaining explanations is perfectly possible, it does not terminate in a substantive 'knowledge that' account of the relevant terms. In the end, such explanation must terminate in a process of 'seeing what was meant' which is independent of further explanatory analysis. This terminus in explanation means that there must be some form of agreement between learner and teacher which exists prior to language. This is what is meant by 'agreement in judgements'. [35]

It is against this background of agreement in judgements that individual and community usage can be said to be regulated. The individual must accord with communal use over a sufficiently extended number of occasions. The community, in turn, must keep faith with its own usage. The combination of an agreement in judgement plus specification of a particular rule determines, for anyone who shares that agreement, how the rule should be applied. It follows that wherever the community demonstrates such agreement in judgement it ought, from the perspective of that agreement in judgement, to continue in some particular way. If it diverges from this way then it is, by its own lights, mistaken. If such divergence occurs over an extended period, this means either that the community is no longer bound by the rule in question - either because it has adopted another rule or because, from the perspective of someone whose usage does not so diverge, the community fails to display agreement in judgement.
This, then, is the sense in which we should view the compatibility of CT and RT and the consequential 'weakening' of the CT and RT theses (in which communal usage is seen to represent a standard which the community itself must obey). Communal usage is a standard of correctness. It only makes sense to view communal usage as a standard of correctness at all where there is already agreement in judgement - there is neither an external nor an internal viewpoint of the kind mooted in §3 from which the standard can be substantively stated of the kind employed in the previous section. But where there is such agreement in judgement, we are already disposed to see the explanations, corrections and so on associated with a particular linguistic expression as pointing to a certain way of applying that expression. And this is true even for the community as a whole. So, just because the usage of the community as a whole constitutes a standard of correctness, even communal usage can be thought of as subject to regulatory influences.

Two further criticisms

There is, of course, what will seem to some a cost incurred with adopting RFA. What is on offer now is an account of the understanding of language which states that to understand the meaning of an expression is to understand the rule for its use. But RFA rejects the idea that grasp of such a rule could be, in all cases, substantively described in terms of 'knowledge that'. Now this raises the question of what such a philosophy of language account is supposed to do. A theory of meaning aims at explaining the way we use language by formulating expressions of items of
knowledge which could explain the abilities we typically display whereas RFA seems, at present, to offer nothing by way of alternative.

But this is a misleading way of putting things. RFA's emphasis on rule-following usage establishes that ordinary explanation is the paradigm in establishing the meaning of a linguistic expression. This is not, though, the claim that we can discover what an expression means by finding out what the ordinary person in the street says that it means. First, the ordinary person may not command a clear view of his own practices involving present use of that expression. Second, the ordinary person may discover that his present usage leads, in novel cases, to contradictions. It is for cases such as these that RFA is designed. The goal of such an account is to re-present explanations, reminders of usage, corrections and so on in order to elucidate what ordinary usage is and why, in specific circumstances, it can generate contradiction. Such a re-presentation will include reminders as to the point of our use of the relevant expression - a reminder, from the perspective of our agreement in judgements, of the range of natural interests and goals which are associated with our uses of that expression. This is the aim of the philosophical method of imagining different language-games. [36]

Nevertheless, it might seem as though two important facts are being ignored. First: the seductiveness of the theory of meaning approach rests largely on the fact that it promises to explain how we use expressions which we have learned in a limited range of
contexts in novel contexts. What seems to be offered in response, by RFA, is the mystical suggestion that we just do know how to go on in the future. Second, the theory of meaning offers an analysis of the way the structure of a sentence, by somehow reflecting the structure of the world or the structure of our evidential relationship to the world, determines its meaning and of how this is instanced in our abilities. RFA seems to have nothing to say about any of this.

In response to the first claim, we can note that the air of mystery surrounding the idea of primitive agreement in judgements, which underpins our explanations of being able to go on in the same way with an expression, exists only because the meaning theorist attempts to provide a philosophical explanation for a natural fact. If it were not already a fact about us that we possess some form of agreement in judgements, we would not have the rule-following practices of 'going on in the same way' which we do have. The second claim will take more time. The answer to this second challenge is that the account on offer does have something to say about what such structure-reflecting abilities consist in. The model which will be offered depends on the notion of criteria. However, before introducing this notion in the next chapter, it is worthwhile noting the contrast between RFA and the accounts given in §2, §3 and §4.

The contrast with Kripke's scepticism

Although we criticized Kripke's positive account of language usage, nothing was said about the strategy which underlay his
negative, sceptical programme. But any such programme, which attempts to demonstrate that normativity in the sense captured by RT is to be rejected must, according to RFA, be somehow mistaken.

Kripke's central claim is that meaning is a non-factual affair. His sceptical argument is intended to show that there can be no justification for saying that what I intended in the past settles, now, how I should apply a specific rule. An immediate response to this claim is that, according to RFA, it mischaracterizes the relationship between explanations and applications. If an individual offers up a certain range of explanations, clarifications and the like, then this settles, from the perspective of our agreement in judgement, that in the past he meant to use the relevant expression in the way that we would use it. The response to this put forward on Kripke's behalf was that an appeal to the 'grammatical' nature of the link between rule and explanation only pushes the sceptical challenge one stage back. How, the sceptic now wants to know, can we be sure that we correctly interpret the explanation?

Now since we have already claimed that 'explanations come to an end', we might think to appeal to agreement in judgements. But of course in Kripke's account, there is also agreement in language use - it is a matter of fact that we do not fall to blows over the correct application of 'red' and so on. It is this agreement in judgement which ties together our practices of using expressions, explaining them, providing justifications and the like. But this agreement in judgments as to redness or whatever is a contingent
affair. It follows, the Kripke-style sceptic would conclude, that we cannot ensure correct uptake of a given explanation by appeal to such agreement.

The sceptic's point here would be that wherever appeal to a contingent fact such as agreement in judgement is made, one can always, without fear of contradiction, raise a further doubt. The sceptical conclusion derived from the idea that, for example, although I claim to have meant such and such by the explanation I gave of my use of 'green' there is still uncertainty as to which colour-expression rule I am actually following. Explanations may be grammatically related to the rule which guides their use, but the fragment of applications currently accessible is under-determinative with respect to the question of which rule/explanation they instance, and no further facts, past or present, are sufficient to make up this under-determination. It follows that in offering a given explanation, nothing excludes my having a peculiar interpretation of my own explanation such that, as with the rule in general, I systematically misinterpret the effect my explanations have on those around me. Now suppose we claim that for there to be explanations about colour-term usage, there must already be some agreement in seeing instances of the same colour as the same. The sceptic will point out that this is a contingent, not a grammatical matter. It is only contingent that our agreement in judgements is as it is, while it is in virtue of our having such agreement in judgement that, in our practice, we separate out episodes as 'green' in the way that we do, So it must be possible to conceive, the sceptic urges, of cases in which
others disagree with us in judgement and separate out episodes
differently from ourselves, but by a peculiar quirk of fate offer
the same rule formulations, explanations and the like (up to the
present moment). So, just because it is a contingent fact about us
that we separate the world around us into episodes of greenness in
the way we do, the sceptic would conclude, it is possible that
others separate the world out in a different fashion, a difference
which is, because of the systematically different way those others
interpret rules, explanations and the like, invisible to us.

According to RFA, though, the appropriate response to Kripke is
that his type of sceptical doubt relies either on (a) wrongly
separating out understanding of language from agreement in
judgements or on (b) miscalculating the effect of such agreement
being contingent. For a given expression, we only acknowledge an
individual as linguistically competent once he has demonstrated
that he is able to use that expression, in a range of cases, in
the way that we would use it. But if we feel that he does use the
expression in the way that we would use it - applies it in the
same cases, offers the same kinds of explanations, responds to
corrections in the same way, and so on - we are not in a position
to raise a further doubt. We are, in such cases, unable to see
such explanations as meaning something other than what our
agreement in judgement determines it to mean.

So, as far as (a) is concerned, if we ascribe our rule-use to the
individual, we cannot at the same time doubt whether he shares
agreement in judgement with us. For to explain someone's use of an
expression by saying that he follows the rule we follow is already
to claim that he agrees in judgements with us - agreement in
judgement is prior to such joint following of rules. As far as (b)
is concerned, on the other hand, suppose we did attempt to raise a
doubt about whether the individual employs our rule at all - by
questioning whether he shares agreement in judgement with us. We
might suggest that he has a range of natural responses quite
different from our own, the suspicion being that his nature is
radically different from our own, and so his rules are
 correspondingly different, but that such differences have yet to
emerge. But this doubt is also ruled out. Provided we gather an
appropriately wide range of responses from the individual, the
view from our own agreement in judgement settles for us which rule
he is following. We are naturally disposed to seeing certain
explanations which the individual might offer, for example, as
determining a certain path of future usage. And although this is
only a contingent fact about us, it nevertheless imposes a
particular viewpoint upon us. It follows that if the individual
produces what we regard as appropriate responses over a range of
cases, we are committed, by the natural features which underpin
our agreement in judgements, to seeing such responses in terms of
activity guided by a particular rule.

It is this idea which establishes the regulation element so sorely
missed in Kripke's positive account. Kripke wanted to have it that
our agreement in inclinations was enough to establish the concept
of reliability in future response. It was this notion which gave
point to the whole idea of ascribing understanding to others. But
agreement in inclination proved too weak a notion: reliability in response cannot be established merely by pointing out that present inclinations coincide, unless such present coincidence conditions the way we think about the linguistic capacities of other individuals. It is this latter notion which RFA captures. Agreement in judgement has a normative element which licenses talk of reliability, un-reliability and so on. This stems from the fact that, in RFA, agreement in judgement conditions the way we respond to others' actions in such a way that we see present agreement in inclination as proof that future inclinations will remain in agreement.

The contrast with 'limited communal error' accounts
The major contrast lies, of course, in the fact that the 'knowledge that' format which is abandoned in RFA is illicitly relied upon in the limited communal error account. However, this aside, it is also clear that RFA disagrees with both the 'external' and 'internal' perspectives adopted in §3. The 'external' perspective requires a standpoint beyond agreement in judgments which would provide a non-reinterpretable account of the rules in use in a given language community. The 'internal' perspective implies that the community as a whole has no regulatory commitments. The former claim, according to RFA, must be false: rule-following only makes sense against a background of communal agreement in judgement. But RFA also rejects the latter, internalist claim. According to RFA, the community as a whole does face regulatory commitments: from within a given agreement in judgement perspective, the question of what would constitute
correct usage is already settled - and the standard thereby set applies to the community as a whole as much as to any individual.

The contrast with 'internal relations'

It could be argued that Baker and Hacker are free to make use of the same notion of 'agreement in judgement' which appears in RFA. The implication might be, for example, that the language trainee shares such agreement with his or her trainers, so that he or she is already predisposed to see certain ways of 'taking' a formulation of a rule as natural. Baker and Hacker might then suggest that rule-formulations work on the basis of just such agreement. This would show how the language learner 'breaks into' the circle of normative concepts.

But if this is the notion which carries the explanatory weight in the Baker and Hacker account, we have returned to a position in which it is difficult to make out a distinction between trainer and trainee. If we augment Baker and Hacker's account in this way by appeal to agreement in judgements, the trainee will be taken to agree with the trainer in seeing certain ways of 'going on' as the natural outcome of earlier applications of an expression. But to know 'how to go on' with an expression just is, on the internal relations account, to know the rule for the use of that expression. So if the trainee agrees with the trainer in seeing certain ways of going on as natural, he already agrees with the trainer in how to use that expression. In which case the trainer's role becomes redundant - the trainee already grasps the rule in question.
Chapter 2

What is needed, it seems, is an account which allows for a limited agreement between individuals prior to language learning which still leaves room for the role of the language teacher. This marks out the difference between RFA and the account offered by Baker and Hacker. On RFA, agreement in judgement is a necessary but not sufficient condition for language learning. On the Baker and Hacker account, the internality of the relation between formulation and usage implies that agreement in understanding a rule-formulation just is agreement in use, which precludes the possibility of learning or teaching.

96 CONCLUSION

This chapter began with a rejection of the 'knowledge that' perspective which underlies meaning theories such as those offered by the Realist and the Anti-realist. The solution adopted was to explain language understanding in terms of a Wittgensteinian rule-following model. It was argued that such a model ruled out the meaning theoretic approach. It was also pointed out, however, that in such a model, room had to be made for two theses: CT and RT. The difficulty, at first, was that these theses seemed to represent competing accounts of what it is to grasp correctly the meaning of a word. The resolution of this difficulty only comes, it was suggested, when it is accepted that the theses must be taken in a relatively weak form in which the communal standard of correctness guides the community as much as the individual. In order to achieve this weaker form of CT and RT, however, it was discovered that we must make appeal to the notion of agreement in
judgement. It was this combination of appeal to agreement in
judgement and to a combination of RT and CT which established the
particular account of rule-following made out in §5: RFA.

This left the question of how an account of understanding such as
RFA is to be taken as an explication of the structural aspect of
language use. It was suggested that the key here is the notion of
the criterion. According to RFA, wherever there is a question
about how an expression ought to be used in a particular context,
one resource is to examine typical explanations, corrections,
amplifications, questions and so on that might be involved in
using and teaching such an expression. The point of such an
examination is not to recover, from language users, an account of
what they think the expression means. For it may well be that
their own usage is not, for them, perspicuous. Rather, the goal of
such a study is to gain a clear understanding of how the
expression is in fact used.

Of course, on many occasions such language use may resist
explanation - real or imagined. Our usage may, to us, seem so
obvious that there is no form of explanation or amplification
which comes to mind. It is for such cases that the language-game
methodology devised by Wittgenstein is particularly apt. Wherever
an aspect of our language usage seems resistant to any kind of
explication, we can get a clearer view of what our understanding
of that aspect consists in by imagining the way in which others
who used this aspect of language in another way would have to
differ from ourselves.
The result of such an analysis will be a clearer grasp of the way we use an expression in a given range of contexts; we will be better able to understand what it is about those contexts which make our use of the relevant expression appropriate. In particular, we will have a clearer understanding of the limits on our use of language imposed by our agreement in judgements. In certain contexts, we will recognize that no further explanation can be given as to why a given expression applies in this case. Anyone who shares agreement in judgement with us and grasps the rule for the use of that expression will concur with us in seeing that the expression, in this context, does apply.

What this implies is that there will be a range of cases in which, given our agreement in judgements and our understanding of the rules of language, certain inferences will seem, to us, natural and unavoidable. These inferences will, for us, have a special status. The conclusion of such an inference will not be certainly true, as would be the case in a valid deductive argument from true premises. But it will not be a matter of open doubt, either, as would be the case on many occasions in which one proceeds to a conclusion on the basis of evidence which is, at best, symptomatic of that conclusion. In effect, the special inferences which are associated with the limits of our language use, imposed by agreement in judgements, can be thought of as criterial. It is this notion of a criterion which the next chapter pursues.
CHAPTER 3

§1 INTRODUCTION

Criteria and the traditional sceptic

The previous chapter concluded by suggesting that an important aspect of RFA lies in the use such an account makes of the notion of the criterion. On the RFA account of language understanding, a speaker understands the meaning of an expression when he is able to use it appropriately. The specification of those occasions in which it would be appropriate to use an expression cannot be a substantive specification. But, from within the perspective of agreement in judgements, the proponent of RFA can describe the typical circumstances in which the given expression would be likely to be used. This depends upon an examination of the way the expression is actually used, an analysis of the kinds of explanations, corrections and so on that could appear in teaching a language learner such usage and the deployment of the language-game methodology in assessing the limits, for us, of such linguistic explanation. The result of this analysis is, in many cases, the isolation of a range of criteria of application of the relevant expression. [1] A good way to understand the nature of such criterial specifications is to look at the way they can be used, within the framework of RFA, in providing a response to sceptical arguments. For one point of interest about RFA is that its response to the Kripke-style meaning sceptic spreads into other areas of sceptical criticism, thereby constituting a criterialist challenge to the sceptic in those areas.
Consider an assertion to the effect that somebody other than oneself is in pain. Any account of such talk must allow for the fact that such assertions are defeasible. This has led those of a sceptical bent to draw the conclusion that second and third person ascriptions of pain must always be open to sceptical challenge. 'You may believe that X is in pain', the sceptic will say, 'but you can't be sure'. So long as a pain ascription is defeasible, it makes sense to doubt whether it is true.

Now there are a number of different ways to take this challenge, and the criterialist's goal is, in each case, to close the sceptical gap by demonstrating that the relevant pieces of evidence represent criteria for the use of 'pain'. The general tactic is to show that a description of the evidence in use is intimately related to a description of the individual as being in pain. This relation is taken to be founded in the way the community uses 'pain' in describing, teaching, explaining, correcting and so on. Evidence is criterial if it represents the kind of matter which the typical language user would rely on in providing explanations, corrections and the like.

The most basic form of the sceptic's attack is one in which the appropriateness of the 'pain' assertion is challenged. In this case, we must picture the sceptic standing, for example, amidst the carnage of war and casting doubt on whether it is appropriate to say of the victims strewn around that they are in pain, on the grounds that any claim involving 'pain' must be defeasible. The criterialist response to this initial formulation
is that such a person would demonstrate, to us, a lack of understanding of what 'pain' meant. RFA tells us that the meaning of 'pain' depends on the way the community as a whole uses 'pain'. And the community as a whole treats obvious cases of wounding and injury as typical instances of when 'pain' should be used. It follows that if the sceptic tries to avoid using 'pain' in the very cases where the community would regard its use as appropriate, we would treat his claim not so much as a sensible doubt but as an expression of misunderstanding.

The sceptic need not be in such a position for long, though. His doubt may rest, not on the inappropriateness of 'pain' in such cases, but on the fact that, appearances notwithstanding, the victims among whom he stands may really be pain free. (Science tells us, for example, that victims of massive injury often report no pain experiences.) The sceptic may admit that it is appropriate to use 'pain' in these circumstances while denying, by appeal to defeasibility, that the victims so described really are in pain. But the criterialist now will want some account of what 'really' being in pain amounts to. Clearly, we are to conceive of a set of circumstances on which the truth of 'X is in pain' depends. If circumstances are other than these, then even if it seems appropriate to say 'X is in pain', he is not really in pain. Now in such a case, we must take the meaning of 'pain' to rest on some characterization of these circumstances. For these are the circumstances which mark out correct (as opposed to merely appropriate) usage of 'pain'. This raises the question of how we rely on these 'real' correctness conditions in our usage of
Chapter 3

'pain'. We have, at present, two models: the meaning-theoretic model and RFA. The former was rejected in Chapter Two. But the latter states that the 'real' correctness conditions for the use of 'pain' are settled by communal usage, and we have already admitted that the community would use 'pain' in the present case. It seems, then, as though the sceptic is unable to make out the intended sense of 'really in pain'.

But this will not do, if only for the reason that it now becomes impossible to make out the defeasibility of descriptions which use 'pain'. If there is no conception of correctness other than that supplied by communal usage, then whenever a typical communal member would describe X as in pain, X is in pain. This clearly rules out the idea of defeasibility which seems central to speaking of pain. However, the criterialist must give some account of such defeasibility. His general response to this is to claim that one can always make out the idea of defeat by alluding to some special, atypical aspect of the background context in which the use of 'pain' takes place. This separates uses of 'pain' into two kinds: uses of 'pain' in normal contexts, and uses of 'pain' in special contexts. The notion of defeat attaches only to the second type of usage. [2]

But now the sceptic's challenge can be resurrected. [3] It seems, he will claim, as though there are now two conceptions of correctness. The first of these says that a statement is correct if it is criterially supported, while the second is the notion of correctness required to make sense of the defeasibility of those
criteria. But why, the sceptic will ask, should we accept that the relevant statement is supported by the available evidence, when we have, by hypothesis, no reason to suppose that the statement is not, in the second sense of correctness just formulated, incorrect? If the making of a statement involves a commitment that one would regard the future occurrence of certain evidence as defeating the present claim, then it seems that the criterial support is required to provide a warrant both for the present affirmation of the statement and for the belief, implicit in this affirmation, that no contrary evidence will subsequently arise. What the sceptic demands, in other words, is that the criterial evidence provide support both for the claim itself and for the background assumption that conditions are, in the relevant respects, normal – that they will not turn out, subsequently, to have been special in some problematic way.

At first sight, this looks like a strong blow against the criterialist's position. The criterialist account seems bound to fill out its notion of defeasibility by formulating two different kinds of warrant in the making of an assertion through appeal to the normal/special context distinction drawn earlier [4]:

(1) Correctness of Background
A principle of inference based upon a criterial link has its own warranting conditions – those in terms of which defeasibility is made out.

(2) Correctness of Assertion
The relevant rule determines whether the available evidence is such that one is warranted in making the associated assertion.
And now the sceptic's complaint is that the available Assertion evidence is insufficient to guarantee the Background warrant.

The criterialist's general response to this is to argue that judgements require warrant only at the level of Assertion evidence. A speaker can only properly issue a judgement which involves some set of concepts if he understands those concepts. Understanding a concept takes place against a background of agreement in judgements. To understand the nature of this judgement therefore requires agreement in judgements between speaker and hearer. Where there is such agreement in judgements, speaker and hearer alike will respond to background circumstances in the same way. It follows that so long as a hearer understands the judgement made by a speaker, he will agree with that speaker in seeing certain forms of Assertion evidence as especially good grounds for the judgement issued. And what this means is that the criterialist is able to demonstrate that the sceptical challenge cannot spread to question the normality of the relevant Background conditions - the conditions which warrant the application of the principle of inference - even though it is accepted on all sides that these latter conditions are the matters in terms of which the notion of defeasibility is made out.

Of course, matters of Background evidence can still be open for discussion - the possibility could still be raised that Background conditions are, in fact, special. It is, however, enough to make the criterialist's case if the suitability of Assertion evidence is seen to be independent of a consideration of the Background
evidence. For if Background evidence must be held to be independent of the question as to whether an assertion is correct, then no sceptical challenge to that assertion can be raised by attempting to undermine evidence of the Background kind.

But why should the sceptic accept that Background evidence is independent of Assertion evidence? The criterialist has to provide some kind of proof that Background issues must be independent of Assertion issues.

§2 TWO ACCOUNTS OF ASSERTION/BACKGROUND INDEPENDENCE

A problematic account: internal relations
The criterialist must demonstrate that Background matters are independent of Assertion matters. This highlights the difference between the criterialism of the RFA account and the form of criterialism offered by the internal relations account of Baker and Hacker. On their account, it is because we are willing to make pain assertions, on the basis of Assertion evidence (such as foot-holding behaviour), without raising the issue as to whether Background evidence is available in sufficient strength, that we regard such behaviour as internally related to pain. It follows that even though Background evidence may be defeasible, this does not spread into the question of whether a body of Assertion evidence is good enough to warrant the making of an assertion involving the relevant concept. The reason for this [5] is that, for there to be a sceptical challenge, the concept utilised in the raising of the sceptic's doubt must be clear. But each concept is
determined by the internal relation between the rule for the use of this concept and its applications and, in some cases, by the similarly internal relation between a given type of behaviour elicited by an individual and states such as pain, for example. Consequently, if there is to be a challenge at all, these internal relations must be held stable. But to achieve this stability the Background assumptions which would be involved in any attempt to formulate the rules of grammar involved must be unchallenged.

It follows from this that an attempt to raise a sceptical challenge which moves from defeasibility of Background evidence to question the suitability of the relevant assertion would fail to raise a coherent doubt at all: instead, it would merely "disrupt the concept". So in order to get the sceptical argument going, Background evidence must be treated, in principle, as independent of Assertion evidence, even though Background evidence is admitted to be defeasible. It is for this reason that the sceptic's challenge - the challenge raised via the claim that defeasibility requires two distinct species of correctness conditions - seems to miss the mark.

We can summarise this position in the following way. The sceptic traditionally relies on there being a gap between the available evidence for an assertion and the truth of that assertion - and this is taken to motivate a denial of any attempt to claim that the assertion is certain. Now on the Baker and Hacker view, there is a distinction between the obtaining of evidential conditions, even of the criterial kind, and the truth of an assertion. But
this distinction cannot motivate a doubt in respect of Assertion evidential conditions, because the conditions which are involved in the defeasibility of criteria must be independent of the conditions which settle the issue as to whether an assertion is dubitable. The independence of Assertion evidence from Background evidence rests upon a difference between questioning Assertion evidence and questioning Background evidence. We can cast doubt on an assertion if we think that the Assertion evidence is not good enough. But we cannot similarly cast doubt on the assertion by drawing attention to the defeasibility of Background evidence. If we try to raise a doubt on that basis, we disrupt the concepts involved. [6] It follows that once questions as to whether Assertion evidence is suitably strong have been settled, doubt is ruled out.

The difficulty revealed
The problem with the Baker and Hacker account, however, is that there ought not, in fact, to be very much difference in the results of questioning the suitability of the two kinds of evidence. The claim was that sceptical questioning of Background evidence may in a self-defeating way disrupt the concepts required for the raising of a sceptical challenge. But the same result may ensue, if the Baker and Hacker view holds, if evidence at the Assertion level is questioned. [7]. According to Baker and Hacker, evidence such as pain behaviour, say, is internally related to the concepts involved in the making of assertions about pain. Consequently, if someone questions whether Assertion evidence which is strong enough to be criterial really is evidence for the
application of the relevant concept, then this ought to make it unclear as to which concept is being employed in the doubt. It follows that the potential for the disruption of concepts which is said to attend questioning of Background evidence is a feature which Background evidence shares with Assertion evidence.

There is no point in denying that such an awkward question could arise. For the sceptic's traditional ally, the Realist, is likely to make just such a move. It is not, he will say, that I suspect anything awry in the Background evidence. It is just that I feel doubtful as to whether this individual exhibiting behaviour which is criterial in regard to 'pain' really is in pain - I feel that the fact of the matter may (perhaps completely undetectably) vary from the way the evidence, in this case, points. In such a case, it seems, the internal relations proponent must insist that here, too, a concept is being disrupted.

What this suggests is that, as far as the threat to the comprehensibility of the sceptic's challenge goes, the openness to question of Background evidence is on a par with that of Assertion evidence. But the independence between Background evidence and Assertion evidence is supposed to consist in the fact that the raising of questions about Background evidence disrupts concepts in a way that the raising of questions in regard to Assertion evidence does not. The fact that there is no such difference shows that, so far, the sceptic has been faced with no reason as to why he ought to accept the claim about the independence of Assertion and Background evidence.
Of course it might be claimed that this raises no problem for the internal relations account. All that has been shown is that the sceptical challenge will fail both when it rests upon a criticism of Assertion evidence and when it rests upon a criticism of Background evidence. However, this claim would still entail rejecting the picture which was offered in defence of the original Assertion/Background independence claim. The original claim was that if one wants to avoid disrupting concepts, one can question only Assertion evidence, not Background evidence. This was supposed to be what made Background evidence independent of doubts about Assertion evidence. However, we are now saying that the sceptic disrupts concepts both when he questions Background evidence and when he questions Assertion evidence. It follows that our original claim must now be withdrawn. But this independence claim was required to stop sceptical doubts spreading from the defeasibility of Background evidence to the doubtfulness of Assertion evidence. Consequently, the internal relations account must offer some explanation of independence other than that offered by Baker and Hacker. If not, it must either abandon its anti-sceptical stance, or jettison its espousal of the defeasibility of criteria. [8]

One tempting reply to this might be that Assertion/Background independence can be made out in terms of the internal/external distinction. We might claim that Background evidence conditions ought to be treated as externally related to the rules of grammar involved. Unfortunately, however, this seems surely wrong. If someone claimed to grasp the rule for the use of pain concepts,
but showed no capacity to divine cases where background evidence threw pain assertions based on, say, behavioural evidence into doubt (e.g. the presence of film cameras and film crews, clapper boards, microphones and so on), this would be taken by us as grounds for doubting whether that individual really did understand pain-talk.

The sceptic wanted to use the defeasibility of Background evidence as a springboard for a sceptical challenge. The internal relations account, offered by Baker and Hacker, suggested that this would disrupt the concepts involved. But it transpired that this threat of disruption is equally a feature of the sceptical challenge about Assertion evidence. And the difficulty this raises is that the method suggested by Baker and Hacker for differentiating Background from Assertion evidence, such that they can be described as independent, will not work. If the internal relations proponent tries to retain the notion of independence by claiming that the ability to recognise Background evidence is externally related to a grasp of the rule, then this seems to involve precisely the kind of disruption of concepts which the criterialist said was a consequence of the sceptic's argument.

Baker and Hacker do claim [9] that doubt cannot be justified by the intelligibility of Background conditions being defeating conditions, but only by such Background defeating conditions actually being present. But this surely cannot be right. The presence of defeating conditions cannot be appropriately described as a cause of doubt. If conditions obtain which defeat an
assertion, then the proper response must be to retract that assertion, not to express it doubtfully.

Of course, there could be occasions in which such doubtful expression is appropriate: occasions, that is, in which there is doubt as to the nature either of the Background evidence or of the Assertion evidence or of both. If the Assertion evidence is questionable, then this will be a straightforward case in which doubt is an appropriate response. If the Background evidence is questionable, then one will be uncertain as to whether defeating conditions obtain which undercut the (critically strong) Assertion evidence. But it is not clear, on the Baker and Hacker account, why this latter case constitutes a mere 'intelligibility' rather than a full-blown doubt. Of course Baker and Hacker will want to argue that there just is a difference between its being intelligible that such Background conditions are inappropriate and there being a question — or doubt — as to whether they are. But the sceptic will surely now want to know that in which this supposed distinction consists.

RFA's criterialist account

The difficulty for the internal relations account stemmed from the following sceptical challenge. There is no reason to suppose that Background correctness conditions are independent, in the requisite sense, of Assertion correctness conditions, and so our acceptance of statements on mere criterial evidence, must be nothing more than human rashness. The internal relations reply to this was that, were the Background/Assertion independence
abandoned, our concepts would be disrupted. But there turned out to be no reliable way in which such a claim could be maintained.

The proper response to this sceptical challenge, from the RFA perspective, is that questioning Background/Assertion independence disrupts, not concepts, but the whole practice of wielding concepts in judgement. The disruption is not that of the concept but that of the idea of justification. According to RFA, if we use an expression according to the rule for its use, we presuppose that Background matters are contingent and are yet not open to question by us - they constitute a form of agreement in judgement. Now all there is to the community's following a rule is that its applications make sense when seen against this background of agreement in judgement - this is what is meant by saying that there can be no further substantive account of language use. It follows that so long as the sceptic shares our agreement in judgements, he is unable to raise a coherent doubt as to whether, given what the community takes as criterial evidence that such and such is the proper way to apply a given expression, such and such really is the correct way to apply that expression. This does not mean that there could not be differences between ourselves and the sceptic over how to apply such a term. But such differences could only raise completely general worries about our respective usages, evading all explanatory resource. We would be forced to view the sceptic as somehow differing from us in respect of agreement in judgements. In which case there is point in saying that, from our perspective, the very idea of the sceptic's judging at all would be disrupted.
Of course the ability to employ a language rule depends on understanding when, in terms of what is known about background evidence, that rule is applicable and when it is not. So Background evidence is at least relevant to the issue of whether a given case constitutes correct application of the relevant expression. And this seems to show that the defeasibility of Background matters can, after all, generate a sceptical doubt. But on the RFA account, the relevance of Background matters stems from the way communal usage acts as a standard of correctness. It does so by representing a rule for the expression which, against a background grasp of what counts as doing 'the same' grounded in agreement in judgements, settles the issue of how to apply the relevant expression in particular cases. [10] It follows that to share the community's rule at all is already to share agreement in judgement with its members.

The key to an understanding of this distinction between Background and Assertion matters rests in the way the RFA entails a form of criterialism. We can see this by examining the way RFA represents the connection between an explanation of correct judgement and the account of meaning. Anyone interested in explaining what constitutes correct judgement must show how the process of judgement proceeds from a set of grounds, \( \{A_1, A_2, \ldots, A_n\} \), to a conclusion, \( P \). This explanation will involve providing an exposition of some principle, \( \{A_1, A_2, \ldots, A_n\}/P \), which is understood by those competent in making the relevant judgement and which demonstrates how \( \{A_1, A_2, \ldots, A_n\} \) represent grounds for \( P \). [11] What is of particular interest here is that the principle of judgement
involved - \( (A_1, A_2, ..., A_n) / P \) - may, depending on the account of meaning in use, be associated with the account offered of the meaning of \( P \). Suppose, for example, we set out the framework of explanation required in the following way.

\[
\frac{(A_1, A_2, ..., A_n), (A_1, A_2, ..., A_n)}{P}
\]

In order correctly to judge that \( P \), an individual need only have satisfied himself that \( (A_1, A_2, ..., A_n) \) are fulfilled and understand the principle expressed by \( (A_1, A_2, ..., A_n) / P \). On many accounts, this is justified by the claim that the principle involved expresses the meaning of \( P \) in terms of \( (A_1, A_2, ..., A_n) \).

For the Realist and the Anti-realist who relies on the idea of conclusive verification, the expression of the meaning of \( P \) is a truth-conditional one - the principle \( (A_1, A_2, ..., A_n) / P \) isolates the truth conditions of \( P \). It follows from this that the linkage between being satisfied that \( (A_1, A_2, ..., A_n) \) are fulfilled and judging that \( P \), expressed by means of the schema, is a truth conserving one for the Realist and the conclusive verification Anti-realist. An indication that \( (A_1, A_2, ..., A_n) \) are fulfilled merits the conclusion that \( P \) only when the truth of \( (A_1, A_2, ..., A_n) \) rules out the falsity of \( P \). Now on such accounts, principle \( (A_1, A_2, ..., A_n) / P \) - the principle which expresses the meaning of \( P \) and which thereby justifies the move in judgement from \( (A_1, A_2, ..., A_n) \) to \( P \) - is one which rules out defeasibility. But this need not always be the case.
Both the criterialist and the Anti-realist who relies on the theoretical notion of defeasible justification conditions emphasise the importance of defeasible grounds. Their version of the \(<A_1, A_2...A_n>/P\) principle cannot, therefore, be a truth-conserving one. On this version of the principle, a correct judgement can be one in which the assertion of \(P\) is defeasible. There is, however, a further consideration which separates out these two positions. Even for the Anti-realist who emphasises defeasibility, the relationship between the speaker and the \(<A_1, A_2...A_n>/P\) principle is one of the individual's 'knowing that' the principle holds; \(<A_1, A_2...A_n>/P\) is just one of the many theorems implicitly or explicitly known by the individual. The criterialist rejects this account because it involves a mistaken theory of meaning approach and because it thereby opens the door to the sceptic.

The schema on offer is supposed to explain what counts as warranting the move from \(<A_1, A_2...A_n>\) to \(P\), and so it must show how the principle involved contributes to a specification of the correctness of judging that \(P\). But just because this principle is one which makes use of defeasibility, the schema must take account of what it is that makes judging that \(P\) open to mistake. If the schema is to elucidate defeasible judgment, it must offer some way of distinguishing between those individuals whose judgements are sustained and those others whose judgements are defeated. And it must do so by demonstrating what it is about the meaning of the expressions involved which accounts for such success or failure. Now the Anti-realist specifies grasp of meaning in terms of

127
'knowledge that'. It follows that his explanation of such success or failure in terms of the meaning of the expressions involved must proceed by stipulating that which is implicitly known by the speaker. But such a stipulation involves not only a description of \( (A_1, A_2 \ldots A_n) \) but a description of those further grounds, \( (B_1, B_2 \ldots B_n) \), in terms of which the move from \( (A_1, A_2 \ldots A_n) \) to \( P \) can be said to be defeasible. Moreover, since, for the Anti-realist who relies on defeasible assertibility conditions, defeating conditions are themselves defeasible, the description provided must extend to some further set of conditions, \( (C_1, C_2 \ldots C_n) \) which in turn defeat \( (B_1, B_2 \ldots B_n) \). And so the defeasible assertibility conditions Anti-realist faces a dilemma. His claim is that the meaning of an expression can be substantively stated by giving conditions for its correct, albeit defeasible, use. But on the one hand, this seems to involve a regress of conditions which could never be so stated. On the other hand, attempts to halt this slide into regress at any given stage would allow room for sceptical manoeuvre. Just because the conditions in use, say \( (B_1, B_2 \ldots B_n) \), are defeasible, any refusal to countenance a need for specification of further background conditions would be dubious. For the sceptic would want to know why, given that the individual involved is depicted as not knowing whether \( (B_1, B_2 \ldots B_n) \) are defeated by the obtaining of \( (C_1, C_2 \ldots C_n) \), there is supposed to be no room for doubt about that individual's judgement that \( P \).

The criterialist's alternative position is that the proper account of the principle involved - \( (A_1, A_2 \ldots A_n) / P \) - is one which makes
use of RFA. An individual comes to understand the principle when he understands the rule for the use of the expressions which make up P. This rule cannot be substantively stated - there is no authoritative interpretation of the rule which resists further re-interpretation. It follows that principle \( \langle A_1, A_2, \ldots, A_n \rangle / P \) is not a substantive statement of the meaning of P. For someone to understand the way fulfilment of \( \langle A_1, A_2, \ldots, A_n \rangle \) warrants assertion of P, he must already be able to respond differentially to \( \langle A_1, A_2, \ldots, A_n \rangle \) in the kinds of situations in which this warrant would be undermined. The presentation of an explanatory schema already presupposes, in other words, that the presenter and the recipient of the presentation agree in judgements - they are pre-disposed to react to circumstantial variation in respect of the background context of \( \langle A_1, A_2, \ldots, A_n \rangle \) in the same way.

On the criterialist account, this means that for there to be an Assertion standard of correctness, there must already be agreement in responses to variation in Background circumstances. And it is in this sense that Background matters may be said to be independent of, though relevant to, Assertion matters. The criterialist offers an account of judgement which is inextricably wedded to the associated account of meaning. This latter account expresses the meaning of an expression by showing how the rule for its use is to be followed. It is this feature of the meaning account which explains the peculiar qualities of criterial judgements. The idea of following such rules only makes sense against a background of agreement in judgements. This agreement rules out the possibility of our raising genuine doubts as to
whether some criterially supported case of applying a given expression really is such a case. This exclusion of sensible doubt represents the foundation of criterial judgements. Attempts to repudiate this exclusion by raising such doubts carry with them the destructive consequence that the notion of rule-following is disrupted. But the schematic formulation of judgement just outlined requires that there be some account of the meaning of the expressions involved in such judgement - an account expressed in rule-following terms. It follows that to raise the kind of doubts which disrupt the notion of rule-following is to disrupt the concept of judgement. It is not, as Baker and Hacker might suggest, any specific concept which is disrupted by such an attempt to raise doubt. It is the very practice of wielding such concepts in judgement which is disrupted. This is the sense in which RFA can be said to generate a form of criterialism.

83 CRITERIA AND THEORY

The Analysis of Meaning

On RFA, then, to understand an expression is to understand the rule for its use. It is not the job of the philosopher of language to give a substantive account of these rules, but he can, nevertheless, indicate the criteria with which such rules are typically associated. Specification of when it is criterially justified to use an expression is, in certain respects, an answer to some of the sceptic's challenges. This answer does not consist in providing some account of when a particular expression should be used which gives a complete, substantive specification of all
the contexts in which it would be correct to employ the expression (ruling out the possibility of error were this specification only known by the speaker under sceptical attack and thereby ruling out the possibility of such attack). Nor, on the other hand, does the answer consist in demonstrating that the sceptic's use of the concepts involved is somehow self-defeating - as though any attempt to raise a sceptical doubt involves use of expressions which are suddenly, to the hearer, un-intelligible. There is no suggestion that concepts become disrupted in this way. Instead, criterialism provides a way of interpreting 'understanding a linguistic expression' which sets it inside a wider context of agreement in judgements. From the perspective of such agreement, the sceptic's challenge fails to represent a successful act of judgement. In attempting to raise his challenge, the sceptic demonstrates that he misunderstands the limits of justification which the notion of agreement in judgement imposes upon our language use.

What is needed now is an example of the way this RFA-inspired criterialist perspective can be applied in specific cases. The rest of this chapter is given over to examining one such case: the way criterialism provides an explanatory framework for the social sciences. There are two reasons why the link between criterialism and social science is particularly important.

(i) RFA pictures grasp of meaning as a practical ability. This ought to imply that the framework which explains meaningful language use can be used in explaining other meaningful, non-linguistic activity. But the field of study of human meaningful
activity is the domain of the social sciences. This suggests that RFA should carry consequences for the nature of the social sciences. (This, of course, is not a claim unique to RFA - meaning theorists also argue that their explanatory framework extends beyond accounts of language use to accounts which explain the meaningfulness of human actions in general. [12])

(iii) The picture of linguistic abilities offered by RFA relies upon the idea of agreement in judgement. But it is accepted that whether communities of others will display the agreement in judgements which we display is a contingent matter. This raises the issue of how best to frame questions about individuals from other cultures whenever there is uncertainty about their concept use. If it ever makes sense to question the meaning that a particular activity has for members of some foreign community, how best is that question expressed through formulation of theory, and how is it to be resolved by observation? We need, in other words, an account of what, from the rule-following perspective, constitutes the relationship between theory and observation in the social scientific study of communities other than our own.

In order to assess the nature of a criterialist framework for the social sciences, we must begin by examining the relationship between criterialism and the idea of scientific theory in general.

Criteria and theory

The position of the criterialist may be usefully contrasted with that of the meaning theorist. The meaning theorist who emphasises
substantive accounts of understanding has a number of options. He can claim that theories are to be interpreted in terms of semantic realism or anti-realism, in the sense outlined in Chapter One. Here the debate is about the extent to which grasping the truth of a theory is independent of whatever could be discovered by observation. For example, an extension of the 'underdetermination' conception in classical mechanics to the corpus of scientific data might, if such a thing could be properly formulated, suggest to the Realist of Chapter One that the truth of a theory is never conclusively established by appeal to observation. To the Chapter One Anti-realist, (one man's detachment being another's contraposition) it might suggest that the notion of truth, separated out from observation, is inapplicable to scientific theories. Now it has already been suggested that the general theory of meaning programme faces difficulties. But this does not necessarily dissolve the scientific realism/scientific anti-realism debate. For even within an account of the sort offered by RFA, there is room for discussion over the issues of (non-semantic) realism and (non-semantic) anti-realism.

The point is that we might refrain from semantic debate of the sort which distinguishes Realist meaning-theorists from Anti-realist meaning theorists and still raise the question (along the lines suggested by van Frassen [13]) whether scientific theories are intended by the scientist to be literally true or merely empirically adequate. This distinction can be expressed by separating out two views on scientific theories: (a) good theories state what is true; (b) good theories state what is true about
what we observe. The (non-semantic) Realist is committed to (a), his Anti-realist opponent is committed only to (b). Now Crispin Wright points out [14] that this raises a difficulty for the Realist. Traditionally, the Realist denies that there is a distinction between the theoretical and the observable, and so denies the Anti-realist's purported reduction of theoretical terms to observation statements - the Realist is being realist precisely about theoretical terms. But to be a Realist, he must make out the (a)/(b) distinction. So he must establish what it is for a theory to state the truth, and this must be somehow related to the claim that what is observable is true.

This requires that what is observed be taken to be true. But now it is an embarrassment to the Realist who denies the observation/theory distinction that observation statements themselves are required to be realistically true. For if so, they are defeasible. And this means that acceptance of them is a matter of evidence. But evidence must be represented by further theory - and so a damaging regress or vicious circle threatens. Wright's conclusion is that there are only two alternatives to the route which ends in circle or regress, neither congenial to the Realist. The first is to abandon (a) as a viable statement of realism and accept that observation statements might be true undetectably. The other is to deny what is common to both (a) and (b) - to deny the claim that good theories are involved with truth at all, whether truth simpliciter or truth of what is observed. The observation reports with which theory must accord are not, on this latter view, describable as true at all, in any sense which the realist would
recognize. Instead, a good theory is one whose features include that it requires least adjustment to our present world view, as conditioned by the rest of our theories.

Of course the Realist in question might abandon his hostility to the observation/theory distinction. But now, Wright points out, he will be hard pressed to defend his realism against Van Frassen's anti-realism. The Realist must argue that for a theory to be realistically true, it cannot be reduced to observational terms. As stated earlier, it is precisely the theoretical terms that the Realist is being realistic about. What it is for a theory to be true cannot just amount to what it is for it to fit in with true observations. But, if he is to avoid semantic realism, the Realist must accept that the truth of the theory is somehow related to such observation. The Realist is, in this respect, 'optimistic' in Wright's sense. But now the Realist seems unable to justify the supposed relationship between the truth of a theory and its fitting in with observation. The difficulty is, then, that the Realist has to view the truth of a theory as no more than, but no less than, the truth of the relevant observation statements.

[15] It seems, then, that the meaning theorist who espouses non-semantic realism is in a tough spot. The conclusion of both Wright and van Frassen is that the appropriate response is to abandon realism in the relevant sense. But, as the preceding chapter made clear, there is an alternative: to abandon the Realist's commitment to the meaning theoretic approach.
According to Wright, the Realist ought to accept that observation statements themselves are in some sense required to be realistically true, if they are to count as evidence for the realistic truth of a specific theory. But, Wright continued, this requires that belief in their truth be somehow evidentially motivated. And such motivation, given the Realist's rejection of the observation/theory distinction, must involve appeal to the truth of background theory and hence must involve an implicit appeal to a regress of justification (or appeal to a circular principle). For the criterialist, however, there is nothing to the quite general claim that for every statement there can be a substantive account of its justification. To judge that a particular statement has one truth value or another always takes place against a background of agreement in judgement. It follows that it is false to claim, as Wright does, (16) that if a statement is to be held or accepted as true, it must be warranted by (theory-laden) evidence. There will be always be cases in which observation statements are held to be criterially justified - justified by evidence in such a way that the possible defeat of such justification, though a live possibility, leaves no room for us to doubt that observation. It is equally false, therefore, to claim that non-semantic Realism which espouses an intimate connection between theory and observation must lead to a damaging regress or a vicious circle.

Two consequences follow from this. In the first place, the criterialist's appeal to agreement in judgement, which establishes that judgements always take place against a conceptual background
which is independent of such judgement, does not involve a
damaging regress or a vicious circle. In the second place, the
appeal demonstrates that there is no easy way to establish that
the criterialist is committed to an anti-realistic stance in
science. Criterialism is often depicted as a form of anti-realism
whose assertibility conditions have the peculiar feature of being
defeasible. The criterialism outlined here is not of this kind
because it explicitly rejects the meaning-theoretic framework
required for such a stance. Criterialism, in this sense, is not a
theory about semantic realism or semantic anti-realism. This in
turn is independent of the question whether the criterialist is a
Realist in van Frassen’s sense: one who regards the statements of
theory as literally true.

Indeed, one way to defend this kind of realism against Wright’s
attack is to make appeal to the notion of agreement in judgement
which lies at the heart of the criterialist account. The position
of the Realist, as outlined by Wright, is that he must produce
theories which are intimately related to observations but are not
reducible to such observation. This seems to create a problem. If
theory is related to observation, then the truth of theory depends
on the truth of observation statements. But such statements are
defeasible, and so require further evidence of a theoretical
nature, leading to a regress. In effect, then, Wright’s challenge
to the Realist is the same as his challenge to the criterialist:
demonstrate how a statement can be protected from doubt when it is
held to be defeasible. It is, then, open to the (non-semantic)
Realist to answer this challenge by making the same appeal to
agreement in judgements which the criterialist makes in response to Wright's anti-criterialism.

Of course, there might be other reasons to suppose that the Realist and the criterialist would find one another uncomfortable bed-fellows. The distinction between the van Frassen Realist and Anti-realists lies partly in their different responses to the notion of what counts as *observable by us*. Both accept that even observation language is theory-laden, and they also agree that what counts as observable is a contingent matter and might, for us, have been different. But they disagree over whether this means that there can be no distinction between the observable and the unobservable. To the Realist, 'observable' means observable under some idealization. Any state of affairs might be described as observable, with appropriate idealizations of the relevant observer. So there is no scope for the Anti-realists to formulate his thesis that our notion of theory should be restricted to what it says about the observable - the content of a theory must be made out in terms of what it says to be true. To the Anti-realists, 'observable' must, for us, mean observable *by us* - actual observability. And this means that what counts as observably true does not collapse into Realist truth. Now it is this appeal to idealization, on the part of the Realist, which will concern the criterialist. For the Anti-realists, a specification of the limits of the observable will be an empirical affair (one which might even involve the very theory under review, by means of a 'hermeneutic circle'). But a specification of that which is 'observable' in the Realist's sense must, it seems, resort to the
same kind of substantive specification which the criterialist found suspicious in the case of the semantic theorists.

That the criterialist might find himself, eventually, in disagreement with the (non-semantic) realist is, then, a live possibility. But there is, at least, scope for debate - scope which is opened up by an affirmation of the centrality of the criterialist's appeal to agreement in judgement.

94 RULES AND SOCIAL SCIENCE

What the foregoing suggests, then, is that the criterialist perspective offers one way in which to conceive of the relationship between theory and observation in science. Theories aim at stating the truth and the extent to which they do is settled by observation. The statements in which the reports of such observation are couched are defeasible, but this does not imply that such statements always stand in need of further support. Moreover, on the account just offered, it is still an open question whether the appropriate view of science is a Realist or Anti-realist one (in terms of the distinction drawn up by van Frassen).

So the general picture we have of science is one in which recorded observations are explained and further observations predicted. But whether the link between explanans and explanandum (or between initial conditions plus covering laws and prediction) is taken to be deductive or probabilistic, our explanation will be expected to
provide some sort of account of \textit{why} the appeal to this explanans (or set of covering laws plus initial conditions) is supposed to satisfy us. Now (returning to what was said in §2) as far as the social sciences are concerned, this raises the question for the criterialist of what it is about appeal to social scientific theories which provides this \textit{explanatory} background. Since the social sciences usually deal with the meaningful activities of human beings, we can expect his answer to lie in what such theories say about those meaningful activities.

There are, of course, two general perspectives on what a social scientific explanation of meaningful activity should be like. On the first, social scientific theories are characterized as isolating \textit{causal mechanisms}. On the second, such theories are said to capture the \textit{reasons} behind social activities. On either view, however, it is supposed that viewing the explanans (or initial conditions taken with covering laws) as causes or reasons will count as showing \textit{why} appeal to that explanans (or covering law/initial conditions) makes the occurrence of event E (or expectation of the future occurrence of E) sensible. It follows that before the criterialist view can be taken as setting out a framework for the social sciences at all, we must be certain that its account of meaning at least allows for explanation in terms of either causes or reasons, even if it does not settle the question of which is to be preferred.

But this highlights a particular danger in adopting RFA which stems from its emphasis on human agreement in judgement. For there
may be a suspicion that, since it is grounded in our agreements in
disagreement in judgements between ourselves and those others
involved in the activity. It would be unconvincing, for example,
to take as a given that the whole realm of rational action is
somehow restricted solely to our own range of agreement in
judgements. What this suggests is that there ought, therefore, to
be scope for discussion of communities other than ourselves which
does not, from the outset, impute irrationality to all of the
activities of the members of that community. The question now,
though, is whether the criterialist can provide for such scope.

The difficulty is not one of outright relativism. The social
scientist's job is to formulate theories whose worth is measured
by observational fit. The relevant observations will be
observations of meaningful activities - activities which will be
grounded in the subjects' own agreement, with one another, in
judgements. This need not, however, open a relativistic gap. For
us, nothing is added to the expression 'they follow rule R' by the
qualification 'or so it seems to us, given our agreement in
judgement'. To claim otherwise is to site the social scientist in
the position of the sceptic - to view the social scientist as
supposing that an observed instance of apparently following R
might, given sufficient disagreement in judgement between the
social scientist and his subject, really be an instance of
following some other rule, R'. This was essentially the worry the
sceptic tried to raise about the language learner in § 1.
Accordingly, the same reply is in order. The attempt to generate this relativistic gap fails because it relies on a mistaken link between judgement and justification. It is only because we cannot see certain complex patterns of intentional activity as other than instances of following a given rule that we are able to take part in the practice of judgement. Thoroughgoing relativism, like thoroughgoing scepticism, undercuts the very act of judging with which the relativistic (or sceptical) challenge is made.

The difficulty is, rather, to do with a deeper problem peculiar to the specifically social sciences which the foregoing shelved - a problem directly associated with the fact that the realm of study of the social scientist is the domain of meaningful activity. It is implicit in the criterialist perspective that the notion of rule-guidedness is one in which there is no substantive specification of standards of correctness. It follows from this that there can be, for the criterialist, no general account of rule-following in which all of the rules involved are understood in terms of a simultaneous authoritative interpretation. In the case of the social scientist, this means that the rule-guidedness which underpins the meaningfulness of the activities he observes is not a matter which can be captured by authoritative interpretation. But if there is no authoritative interpretation to be had of the rules by which the subjects of a social scientific study live, then this seems to imply a certain scepticism about the results of such study. How can we ever be sure that the interpretations offered by the social scientist correctly capture the meaning of the subjects' activities, when it is already
admitted that, for the rules which guide those very activities, there can be no authoritative interpretation? The difficulty is that there seems to be no standard by means of which to gauge the correctness or otherwise of the social scientist's analysis.

What is needed, it seems, is some account of what, according to RFA, constitutes the correctness or incorrectness associated with a particular social scientific theory, whether it be regarded as cause or reason based.

**Correctness and Ethnomethodological Analyses**

The potential problem is that, because there can be no authoritative expression of the rules in use within a community, the social scientist's theoretical account of such rules, framed on the basis of observations of behaviour, seems to meet with no criterion of correctness. This has prompted some social scientists to accept that there can never be, in these terms, any conception of correctness associated with social science while denying that this state of affairs constitutes a problem. The fundamental concept used in making out this case is the idea of *reflexivity* - a concept often associated with the ethnomethodological outlook. [17] An ethnomethodologist will explain the concept in terms of the distinction between appearance and underlying pattern or structure. For the ethnomethodologist, reflexivity refers to the essential co-mingling of these two aspects of experience.

Suppose the ethnomethodological appearances involved are participants' accounts of their own actions and beliefs. The
ethnomethodologist's purpose is to show how such accounts can be related to an underlying structure, in the form of a series of rules, which might be seen as lending sense to the participants' activities. To the ethnomethodologist, the identification of such a structure is an interpretative activity which spills over into the understanding of the accounts—and so accounts and underlying structures are, in a sense, partially self-determining. But the ethnomethodological activity of 'showing how such accounts can be related to an underlying structure' is itself the giving of an account. And so, reflexively, one introduces a new analytic task if one poses the question of how the ethnomethodologist's account can be understood in terms of its underlying structure; one's own understanding of the ethnomethodologist's account will depend, in part, on the structure one takes to be underlying the ethnomethodologist's practices.

So the meaning of accounts and the nature of the underlying patterns for which they are 'surface' documents come, to the analyst, as a package. And this is as true of the study of ethnomethodology as it is of ethnomethodology's studies. As a result, one's assessment of the ethnomethodologist's account will vary with the results of one's own interpretative work. Consequently, a statement of what a particular account means is wrongly construed if it is taken to be a statement about a determinate issue. For the meaning of that statement is itself open to renegotiation, in that the ethnomethodologist cannot avoid doing interpretative work when dealing with accounts, and we cannot avoid interpreting when we deal with the
ethnomethodologists' accounts, and *others* are similarly committed to interpretation in dealing with our accounts, and so on. Accordingly, there is just no content to talk of what an account (or statement) really means.

This is the crux of the reflexive attitude, and it shows why the issue of reflexivity is so closely linked to the problem which we started with. The difficulty we noted at the outset was that social scientific theories about the meaning of subjects' activities seem to lack any criterion of correctness. Because meaning cannot, in general, be captured by substantive formulation, the meaning of subjects' activities cannot be so captured. But if they cannot be so captured, then the social scientist's theories about the meaning of such activities cannot be gauged against such a substantive formulation of that meaning, and so there is no criterion of correctness for such theories. The idea behind reflexivity, on the other hand, is that there can be no authoritative statement of the meaning of the activities of social science's subjects. There can, likewise, be no authoritative statement of what the social scientist's theoretical pronouncements mean either. Accordingly, the same conclusion is reached: there can be no criterion of correctness against which to measure the social scientist's theories. For the ethnomethodologist, however, this does not constitute a problem, but merely marks out a boundary for social scientific explanation.

Of course, so stated, the ethnomethodological version of reflexivity and its acceptance of the idea that social scientific
theories do not aim to report the 'real' meaning of subjects' activities or the 'real' meaning of their accounts of those activities is bound to seem unsatisfactory. In general, the difficulty here lies in the fact that we seem to be faced with the claim that all talk of meaning is empty of content. And such a claim seems to leave the ethnomethodologist faced by a paradox. For it now seems that the texts represented by the formulations of ethnomethodology itself purport to be discussions of meaning which come to the meaningful conclusion that talk of meaning is without real meaning. For example, one way to take the ethnomethodologist's reflexivity claim that accounts can always be re-interpreted is to see it in terms of the following statement: R "No statement has a settled meaning - any statement can be interpreted as meaning whatever you want."

In the first place, this seems an unsatisfactory edict because it relies on there being content to talk of meaning - it seems to presuppose the very thing denied. However suppose there was some way to re-cast the statement, R, so that this apparent paradox or circularity was avoided. It would still seem unsatisfactory. And the reason for this is that it conflicts with our everyday idea that statements have content - that they tell us something about the world. On any sensible view of meaning, making statements is taken somehow to involve us in some commitment, C, to behaving in the future, or in counterfactual circumstances, thus and so. Any sensible view ought also to include a requirement, T, that statements aim at saying something true (however 'true' is taken) about the world, around which our commitment is centred.
Now the problem with R, however it is formulated, is that it seems to rule out both C and T. If there is no determinate content at all to the idea that by my statement, I meant so-and-so, then there is, equivalently, no content to the idea that my statement commits me to behaving in any specified way under given circumstances. And similarly, if there is no content to meaning-talk, there is no content left over for talk of truth, for any picture of truth will involve some matching of what was said in the statement with the way the world is (even if the way the world is is, itself, not a matter in principle separable from our means of investigating the world). But the whole idea of matching, whether founded in correspondence or coherence, is lost if one element of the match is systematically indeterminable.

So the argument against R is not just that R seems to rely on there being stable content to talk of meaning - a stability which R itself denies. The argument is that this stability has to be in some way a generalised feature of all statements, with the consequence that even if a non-circular version of R can be produced, R will be false. [18] In effect, the ethnomethodologist's reflexivity account faces the same problem with which we started: the social scientist's theoretical statements ought to face some criterion of correctness and yet the account of meaning on offer seems to rule this out. Consequently, the idea that the meaning of a text is always an indeterminable matter of re-interpretation - that the string of accounts of accounts of accounts is an infinite 'ongoing' one - must be, in some way, modified.
The Strong Programme and Convention

The strong programme in the sociology of knowledge can be seen as an attempt to provide such modification. [19] For the proponent of the strong programme, reflexivity is a desideratum of sociological theorising. According to 'Strong Programmers', sociology must obey certain strictures in regard to the way it treats its subject matter - and these strictures are to continue to be applied even when the sociologist reflects on sociology itself. It is this self-application of the 'proper' rules for sociological theorising that the Strong Programmer means by 'reflexivity'. But this does not, contrary to the claim of the ethnomethodologist, require an acceptance of the idea that statements have no content. Unlike the ethnomethodologist, the Strong Programmer sees such a result as insupportable. However, according to the Strong Programmer, a conception of meaning grounded in convention is still available.

The Strong Programmer claims that past experience and usage cannot determine future usage and that, because of this, an individual's application of a term must be taken as 'formally...a contingent judgement'. From this it follows, in turn, that where there is communal acceptance of some use of a term, this can only be as the result of a communal judgement. Such judgements can be either routine, in which case they are registers of existing conventions, or they can be the result of active negotiation, in which case they represent the formation of a new convention. This leads, says Barnes, to a radical view of the conventional character of knowledge. On this view, the meaning of concepts does not pre-determine future applications, and so, consequently, truth and
falsity are not 'inherent properties' of the statements which involve those concepts. Instead, concept application is to be viewed as explicable in terms of conventions grounded in communal interests. A community applies a concept in a given way because they stick to a particular convention. And that convention is stuck to because it meets communal needs.

The question now is whether appeal to convention shows that we could retain claims C and T (or some acceptable modification of them). On Barnes's view, the meaning of a statement is not determinate enough to settle the issue of whether what the statement says matches up with how the world is. The question of whether a statement is true would only be resolved once a (communal) decision about the meaning of that statement had been made. And so the claim here is that the potential regress of accounts represented by R would be stemmed by conventional, communal fiat. But this picture is problematic.

The problem lies in the idea that our belief about the truth of some statement is settled by appeal to convention - that talk of conventions could supply us with an adequate version of C and T. Consider the question whether the statement 'the sun was shining in Glasgow last week' is true. The normal view, when the veracity of such a statement is put into doubt, is that a careful assessment of the evidence will allow us to decide on that statement's truth or falsity. On this normal view, we believe the statement to be true (or false) because the evidence points in one direction or another. And others make sense of what they take to
be our commitments by viewing our statements against the background of evidence supplied by context.

Our judgements are, in this way, sensitive to evidence as to how the world is. According to Barnes, however, even after all the evidence is in, we still have to make a further decision, based on convention. At any time, it is possible that communal usage of concepts such as 'shining', for example, might swing around in such a way that what had earlier seemed, given the available evidence, the strongest kind of example of an application of the concept might be judged to no longer count as such an application.

Now what is wrong with this position is that it seems to locate our trust in the veracity of a statement in our conventions rather than in the evidence. Similarly it seems to sever any relationship between the notion of having a commitment and what this entails in terms of how the world turns out. But on the first hand, it seems unlikely that we, as language-users, would be willing to accept into our midst an individual who demonstrated a lack of concern with the evidence when attempting to assess whether a particular statement was true. Suppose some individual happened to concur with us in judgements about whether the sun is shining. If it became clear to us that he or she formed those judgements without ever paying attention to the weather, we would be unwilling to accept that that individual genuinely understood the concepts
involved. On the other hand, it is not clear that on such a picture we even have a recognisable grasp of what is meant by having commitments. The crux here is Barnes's insistence that free decision is always involved in concept application. For it seems unlikely that we could retain any notion of being in some way committed by what we say if what we mean when we apply concepts (or use concept terms) is the kind of through and through conventional matter suggested by the strong programme.

Of course Barnes might claim that the important contingent decision is that made by the community, [20] and so one's commitment is responsible to the community's contingent judgements. But if each individual's application of a concept involves a contingent judgement, this leaves the analyst with the problem of explaining why it should be that the patterned agreement in use which is an obvious feature of language-communities exists. In everyday life, individuals have no difficulty in bringing it about that they understand the same thing by a given expression. The purported explanation of this agreement is that it is the result of communal convention arising out of common interests. But it is not clear how this explanation is supposed to work. Since any decision as to when a convention applies is, itself, a contingent matter, the Strong Programmer seems to fall inevitably into an irreversible regress. [21]

These difficulties stem from the fact that we expect our account of meaning to demonstrate the link in meaning between the evidence for a statement and the statement itself. It will be remembered
from §2 that, even on the criterialist account, the link between 
(A₁, A₂, ..., Aₙ) and P is supposed to be a meaning-reflecting link. 
On the Strong Programmer's account, however, the link between 
(A₁, A₂, ..., Aₙ) and P becomes a contingent affair, mediated only by 
convention.

The criterialist approach

We began by outlining a problem which the criterialist account of 
meaning seems to raise for the social scientist. If there can be 
no substantive account of the rules which ground the meaning of a 
community's activities, then the social scientist's theories, 
which aim at depicting the meaning of such activities, seem to 
have no criterion of correctness. The ethnomethodologist accepted 
this and claimed that it did not represent a problem. The 
difficulty with this response is that, if it can be non-circularly 
stated, it makes the existence of meaningful activity, including 
linguistic activity, mysterious. The point of making statements, 
for example, would be lost, since the making of a meaningless 
statement involves no sort of commitment as to how future evidence 
might turn out. The Strong Programmer, on the other hand, accepts 
that the potential lack of a criterion for social science's 
thories does represent a difficulty. His response is to argue 
that the meaning of a community's activities can always be 
expressed by appeal to convention. In consequence, there will 
always be some formulation of this convention which will represent 
the kind of criterion required. The social scientist's theory will 
be successful just to the extent it correctly captures the subject 
community's conventions. The difficulty with this response is that
Chapter 3

it relies on the idea of the community's judgements being warranted by appeal to its conventions, rather than by appeal to the evidence available to the members of that community.

Neither the ethnomethodologist nor the Strong Programmer offers an appropriate response to the difficulty we started with. The ethnomethodologist throws out the meaning baby with the lack of criterion bath-water. The Strong Programmer offers a criterion for the social scientist at the expense of wrongly viewing the relationship between statement and evidence as a contingent matter. What is still needed, then, is some picture of communal activity which avoids the attempt to present substantive formulations of the meaning of such activity while nevertheless allowing for some notion of correctness against which social scientific accounts of such activity can be gauged.

Such a picture is immediately available to anyone adopting the criterialist perspective afforded by RFA. For the criterialist, the distinction in meaning between the social scientist's statements and the (potentially meaningful) activities of his subjects is that the social scientist's statements are couched within our language. There is, then, no scope for raising the sceptical issues which underpin the original problem. It is true that there can be no substantive, authoritative interpretation of the rules which account for the meaningfulness of the subject community's activities. It follows that the social scientist's theories cannot be gauged against such an authoritative interpretation of the activities of his subjects. But this does
not mean that there can be no criterion of correctness for the social scientist's output. If he ascribes a certain rule to the participant community, he thereby commits himself to the claim that, should such and such a set of circumstances come up, they will do so and so. If the participants do otherwise, the social scientist's ascription will be taken to be in error.

**Summary**

What criterialism offers the social scientist, then, is a framework in which it makes sense to talk of social scientific theories explaining and predicting observations of the meaningful activity of its subjects. Moreover these theories, contrary to the worry raised earlier, do face standards of correctness. The social scientist's job is to make sense of a community's activities. He records observations of those activities and offers up theories which make sense of those observations - theories which ascribe certain rule-following practices. This sense-making activity eventually reaches a terminus of justification. It follows that in one sense this rule-ascription must seem correct to us: if the social scientists describes participant behaviour in a particular way this will, to us as much as to him, mean that they are following such and such a rule. But this still leaves open the question of whether the theory on offer is correct. The social scientist's theories utilize rule ascription in order to explain and predict observations of the meaningful activity of its subjects. It is in respect of such explanation and prediction that such theories can be mistaken. Ascription of a given rule means that the subjects will, in particular circumstance, do so and so.
If they do otherwise then, according to the viewpoint afforded by our own agreement in judgement, we must regard the ascription as mistaken. This does not, of course, mean that we question the link between the rule and the behaviour originally described (to do so would be to question a 'grammatical' truth). It means, rather, that we question the veracity of that original description in our rejection of the rule-ascription involved.

§5 SOCIAL SCIENCE AND THEORY OF MEANING

What the present chapter has suggested, then, is that RFA offers a framework for understanding the nature of the social sciences. The perspective on offer demonstrates how theory and observation come together on a basis of agreement in judgement while, at the same time, illustrating the limits to sceptical and relativistic criticisms which such a basis might seem to engender.

The preceding sections represent attempts to extend the account of meaning offered for language use to the wider realm of meaningful activity in general. It was pointed out that this is also a goal of the meaning-theoretic approach. It might be thought, therefore, that such an account, even if its unsuitability as an account of language understanding is accepted, might independently represent an attractive model for understanding how the social sciences represent meaning. It is this thought which is pursued in what follows.
One of the most influential exponents of this idea is Philip Petit. Petit's claim is that there are two structures involved in social explanation - broad explanation and narrow explanation:

"We can... (define) the broad-narrow distinction as follows. A token or particular state is narrow if the presence of that type of state is guaranteed by the context-independent character of the subject; otherwise the state is broad."

[22]

Petit also distinguishes between two types of accounts of action explanation: accounts which rely on regularising and accounts which rely on normalizing. His overall claim is that broad minded explanation is dispensable to regularising accounts of action explanation, but is indispensable to normalising accounts of action explanation. Roughly speaking, normalizing accounts go beyond mere regularization by offering an explanation of events in terms of some normative principle.

What this approach requires as a basic pre-requisite, then, is a clear distinction between what counts as a normalizing explanation and what counts as mere regularization. The initial difficulty, however, is that this distinction proves more elusive, on Petit's model, than might at first be supposed. Suppose some event, E is individuated, as is some antecedent situation, C, together with some principle, C→E. The regularising thought is:

"if the principle generally obtains and the antecedent is satisfied in this case, then the consequent is fulfilled too."

The normalizing thought is:

"if the principle is generally to obtain, and the antecedent is satisfied in this case, then the consequent has to be fulfilled too."

In regularizing, E is seen to exemplify a pattern (enunciated or
evidenced by C→E) which is how the world regularly works. In normalizing, by contrast, E is shown to be something that had to happen if the world is to continue to satisfy C→E.

The difference here is supposed to be captured by the idea that in the latter case, E is something that has to happen if the world is to continue to satisfy C→E. That is, it is because E is necessary for the satisfaction of C→E that E occurs. Now Petit chooses, as an expression of 'E is necessary for C→E', the formula: (C→E)→E.

It is noteworthy that on this formulation, the conditions which fulfil the regularising thought are the same as those which fulfil the normalizing thought. If we take the assumption which forms the background to a normalising explanation to be the assumption that if it is the case that if the principle C→E holds the E, then E occurs, then on Petit's formulation this can be written as: ((C→E)→E)→E. This will be true except when C is true and E false. Equally, the assumption against the background of which the regularising explanation is offered is: C→E. And this, likewise, is true except when C is true and E false.

Now the difference between a normalizing explanation and a regularizing explanation, according to Petit, comes in the assumptions against the background of which the explanation is offered. But so far, the preferred versions of the two background assumptions have the same truth conditions: the normalizing assumption is true iff the regularizing assumption is true (the antecedent of the claim 'E is necessary for the satisfaction of the law C→E', (C→E)→E, is true only when the antecedent of C→E,
Chapter 3

C, is true, since C→E→E is true both when E is true and when E is false, provided that C is true – the content of E is unimportant). The two statements of the differing assumptions have the same consequent, and their antecedents have the same truth conditions, and so it follows that they must have the same truth conditions.

This means that there has to be more to the characterization of the necessity involved in the normalizing account. Now what Petit seems to rely on, in providing such further characterization, is the idea that there will, on normalizing accounts, be a 'deep-seated reason' [23] for thinking that, given C, E will follow. The idea here would be that where E is some behaviour elicited in response to a situation C, we could not conceive of the actor involved as, in the appropriate sense, similar to ourselves if he refrained from E. The link between C, E and our conception of actor as like ourselves in relevant respects, would be an a priori one.

Petit himself casts this a priori aspect in terms of the account of action set out by Paul Churchland [24]:

(1) X wants O
(2) X believes that A-ing is a way for him to bring about O under the circumstances
(3) there is no action believed by X to be a way for him to bring about O, under the circumstances, which X judges as preferable to him, or as more preferable to him, than A-ing
(4) all other things are equal
(5) X knows how to A
(6) X is able to A
(7) then X A's.

where X ranges over agents, O over states of affairs and A over actions. Here, X's A-ing is supposed to follow necessarily from
(1) to (6) in virtue of the fact that were (7) not to occur, we
would not question the principle that if (1) to (6) then (7).
Instead, we would supposedly assume that, contrary to appearances,
one of (1) to (6) was not in fact instantiated. Our deep-seated
reason for thinking that if C then E will occur is presumably to
be explicable, then, in terms of the claim that we have no
conception of what it would be to conclusively verify (1) to (6)
even where (7) is false.

This interpretation would tie in with Petit's claim that there can
never be normalizing principles which enunciate, rather than
provide evidence for, patterns. [25] But it does raise the
question of what such an account aims to achieve. The overall goal
for the meaning theorist is to provide an account, in terms of
'knowledge that' of what an understanding of (in this case) the
meaning of some activity consists in. But either such an account
will be, in the manner discussed in chapter two, substantive or it
will not. If it is not, then it is hard to see how, in the present
instance, the account can be genuinely meaning theoretic. There
ought to seem no reason, on such an account, for setting out a
framework which could allow for an infinite number of instances,
of the form 'if he means by activity A what we mean by activity A,
then on occasion C he will do what we would do in C, E'. The
fundamental requirement of finite axiomatization seems to become
redundant. If, on the other hand, the account is intended to be
substantive, it becomes important to assess just what kind of
substantive explanation is on offer for that in which a grasp of
the supposedly a priori relationship between (1)-(6) and (7) is
purported to consist. The criterialist, who eschews a completely substantive formulation, is able to appeal to agreement in judgements. The meaning theorist seems to have no such recourse.

96 CONCLUSION

The Story So Far

The preceding chapters have argued for an RFA account of meaning which rejects the meaning-theorist's 'knowledge that' framework. The perspective offered in its place has characterized grasp of the meaning of an expression in terms of the ability to follow a rule for the use of that expression. The notion of rule-following which lies at the centre of this perspective has two elements: a regulation element and a constitution element. The former requires that language users, individual and community alike, be responsible to some standard of correctness. The latter requires that this standard of correctness be nothing over and above the community's usage of the expressions governed by that self-same rule. This apparent paradox is resolved by appeal to the idea of communal usage as a general standard of correctness. In order to make out this appeal, the rule-following account, RFA, must reject all attempts to provide a substantive, meaning-theoretical account of meaning and rely instead upon the idea of agreement in judgements underpinning all rule-formulations.

A consequence of this is that the RFA account is able to make use of the notion of criteria. This allows the proponent of RFA to extend the counter he raises against the meaning-sceptic to more
general sceptical arguments. This in turn provides a framework for
scientific theorizing. By adopting the criterialist perspective,
the scientist is able to portray scientific theories as attempts
to convey truth about the world. The conception of truth involved
depends on the idea of such theories being responsible to
observational outcomes. But although observational statements are,
themselves, defeasible, the adoption of the criterialist
perspective means that such defeasibility does not result in a
chronic lack of justification for these observation statements.

The criterialist perspective is especially useful to the social
scientist whose observations and theories centre specifically on
subjects' social activities. The abandonment of the substantive,
'knowledge that' framework of explanation might have suggested
that such theories are problematic. Since the goal of such
theories is to capture the meaning such activities have for their
participants, the discovery that the rules which underpin such
meaning cannot be fully formulated might have seemed to leave the
social scientist's theoretical efforts with nothing to aim at. But
this is a mistaken picture. The goal of the social scientist is
not to provide substantive, meaning-theoretical accounts of the
rules which underpin the meaning found in the social activities of
communal members. That there can be no such account is, therefore,
irrelevant. The social scientist's task is to formulate theories
which make best sense out of the observational data. This sense-
making process occurs within the social scientist's own language
community and so relies upon the kind of agreement in judgement
which is the foundation of the criterialist account. This offers a
sense in which such theorizing can be open to criticism. In making sense of the activities of subject communities, the social scientist ascribes rules to them, and the correctness, or otherwise, of such ascription will show up in the observational data recorded. (26)

The Next Episode

The essence of RFA is that from within the perspective provided by agreement in judgement, human communities can be regarded as setting out their own standards of correctness in the production of meaningful behaviour. This suggests that what it is for there to be a standard of correctness is, in certain respects, a dynamic affair. Communities are constantly re-creating their own standards in the ceaseless bustle of normative activities—explaining, disputing, correcting and so on. But at the same time they are always responsible, in a general fashion, to their own usages on other occasions. This bipartite conception of communal standards of correctness in concept application, with its emphasis on the regulatory as well as constitutive aspects of communal rule-following, represents a formal constraint on social-scientific theorizing. Wherever the social scientist attempts to explain subjects' activities by characterizing the meaning of those activities, he must demonstrate the creative part the community as a whole plays in establishing that meaning and also show the way in which such creative communal activity is itself subject to regulatory forces. It follows from this that any social scientific theory which rules out either CT or RT must be viewed as, at best, incomplete. To see the effect of such constraints, the next
Chapter 3

chapter examines one particular social scientific theory which aims at characterizing the meaningfulness of social life.
CHAPTER 3: APPENDIX

The ethnomethodologist/Strong Programmer debate continued

Because he thinks that it is false to assume that our use of concepts is determined by their meanings or by the rules for their use, Barnes draws the general conclusion that our usage of those concepts is open-ended. [1] It is we, he says, who determine their application. And this determination is to be understood as taking place against a background in which any application of the concept involved could be justified, given sufficient ingenuity. [2]

However according to Barnes, this raises an apparent problem. If a concept's application is wholly determined by us, then it might seem as though all talk takes on an air of self reference, in that in the giving of any account the subject matter involved is invariably a product of our earlier judgements on how to apply the concepts involved. And this seems to push us to the conclusion that: 'all that can be referred to is speech...one cannot refer to things...the sinews of all realities are cut asunder'. [3] Indeed, says Barnes, this is precisely the conclusion to which the ethnomethodologist is committed by his or her approach to reflexivity. [4]

But according to Barnes, this cannot be right. The difficulty involved in treating discourse as self-referentially 'internal' in this way is that it seems to lead to a conceptual void. If the meaning of accounts, and of accounts of accounts, is essentially indeterminate, in the sense that no justification can be given for attributing one interpretation to an account rather than another,
then this seems to rule out the idea that there can be any content to the question of what so-and-so meant by such-and-such an account. So ethnomethodologists, as a result of their determination to treat the meaning of all accounts as systematically variable, overemphasise the importance of the study of accounts, and end up by committing themselves to an empty practice.

But, Barnes says, such a self-referential abyss can be avoided. Although it is wrong to say that an individual's applications, within an account, of a given concept are determined by its meaning, this does not mean that any interpretation of what the participant is doing (when he or she applies a concept) will be as good as any other. A participant's concept applications are not totally random. Whether an individual applies a concept correctly is a question that can be answered by appeal to what the community as a whole would say. In this respect, the community acts as a standard of correctness for concept use against which individuals' applications can be gauged. Of course, Barnes argues, this communal standard is a conventional affair. [5] Because the meaning of the concept does not determine its applications in advance, the community's choice of a given application as correct is an arbitrary matter. But this arbitrariness is not the destructive 'anything goes' arbitrariness of interpretation favoured by the ethnomethodologist. For the existence of a given convention in a community is itself an explicable affair. Each community has a set of goals and interests which influence its view of the world. These goals and interests exert a causal force
on community members in the applying of concepts, inducing them to reject some applications and accept others. And so the ethnometodologist's slide into talk of accounts of accounts of accounts in which 'speech can never do more than refer to itself' can be halted. It makes sense for an analyst to talk of the meaning of an account if he or she can give an explanation of such talk - an explanation which will refer to the way participants' interests can be said to cause the account.

The ethnometodologist's reply to the Strong Programmer

Unsurprisingly, the ethnometodologist has been unwilling to let matters rest here. One counter-criticism [6] which the strong programme faces is the claim that it fails to remain true to its reflexive colours. Since it is a central tenet of the strong programme that the concept of explanation is grounded in the notion of the causal force of participants' interests, it is clear that a properly reflexive strong programme would accept that the pronouncements of that programme itself are likewise explicable. But then this seems to require of the strong programme that it treat its own talk of interests as, itself, the product of active interpretative work - work done in the maintenance and pursuit of the Strong Programmers' own interests and goals. But while they are willing to pay lip-service to this reflexive requirement, the criticism is, the Strong Programmers fail to follow this through into reflexive practice. The result is that they treat interests not as analysts' (or participants') constructions, but as analytically unproblematic givens. The complaint is, then, that the Strong Programmer's texts illegitimately evade the regress of
account of account of account by surreptitiously placing in the
background the interpretative work involved in the isolation of
participants' interests. [7]

The Strong Programmer's response is that this misses the point. It
is common ground that in principle any interpretation of an
account can be justified. And this holds for Strong Programmers'
accounts too. So the question of whether a particular appeal to
interests is acceptable is not an issue which has a final, all-or-
nothing answer. Thus Barnes, for example, admits [8] that the
pronouncements of Strong Programmers lay no claim to
epistemological priority; interests are not held to be an element
of participants' lives whose existence can be indefeasibly
established. This, the Strong Programmer will assert, is
explicitly stated in the strong programme's formulations. So the
ethnomethodologist's criticisms are misplaced.

But the difficulty with this strong programme response is that it
may appear unconvincing. The Strong Programmer may claim that his
appeal to interests is not intended to betoken an especially clear
form of epistemological vision, but this is how it seems bound to
be taken, for it is this appeal to 'real' causes which gives the
Strong Programmer's explanations their spurious charm. [9] So it
seems that the Strong Programmer is faced by a dilemma. If the
explanation of accounts in terms of interests is retained as is,
the Strong Programmer will seem committed to a form of analysis
which makes illicit appeal to analytically 'unproblematic'
entities. If, on the other hand, the Strong Programmer makes 'genuine' attempts to relativise his or her employment of causal, interests-talk, then strong programme accounts will come to seem as fragile, in terms of having a determinable sense, as all other accounts, leaving the Strong Programmer in more or less the same position as that attributed by the strong programme itself to ethnomethodology - a position which, the Strong Programmer claims, is untenable.

This reformulation of reflexivity as a problem for the Strong Programmer leaves some room for a strong programme rejoinder, through an emphasis on the contingent nature of participants' interests and goals. That participants understand an aspect of their world in some particular way and apply their concepts accordingly is, the Strong Programmer argues, a conventional affair. Nothing about the meaning of concepts can be seen as forcing participants' to apply their concepts one way rather than another. That they do so is merely a matter of communal convention. And that a community has a given convention is explicable in terms of those of its goals and interests which cause or give rise to their conventions. This much is now familiar. What this shows, the Strong Programmer admits, is that interests are the theoretical posits of the analyst's theory. Because their existence is inferred from participants' accounts, any existence claim regarding participants' interests will always be defeasible. Because, that is, participants' accounts, conventions and interests are, for the analyst, inter-linked, the question of whether the analyst has correctly identified the
meaning of an account (or has correctly identified the interests from which it derives) will in principle always be an open one. As such, it may be that the existence of a particular interest, and the supposition of its causal influence, will come to seem more, or less, likely as an explanation of participants' lives as further analytic work is done. Such further work may include analysis of the analyst's own interests, and later assessments of the credibility of these posited participants' interests may rest upon this analysis of the Strong Programmer's interests. But this shows only that, in the light of later evidence, the analyst's earlier interpretations were wrong. It does not show that, in the light of earlier evidence, those interpretations were unfounded.

However this conclusion fares as a riposte to the ethnomethodologist's criticism, it is clear that, as a version of how Strong Programmers employ interests as analytical tools, the argument for the conclusion relies heavily on the claim that, by appeal to communal convention, sense can be made out of participants' judgements as to concept applications, even if the knowledge claims generated are defeasible. The Strong Programmer's claim [10] is that, although a participant's employment of concepts is not determined by the meaning of those concepts, it makes sense to consider that individual's contingent judgements about the application of a concept as answerable to a criterion of correctness laid down by communal convention. The analyst can make out a case for having divined the nature of such conventions if he draws a picture of the kinds of interest that might be said to motivate the maintaining of those conventions. This would justify
him or her in identifying a particular participant's concept application in a given situation, thereby providing grounds for a claim as to what relevant participants' accounts might mean. If this is the correct way to understand the strong programme, however, then it seems that, independently of whether the strong programme has a problematic epistemological (or even ontological) commitment to interests, there is scope for raising the question of whether that programme's appeal to communal convention can do the job required of it.

The task here is to consider whether the application of concepts (or the use of concept terms) could be the kind of through and through conventional matter suggested by the strong programme. The claim is that applications are never determined by meaning - each individual's application of a concept involves a contingent judgement. Now obviously this leaves the analyst with the problem of explaining why it should be that the patterned agreement in use which is an obvious feature of language-communities exists. The Strong Programmer's explanation of this agreement is that it is the result of communal convention arising out of common interests. Whether this is a coherent reply remains, however, to be seen.

The strong programme's appeal to convention came as a consequence of its insistence on the idea that any and every act of concept application involved a contingent decision. [11] The failure of the strong programme's appeal to convention implies that, if the intuitively appealing notion of a generalised agreement in language use is to be retained, the Strong Programmer's commitment
to the prior idea of contingent judgement must be rejected. The
Strong Programmer's belief that concept application is always a
contingent matter stems, it will be remembered, from the claim
that neither the meaning of a concept nor the rules for the use of
a term can be said to determine specific applications. [12] The
contrary of this argument, it is asserted, is 'everywhere denied'.
[13] Consequently, one way of arguing for the rejection of the
Strong Programmer's claims about the contingent nature of concept
application is to demonstrate that there is something wrong with
the claim that the rule for the use of an expression cannot
determine what will in the future count as the correct use of the
term - that what is to count as correct application of a concept
is not always a matter for arbitrary decision.

Barnes states that 'it is false to assume that usage is determined
in advance, by meanings, rules, norms, logic or similar entities',
and claims that this conclusion can be drawn from Wittgenstein's
later philosophical work on rule-following. [14] But one outcome
of the discussion in Chapter Two was the conclusion that, from the
perspective of our own agreement in judgement, it does make sense
to view meaning rules as somehow settling what correct usage
amounts to. It follows that, contrary to Barnes's claim, one could
make appeal to a rule-following account without relying upon the
idea of contingent decision (whether conscious or habitual) at
every step in the use of an expression.
CHAPTER 4

§1 INTRODUCTION [1]

Social scientific explanation proceeds by describing a variety of observations and positing relationships between them explicable in terms of an overall theory. Whatever the theoretical terms, however - e.g. motives, unconscious desires, ideological fixations - the explanation can only proceed against the background claim that the activity captured by means of such terms is, for both actors and social scientist, a meaningful one. The social scientist must therefore provide his explanations by reference to the constitution and regulation aspects of the communal rule-following practices which underpin the meaningfulness of social life. His theories are not designed to be substantive, meaning-theoretic formulations of such communal rules but they are intended to provide an explanation (to someone who already shares with him an agreement in judgements) of the nature of the rules which the relevant community can be said to be following, together with an account of the point which such rule-following practices have for the communal members involved. The suggestion of the previous chapter was that the criterialist perspective offers one way in which this can be achieved.

One difficulty with a rule-following account of the meaning which events and activities have for their participants arises from the fact that, in many cases, two separate notions are conflated in such explanations: (a) an outline of the social rules or social
conventions a particular society follows (e.g. a register of laws, hierarchies, taboos and so on) and (b) an outline of the rules through which we can express the meaning of the concepts employed by communal members (including the concepts through which the social rules and conventions may be expressed). The difficulty with such a conflation is that the two types of rule are fundamentally different. It is quite acceptable to conceive of laws, taboos and the like as receiving an explicit formulation which stands in no need of further interpretation. [2] To anyone who understands the language in which the formulation is couched, the requirement laid down by the rule may well be completely fixed. According to the arguments of the previous three chapters, however, the same cannot be said of the rules through which the meaning of concepts may be expressed. If the social scientist aims to provide an account of the meaning which a range of activities has for communal participants, a quite different explanatory framework must be used.

Whenever these different types of rule are conflated, problems inevitably ensue because the social scientist attempts to account for the meaningfulness of communal activities by adopting the substantive formulation framework which is only suitable for the expression of rules such as communal taboos and the like. As a result, the rules which guide the application of concepts within the society may be given a spurious 'non-re-interpretable' substantive interpretation in order to comply with the regulation requirement (expressible, perhaps, by means of some supposed communal convention, in the fashion of the Strong Programmer).
They may, instead, be given a substantive interpretation which can be re-interpreted in any way seen fit, at the communal level, in order to comply with the constitution element (expressible, perhaps, by adoption of the reflexivity perspective in the fashion of the ethnomethodologist).

This is particularly important since many current forms of practical social scientific theorizing attempt to explain social experience by demonstrating that individuals' understanding is determined by communal forces. In so doing they succeed in emphasising the constitutive role played by the community as a standard of correctness against which individual application of concepts can be gauged. But this emphasis often overshadows the criterialist's other requirement: that the theory illustrate the regulatory standard of correctness to which the community's own concept usage is responsible. The social scientist commonly ignores the idea of standards of correctness at the communal level and concentrates exclusively on the regulatory role of the community in respect of the individual. In what follows, one typical example of such theorizing - the social representations perspective - is examined. The aim is to demonstrate how, at a practical level, the criterialist's notion of communal standards of correctness imposes restraints on the kind of explanation social science can offer.

**Social Representations**

Moscovici's theory of social representations [3] has attracted considerable interest among social psychologists, as well as
stimulating a growing body of research. [4] Moscovici's goal is to resist the individualistic reductionism common elsewhere in social psychology and improve on previous theoretical constructs such as attitudes and opinions. There are, however, a number of flaws in his theory which stem from the way he depicts the role of the community as arbiter of meaning.

Moscovici reports that the term 'social representation' can be traced back to Durkheim's concept of collective representations [5], and he offers a variety of examples to illustrate its use. These include popular conceptions of scientific theories, such as the layman's notion of neurosis [C, p.110], traditional beliefs such as religious or superstitious beliefs [A, p.135], commonly employed explanations of the world such as talk of the 'class struggle' [B, p.952] and popular ideas on societal features such as racial differences. [6] The most famous example is, of course, his study of the pared-down version(s) of psychoanalysis widely used in France during the 1950's. [7] However he is less willing to provide a precise definition of 'social representation' [B, p.957] and on occasions even treats its vagueness as a virtue. [D, p.91]

Nevertheless it is possible to glean from Moscovici's writings a partial list of definitional criteria, although this is a difficult task in that Moscovici's comments on the matter are often fragmentary and are sometimes contradictory. The next section attempts to remedy this by collating these diverse
comments, thereby providing a coherent picture of social representations theory. [8]

§2 SOCIAL REPRESENTATIONS - THE THEORY

Structure, Content and Function

At the most general level, Moscovici describes social representations as mental phenomena (or processes) having both structure, or form, and content [9] and occupying a curious position between understanding-type concepts and perceptions. [C, p.17] Their structure is complex in that representations typically comprise both abstract elements and pictorial elements. The abstract property of a social representation is in most cases outweighed by its pictorial side, as most social representations centre on a figurative nucleus. [C, p.38] A social representation, then, takes the form of a set of concepts and linguistic elements which are either abstract or pictorial - a combination in which the pictorial dominates in most cases.

As is obvious from the list of examples given by Moscovici, the content of social representations varies extensively. What can be said, though, is that this content typically includes both cognitive and affective aspects. Furthermore all representations have a common mode of derivation, stemming either from the dissemination of scientific theory or from other social representations or from inter-individual communication. The intrinsic linkage of social representations to communication is important for two reasons.
First, it underpins the social nature of social representations and their contents:

'The word 'social' was meant to indicate that representations are the outcome of an unceasing babble and a permanent dialogue between individuals...Representations adapt to the flow of interactions between social groups...'

[B, p.950]

Thus it is part of the essential nature of social representations that they are dependent on actual instances of inter-individual communication; in part, social representations grow out of the discussions people pursue in making sense of their social lives.

Secondly, the linkage between social representations and inter-individual communication provides a dynamic for the contents of social representations:

'...what is most striking to the contemporary observer is (social representations') mobile and circulating character; in short their plasticity. We see them, more, as dynamic structures, operating on an assembly of relations and of behaviours which appear, and disappear, together with those of the representations.'

[C, p.18]

As the nature of representations is dependent on their actual use in everyday talk, their content will be continuously changing as individuals employ them in new contexts.

Social representations are therefore complex mental entities which have abstract and pictorial elements. They derive from inter-individual communication and so are essentially social. [C, p.13]

This social aspect ensures that social representations, unlike, in Moscovici's view, other social psychological constructs such as schemata or tacit knowledge systems, can never be static entities, for they are continuously modified by the pressures of use.
The social nature of social representations has implications for their function. Social representations function by providing a meaning for the communicative matters common to a group of speakers. A representation imparts meaning by assimilating objects to general categories of thought. Because they have this codifying effect, social representations allow an individual to pick out some part of the environment in the knowledge that the object so identified will have the same type of meaning for those who share his or her social representations as it does for himself or herself. Experience is constrained by social representations in that the world of an individual is only meaningful to the extent that social representations give it meaning. [C, p.7] This constraining feature of social representations demonstrates that social representations are more than a mere mode of communication; they have the capacity to construct the meaning of what is communicated. For this reason Moscovici thus contrasts social representations theory with information processing perspectives by emphasising that for the social representations theorist, 'factuality is never at the core of the exchanges between members of society'. [B, p.963]

Social representations endow experience with meaning in a constructive manner, and thereby determine the way an individual thinks about and responds to his/her experience. This has two consequences. In the first place, it means that social representations theory provides a principled demarcation of the limits of social groups. If social representations determine meaning, then all of those who share a particular social
representation will agree in their understanding and evaluation of
the world in some respect. Social representations place
individuals in identifiable groups just as surely as do physical
characteristics or occupational status. [D, p.91] Put another way,
for a social representations theorist, what makes a group a group
is exactly the shared representations of the members. [10]

The second consequence of this constructive aspect of social
representations is that they allow people to feel familiar with a
world which might otherwise seem strange and threatening. This is
best understood by considering in detail the way in which social
representations are said to construct meaning. The two key
concepts in such a consideration are the twin concepts of
anchoring and objectification.

**Anchoring and Objectification**

To see how social representations bestow meaning on objects, says
Moscovici, one should examine the way they make an unfamiliar
world familiar by considering the case of a group of individuals
(or an individual group member) confronted by a novel, unfamiliar
experience. Because human beings have an innate desire to make
sense of their world, these individuals can be expected to
assimilate this new experience to their existing knowledge. That
is, the unfamiliar experience produces what Moscovici calls a
'fracture or fissure' in the group's perceptions which incites
their curiosity and induces them to 'take hold' of the unfamiliar
concept. [C, p.29-30] Once the group has taken hold of the
unfamiliar object, an attempt is undertaken to make the
experienced object more familiar - a process which operates through the social representations.

The unfamiliar object is first assigned to one of the categories of thought which makes up the social representation - a form of classification which Moscovici calls 'anchoring'. This anchoring procedure depends upon the fact that social representations are built around a nucleus which is a prototype or paradigm for the category class. In this respect, social representations resemble the more familiar notion of prototypes. [11] When a group anchors some unfamiliar object, it does so by relating it to the paradigmatic or prototypical case which lies at the heart of the relevant social representation. [8, p.953] Once an object has been anchored to a social representation it acquires, in the eyes of the anchoring group, the characteristics and tendencies of the group prototype. This assimilation to the prototype bestows a meaning on the object which it lacked in its unfamiliar state. [C, p.35]

The anchoring process is accompanied by another process which Moscovici calls 'objectification'. Once an unfamiliar experience has been drawn into one or other of the group's social representation categories, its essential elements are reproduced, within the minds of group members, in pictorial form (except in the case of taboo subjects etc.). This image is then seen in relation to those other images which form the nucleus of the prototype and, subsequently, the image of the erstwhile unfamiliar object is amalgamated with those other, nuclear, images. Such
images then become reality for the group, in the sense that when
group members talk about the world they are, in fact, talking
about these images. [C, p.40] In this way, the capacity of social
representations to make the unfamiliar world familiar allows them
to operate as entities which construct or bestow meanings.

The objectification process has both historical [C, p.29] and
prescriptive elements. [C, p.9] It is historical in that the
social representations which are, for us, reality are partly the
result of earlier objectification processes. It is prescriptive in
that the constraining force of such representations is a power
which cannot be resisted:

'Nobody's mind is free from the effects of the prior
conditioning which is imposed by his representations,
language and culture...We see only that which
underlying conventions allow us to
see...representations are prescriptive, that is they
impose themselves upon us with an irresistible force.
This force is a combination of a structure which is
present before we have begun to think, and of a
tradition which decrees what we should think.'
[C, p.8-9]

So the prescriptive power of social representations, stemming from
the objectification process, is a consequence of the influence of
the past over us. The reality of yesterday controls the reality of
today, says Moscovici, such that intellectual activity constitutes
a mere rehearsal or re-presentation of what has already gone
before, in that our minds are conditioned by representations which
are forced upon us.

Moscovici claims that the theory of social representations
explains not only veridical judgement but also illusion. [C, p.38]
As social representations are partly made up of iconic elements,
they are not required to obey the dictates of logic such as the law of non-contradiction. Thus social representations exist which cannot accord with reality — and this means that groups of individuals can impose illusory interpretations on their experiences. Illusion and reality are, says Moscovici, achievements which have the same explanation: both are related to the world of representations.

The Consensual Universe

For Moscovici, then, social representations are mental entities made up of abstract and pictorial elements which function via the categories through which individuals construe their world. And in this way they act as group parameters which separate out individuals in terms of how they see the world. All of this shows, says Moscovici, that 'knowledge of the world' is not the same concept for social psychologists as it is for others, such as those involved in the physical sciences.

Moscovici proposes a fundamental distinction between a 'reified universe' and a 'consensual universe'. [C, p.19-23] The former is constituted by objects whose qualities and characteristics exist independently of individual interpretation and which combine to form brute facts. In the latter universe — the domain of the social psychologist — objective facts as such do not exist. The world of the consensual universe is one in which facts are imbued with meaning for individuals and things are as they are conceived to be: a conception of an object becomes a perception of that object. [C, p.40] What the social scientist must understand, says
Moscovici, is that the scientific model of knowing and learning about the world is inappropriate for study of the consensual universe, for such a model cannot encompass the essential creativity of social representations. [C, p.13]

Examples of such reality construction are, says Moscovici, as numerous as social representations themselves. However the most obvious examples of such constructiveness are probably those in which a scientific theory comes to be disseminated throughout common sense. Moscovici reports, for example [12], that the scientific hypothesis of functional specialization between the cerebral hemispheres has become part of 'folk neuroscience'. Non-scientists have, purportedly, taken up this notion and transformed it into the claim that people really have two minds, a claim which is then used to explain the existence of cultural dichotomies such as 'rationality v. intuitiveness' and 'masculinity v. femininity'. In such a process of transformation, then, a particular scientific theory is seized upon by laymen and is re-created anew. [B, p.964]

The ramifications of such a social representational dissemination and transformation for group membership can be seen by comparing those who see the world in terms of the split brain social representation with those who see it in terms of an alternative such as the common sense version of psychoanalysis.

According to Moscovici, then, social representations shape our collective consciousness, and so social psychological notions such as 'attitudes', 'opinions' and 'social groups' will be inherently faulty unless they come to terms with this shaping process -
unless, that is, they are given a location within the consensual universe. For a social psychology to be worthy of the name, it must deal at some point with the social representations.

**Recap**

The most important aspects of social representation theory are:

(A) social representations have a form which is partly abstract and partly pictorial;

(B) they function by allowing people to make joint sense of an unfamiliar world, and they thereby mark the psychological parameters of groups;

(C) they assimilate the unfamiliar through the processes of anchoring and objectification;

(D) objectification prescribes how an individual will see the world;

(E) the prescriptive force of social representations can be thought of in terms of the past influencing the present;

(F) this influence means that the world is as it is given to us by our social representations — in this sense, social representations constitute our reality;

(G) that our world is constrained by our social representations in this way shows that the social scientist, unlike the physical scientist, must confine his analyses to the consensual universe.

Taken together, these claims provide a powerful argument for the primacy, in social psychological study, of the social representation. However the next section reveals that this argument is, in several respects, importantly flawed.
Chapter 4

§3 DIFFICULTIES WITH THE THEORY

Creating the world

Moscovici's contention that social representations constitute reality [13], lies at the heart of social representations theory. The importance of this reality-constituting thesis stems from its attempt to tackle, head-on, the question of how social psychology can develop theories which take proper account of the active, meaning-constructing nature of human existence. The answer provided by Moscovici is that human beings encounter the world only in terms of meaning-imbued experience - experience constituted by social representations. Now this might, at first, seem radically idealist in tone. But Moscovici does not, in fact, deny that there is a world beyond social representations - he maintains only that the outside world does not directly enter into our social relations. Indeed, according to Moscovici, sometimes representations are mistakenly not recognised as such - they are 'not thought of as symbols but as reality'. [D, p.91] [14]

A problem exists, however, in reconciling this emphasis on the constitutive nature of social representations with Moscovici's claim that there is a world beyond social representations. If social representations constitute our reality, such that all thought is predetermined by social representations [C, p.36; Moscovici, 1983, p.129], then the notion that there is a world beyond the world of social representations would seem to have no content. If 'representations are all we have', then there is no coherent way to understand talk of a world which can be presented
to us independently of those representations. Furthermore, social representations are said to be responsible for both veridical and illusory experience. [C, p.38] But this too is problematic: for, on Moscovici's formulation, we are left with no criterion of correctness which could establish the difference between illusion and reality as separable elements of our experience. Moscovici is certain that social representations can be wrong [A, p.134] but what Moscovici fails to tell us is how to distinguish wrong representations from right ones. If social representations determine our reality, there seems to be no way left in which to make out the claim that nevertheless, for us, some of those representations are wrong.

The difficulty, in other words, is that the community, on Moscovici's account, faces no standard of correctness in its application of concepts. An individual can be mistaken in applying his concepts, relative to the community's usage, but for the community itself, however it thinks it best to apply its concepts to the world just is how those concepts apply. Now if there is any content at all to talk of a world beyond our representations, then it must make sense to think of it as representing some limit to correct concept application. Given the idea of a world beyond our representations, it must make sense to think of the community as a whole as, in some respects, subject to a standard of correctness. But Moscovici's theory leaves no room for this. It follows that Moscovici's talk of a world beyond representations is inconsistent with the account he offers of those representations.
There is a tempting, and superficially convincing, response to this point: the claim that science tells us what the real world is like. Science recovers the unvarnished facts against which the correctness of social representations can be checked. Now, on at least some occasions, Moscovici certainly seems to regard scientific knowledge as independent of social representations as, for example, when he talks of a 'language of observation ...expressing pure facts'. ([C, p.17) There is, however, a crucial difficulty with the suggestion that scientific knowledge could act as a standard of correctness.

If science is to distinguish between correct and incorrect representations, then this seems to invite a form of reductionism. For the question of whether a representation is correct would reduce to the question of whether that social representation adequately reflects what would be the content of supposedly objective scientific knowledge claims. The problem here is that this position treats social representations as a mere level of analysis, adopted for reasons of brevity, clarity and so on; a level of analysis which could, if required, be analysed out into scientific terms (perceptions, cognitions or whatever). However given that the essence of social representation theory is that it rejects the idea that human social life can be fully explained by reference to neutral, factual information, Moscovici would be unlikely to adopt this course. Indeed he seems explicitly to repudiate this view, ([C, p.4) claiming instead that all individuals make sense of their social interactions only through the medium of social representations. ([A, p.129) Since, in line
with recent social studies of science, he also accepts that the practice of science is, itself, a form of social interaction prey to the influences of history and tradition [C, p.28], it seems that scientists, like all other individuals, must rely on their social representations in making sense of their own activities. [15]

In one respect, then, it would be a profound mistake, from the perspective of the social representations theorist, to think of science as an unproblematically asocial realm of activities in which knowledge of 'pure fact' is generated; the scientist is as much trapped in his social world as is the layman. According to Moscovici's own suggestions, not only is it impossible to treat scientific knowledge as a standard of correctness for social representations, but the whole distinction between social representations and scientific knowledge becomes unclear. For the individual, layman or scientist, there are only social representations. As a consequence, the idea that scientific knowledge can be regarded as in any way immune from the effects of social representations, as is suggested by Moscovici's talk of language expressing 'pure facts', must be rejected.

There is a temptation to respond to the foregoing by objecting that, while laymen inhabit consensual universes where the mode of knowledge is the social representation, scientists work in a reified universe in which the mode of knowledge is science and that, consequently, scientists just do gain non-social, objective knowledge of the world. The problem with this position is that the
distinction between reified and consensual universes is, by and large, supposed to be understood in terms of the distinction between knowledge based on social representations and the objective form of knowledge typified by science. It follows that any attempt to explain the difference between scientific knowledge and social representations-based knowledge by way of the distinction between reified and consensual universes will be simply circular.

Even if such circularity could be avoided, the claim that scientists do gain objective knowledge by working in a reified universe would still be problematic. First, it would stand in contradiction to Moscovici's fundamental thesis that all social activities involve social representations and that reality is what is given by those representations. [16] Second, it would raise precisely the same kind of reductionist issue brought out earlier when the suggestion was mooted that science might serve as a standard of correctness for social representations. If there is any form of objective perception of the world available to scientists, then, independently of whether we take it to be a criterion of correctness for social representations, there seems to be no reason why social scientists should be restricted to basing their understanding of social phenomena on the study of essentially social, meaning-imbued social representations - which Moscovici himself describes as 'vague' [D, p.91] - when they could study the objective facts of the matter instead. So there seems to be nothing wrong with the claim that the social scientist need not stop his analysis at the level of the social representation; he or
she could go further and discover the objective facts beyond those representations. And if this is a possibility, then there ought to be no difficulty in reductively construing those representations in terms of such facts. And this would be clearly contrary to Moscovici's own view. [A, p.142]

Of course someone might say that the difference between the social scientist and the non-social scientist is that the former just does not have access to the kind of objective knowledge with which the latter deals. The social psychologist is a scientist [C, p.23] - but works in the realm of social science. However this objection will impress only those who have already forgotten that the distinction between the social and the non-social is precisely the distinction which talk of the scientist's access to objective knowledge, in contrast to social representations-based knowledge, is supposed to explain. [17] In general, then, it seems that Moscovici's discussion of scientific knowledge fits badly with his account of the essentially social nature of social representations, in that any non-circular account of the distinction between scientific knowledge and knowledge based on social representations seems to lead to a species of reductionism.

Recapitulation

So the social representations theorist faces a dilemma. He might argue that there is no scope in social representations theory for the idea of the community's application of concepts being responsible to correction from any other source. But just because the only standard of correctness for such concept application on
offer would then be the standard set for a given individual by his community, it seems as though there could be no content to the claim that the community itself is correct (or incorrect) in its applications. This would rule out Moscovici's claim that the world exists independently of social representations. If, on the other hand, the theorist argues that science offers a competing view of the world, independent of that given via social representations, such that we can view representations as fallible in respect of such knowledge, then the question becomes one of why we should not treat representations as reducible to such scientific terminology - a view expressly rejected by Moscovici.

The realm of the unfamiliar

The general difficulty for Moscovici here is that his conception of science seems to demand a standard of correctness for the community as a whole, represented by 'how the world is', which is ruled out by his account of the way communal usage imparts meaning to the social world. This difficulty re-emerges in Moscovici's account of the functioning of social representations. Individuals are said to make sense of the world by anchoring unfamiliar objects to social representations prototypes. But this implies that, in some fashion, individuals perceive these unfamiliar objects prior to anchoring them. Moscovici himself expresses this notion in a number of different ways, stating, for example, that individuals 'take hold of an unfamiliar idea' [18] and that 'we...draw something foreign...into our system of categories'. [C, p.29] Moscovici also claims that the unfamiliar object world is 'perceived at first in a purely intellectual, remote universe'.

191
Chapter 4

[C, p.38] The problem here is that Moscovici gives no adequate account of how this 'drawing-in' or 'taking hold' is to operate. All thought and perception is, according to Moscovici, based on social representations and so this seems to rule out 'taking hold' of unfamiliar objects as a cognitive or perceptual process. Nevertheless Moscovici insists not only that such a process does occur but that study of the realm of the unfamiliar is, in the last resort, the true goal of the social psychologist [C, p.28], in which case it seems legitimate to ask how the process of taking hold does operate.

This brings us back to the problem of how the world can be experienced independently of social representations. If there is content to the notion of 'taking hold' of the 'unfamiliar', then Moscovici ought to explain why it is wrong to construe human social cognition reductively in terms of that non-social representations notion, given that such 'taking hold' will, by hypothesis, itself be an essential element in social life. Moscovici offers an account of understanding the world in which meaning is settled by communal usage while, at the same time, clinging to a notion of 'taking hold of' the world beyond communal representation which, contrary to the account of understanding on offer, represents a locus of correctness conditions to which community-based explanations of meaning might readily be reduced.

Neither this point nor the earlier doubts raised in connection with Moscovici's account of scientific knowledge should be taken as a plea for the validity of some form of cognitive reductionism.
Both sets of issues are raised solely to illustrate the fact that social representations theory, as formulated by Moscovici, seems to involve a line of argument which can offer no reason why the social ought not to be regarded as in principle reducible to more basic cognitive notions. The worry is: if it is a necessary feature of Moscovici's theory that some people (scientists) often experience the world independently of social representations, and most people sometimes experience the world independently of social representations (in taking hold of the unfamiliar), then why should the social scientist accept Moscovici's claim that talk of phenomena such as opinions and attitudes should be supplanted by talk of entities - social representations - which are in some way essentially social? And if such a reduction can be achieved, this gives the lie to the claim that the community itself settles the proper way to apply its own concepts. It seems as though Moscovici once more faces a dilemma: either he must abandon his account of the way the community settles the content of social representations or he must amend his explanation of how such representations 'hook onto' the world.

History and the social representation
The problems which Moscovici's theory faces with regard to the relationships between reality and illusion, between science and the reified universe and between the familiar and the unfamiliar stem from the fact that the social representations proponent can make no use of the idea of a standard of correctness in concept usage for the community as a whole. It might seem, therefore, that one solution to these problems would be to amend the theory to
make room for such a notion. It is important to be clear, therefore, that such a move is expressly ruled out by the way in which, on Moscovici's account, communal conceptual usage is said to constitute a standard of correctness for the individual's usage.

The key to the constitutive aspect of social representations theory is Moscovici's insistence that social representations historically determine the way individuals see the world and react to it. Our representations are said to grow out of earlier traditions and to prescribe how we experience and understand the world of the present. It is in this historically prescriptive light that we are to take the claim that communal social representations constitute our environment - they constitute our environment because the present world is, for us, determined by the world of the past. [C, p.41] Now it may well seem that, in general, Moscovici's emphasis on the historicism of human experience acts as a valuable counter to perspectives which treat experience in an over simple, de-contextualised fashion. But this way of expressing the fact that the community constitutes a standard of correctness for individuals' concept application is problematic. First, it leaves no room for change in the content of social representations; individuals are regarded as nothing more than victims of their representations, with no power to depart from communal usage. Second, this result rules out the idea that the community itself could be pictured as responsible to the only possible communal standard of correctness offered - the world as
it is 'beyond' social representation. These can be outlined in
turn.

Individuals are victims of social representation

The root cause of the first difficulty lies in the strength of
Moscovici's formulation of the historical thesis. As we noted
above, Moscovici claims that social representations 'impose
themselves upon us with an irresistible force' in such a way that
'nobody's mind is free from the effects of the prior conditioning
which is imposed by his representations'. And this is to be
understood in the following way:

'...our manner of thinking, and what we think, depend
on such representations, that is on the fact that we
have, or have not, a given representation. I mean they
are forced upon us...The peculiar power and clarity of
representations...derives from the success with which
they control the reality of today through that of
yesterday...'

[C, p.10]

If what we think is governed by our social representations, and if
our representations are irresistibly imposed upon us by the
reality of the past, then there is no room left for the notion
that we can nevertheless influence our social representations or
experience change in their content. For the question of the
content of a representation is settled by the past, and it is the
representations which govern us, not we them.

Now it is important to be clear about this claim. It is not
suggested that Moscovici denies the possibility of change in the
content of a social representation; nor that he denies that
individuals influence the nature of social representations through
their employment of those representations. Moscovici explicitly
propounds both of these contentions [B, p.950]. Nor is it suggested that there could not be an account of the past influencing the present which allowed for change and individual involvement. The problem is that Moscovici's formulation of the 'past influences the present' thesis is so strong that it rules out any notion of individual influences or subsequent change. He claims that social representations are both changeable and open to individual influence, but this claim is just inconsistent with what he has to say about the historically prescriptive nature of such representations.

There is a potential response to this, which is to claim that social representations, in an important sense, just are their instances of application by individuals and groups; all there is to a social representation is the occasions on which that representation is employed in communication. This would be one interpretation of the kind of dependence inherent in social representations described earlier. The suggestion here would be that a social representation cannot help but be conditioned by the specific, contemporary uses to which it is put by individuals, just because that social representation is nothing but those uses.

Such a response faces two hitches. On the one hand, it seems to contradict Moscovici's claim that social representations are more than mere collections of specific historical utterances. [C, p.9] On the other hand, it seems to leave no room whatsoever for any prescriptive content. If a social representation just is its instances, there is nothing left over, in the Moscovici account,
in terms of which one can make out the notion of prescription. It
seems, then, that this extreme form of dependence is completely
inconsistent with Moscovici's historical thesis, and thus cannot
be a response coherently open to Moscovici.

*Standards of correctness for the community are ruled out*

The initial difficulty of the passive, victim-like status of
individuals in the face of communal usage results in a second
problem. For it should, by now, be obvious that the prescription
thesis is, in virtue of this first difficulty, incompatible with
the idea of the community being responsible to a standard of
correctness in concept usage in its own right. If it made sense to
think of the community as responsible to a standard of correctness
in this way, then it would make sense to conceive of some
individual who was, relative to such a standard, correct in his
concept applications whilst, at the same time, being in
disagreement with the community at large over such applications.
But the 'victim-status' difficulty reveals that this consequent is
ruled out by the strength of the prescription thesis. Wherever
there is disparity between the individual and the community as a
whole, the decision must always go in favour of the community -
the strength of the 'prescriptive' aspect of social
representations theory rules out any other choice. It follows
that, by the same token, the antecedent is ruled out. In other
words, the strength of the prescriptive thesis prevents the social
representations theorist from augmenting his theory with the
addition of a clause allowing for communal standards of
correctness.
Summary

There are, then, four areas of difficulty with social representations theory.

(1) The stress on the constructive influence of representations on all thought allows no room for the idea of communal correctness or incorrectness. There is no representation-free way of identifying which representations are veridical and which are not. In consequence, there is nothing left to the idea that communal application of concepts is itself a rule-guided matter.

(2) Likewise, the proposed distinction between scientific knowledge and knowledge based on representations fails because there is no coherent way, within the theory, for justifying why scientists should be any more immune from the effects of representations than other people. In this respect, there is no scientific criterion to act as a standard of correctness for communal concept usage. If there were, reductionism would beckon.

(3) The central cognitive mechanism for dealing with novel objects gives no proper account of how the process of 'taking hold' takes place. Moscovici has it that all cognition is based on representations, yet novel phenomena are exactly those with which representations cannot deal. If it could, reductionism would again threaten.
The historical thesis which Moscovici propounds is so strong that it leaves no opportunity for representations to change and leaves people as puppets of those representations. It thereby rules out any attempt to supplement social representations theory with some version of a communal standard of correctness in concept usage.

Most of these difficulties stem from the fact that the social representations theory leaves no scope for the idea that the community as a whole is responsible to a standard of correctness. Moscovici recognizes that our applications of concepts ought to be responsible to something - even at the communal level, Moscovici seems to accept that there is more to the world than our representations of it. The problem is that there is no scope, in a social representations theory, for any conception of correctness at the communal level other than to make appeal to a transcendent 'real world'. But this leaves the social representations theorist at a loss as to how we come into contact with this real world.

94 SOCIAL REPRESENTATIONS QUA REPRESENTATION

The most important feature of social representations theory is that it provides a form of analysis which is able to deal with specifically social aspects of behaviour. By emphasising the way groups and communities make joint sense of their world, the social representations theorist is able to stress the importance of the effects of specifically social phenomena on individual behaviour. Unlike its more individualist rivals, such as attitude theory and
attrition research, the social representations research strategy explicitly makes room for the influence of social and cultural factors which seem out with the explanatory reach of the individualistic perspective. Moreover, it is in virtue of this emphasis on social issues that the social representations analyst is able to provide for an explanation of thought and language which emphasises their constructive roles. Just because we do see the world in terms of what has already been given to us by our social background, we find ourselves creatively making sense out of what might otherwise be puzzling, incoherent experience.

But even if social representations theory paves the way for a social perspective which emphasises both the social and the constructive aspects of human behaviour - especially human linguistic behaviour - the argument of the preceding section suggests that the theory nevertheless contains a number of flaws. Most of these flaws stem from the fact that social representations theory leaves no room for the idea of standards of correctness at the communal level. An individual is pictured as correct, relative to some group, if his representations accord with those of that group. But there is little resource for explaining how the group as a whole might be viewed as correct or incorrect. It was, the suggestion ran, this lacuna which prompted Moscovici to make appeal to his conception of a world beyond the representation.

The social representations theorist does, though, have a further rejoinder. His theory centres on the idea of representation. It follows from this, it might be claimed, that the lacuna can be
filled. That communal usage is itself responsible to some standard of correctness is guaranteed by the fact that social representations are \textit{representations} - in virtue of which they can be thought of as accurate or inaccurate, depending on how the world is. Of course this reply will only impress if the social representations theorist can demonstrate why the representational aspect of the theory is not merely yet another contradictory aspect of the social representations perspective. What is needed is a fuller explanation of the way the \textit{representational} aspect of social representations functions.

\textbf{Representations as 'direct copies'}

The general idea behind social representations theory is that in thinking about the world, all of our joint communications are done through the use of representations which are representational, cognitive states - in such communication, we essentially rely on the representing aspect of these cognitive states. So members of the same social group understand one another precisely because they already share the same representational, cognitive states - they agree in representing the world to themselves in the same way. They carry around with them the same stock of images, metaphors and so on through which their joint experience is channelled, and in consequence each member of the group understands that experience in the same way as the rest of the group. For the social representations theorist, then, we communicate by means of these representational, cognitive states - to the extent that when we communicate about our social world we are communicating about these social representations. And we know

201
that these are cognitive representational states because a social representation represents by being a cognitive copy of something in the world.

In effect, then, the general idea behind social representations theory is the following. Whenever I communicate, I represent the subject matter of my communication to myself and others. Whatever I represent to myself is a (social) representation. So whenever I communicate, what I communicate about is a social representation. But representations are representational, cognitive states. So what I communicate about is a representational, cognitive state.

There is, however, a flaw in this line of argument. What I represent to myself is something in the real world—a book, or a car or whatever. But a book cannot be 'in my head', even if 'representations' can be said to be. So when I say 'representations are representational, cognitive states' I cannot be referring, through 'representation', to the same thing I refer to when I say that 'what I represent is a representation'. It follows that even if it is correct, in some way, to say that all communication is communication about social representations, it does not follow from this that to study such social representations involves studying the form or content of representational, cognitive states. For when we say that communication is about representations, we must mean something different by 'representation' than what is meant when it is said that a representation is a mental or cognitive representational state. So when the social representations theorist says that all
communication involves social representations understood as representational, cognitive states, he or she must mean that there are two separate elements to a social representation. There is an element in the real world and a corresponding element in the mental world.

To see what such a bi-partite social representations theory is like in practice, consider the following example. Suppose I pay a visit to an inner city and that, while there, I experience a variety of unusually violent activity - fires are started, cars damaged and so on. On the traditional social representations view, social representations are representational, cognitive states through which we structure our experiences of a real world. So if, subsequently, I were to describe my inner city experiences by saying 'I disapprove of the disturbances I saw last week' I would, on such a theory, be telling my listener something about myself by using a social representation. The representation I used to express my thoughts would, on this view, have two elements: what I saw in the real world - blazing shops, dented cars and so on - and my mental copy of those events, embedded now within the representation 'disturbances'. In this way, the traditional theory would conclude, I am able to give expression to my experiences of the real world by channeling such experiences through my cognitive representations, in one of which damaged cars, the setting of fires and such like are associated with the notion of 'disturbances'.
If this is to be an adequate account, however, we must discover what the relationship is supposed to be between the cognitive content of the 'disturbance' social representation and the actual instances of fire-setting, car smashing and so on. If such real world events as the damaging of cars are not part of my 'inner', cognitive world, in that my thoughts are not on fire, nor do they have bent aerials, we need to know what the relationship of these events is to whatever it is that does make up my representational, cognitive state. One answer is to say that this representational, cognitive state is a verbal or pictorial copy which directly mimics the real properties of dentedness, of being burnt and so on - on this view, the social representation is a 'direct copy'. Once I opt for a particular representation, I thereby 'grab hold' of matters in the real world because the verbal or pictorial representation provided by the representation directly mimics these real world matters. These real world matters are themselves, nevertheless, independent of such representing so that, ultimately, it is these real properties of the world which my thoughts and behaviour are focussed upon - it is these real world events which my communication is about, although such communication is necessarily channelled through my representations.

This cannot, however, be the whole account of the way the representation and the world interact. The account given so far supposes that social representations are representational, cognitive states which directly mimic some aspect of the real world so that, through such representations, we can
unproblematically refer, in our communication, to the real world. But suppose someone else tells me that an 'anti-apartheid demonstration' has recently taken place. I might well respond favourably to such news — even if the item of news referred, unbeknown to me, to the very acts of car-denting and the like my first informant has just told me about. If my thoughts and behaviour are directed at the real, worldly properties of being burnt, dented and so on, however, then if I disapprove of them, I disapprove of them. But this must mean, if such worldly events occurred at an anti-apartheid demonstration, that I am annoyed by the anti-apartheid demonstration. But by hypothesis, I approve of anti-apartheid demonstrations. On this view, then, it is both true and not true that I approve of some given anti-apartheid demonstration. This means that if we view social representations as directly picturing world events so that, through them, we can unproblematically refer to real world events, we are led into a contradiction. Some modification must be made.

Representations as 'indirect copies'
The difficulty faced by the social representation proponent is, of course, familiar to those used to dealing with intentional concepts. The required modification might seem, for this reason, to be the further claim that social representations work by referring to some aspect of the real world as thought of in some specific way. On this modification, social representations would be viewed not as direct copies of the real world which allow us immediate reference to events in the world, but as indirect copies which only allow us to refer to worldly events as thought of in
one way or another. This would provide an immediate explanation of how I could both disapprove and approve of an outbreak of car-
damaging and so on. Considered as a mere 'disturbance', I might
well disapprove of such goings on. But considered as an 'anti-
apartheid demonstration', I might well commend such acts. On a
social representations account which viewed representations as
representational, cognitive states, one would have to say, then,
that I had two distinct representations - a 'disturbances'
representation and an 'anti-apartheid' representation - which
referred to one and the same set of events. If I had represented
those events to myself by means of the former, I would have
delivered a judgement of disapproval. If I had represented the
same events to myself by means of the latter, on the other hand, I
would have expressed approval of them.

It is unlikely, however, that even this version of social
representations as representational, cognitive states will
suffice. For what we do not have yet is an account of how two
different social representations represent the same event in
different ways. What we need to know is what it is for a set of
real world events to be thought of in two different 'specific
ways'. We need to know what it is to accommodate a given set of
real world events to a 'disturbance' rather than an 'anti-
apartheid demonstration' representation on one occasion and to
accommodate the same set of events to an 'anti-apartheid' rather
than a 'disturbance' representation on another occasion. But this
is precisely what the theory of social representations as
representational, cognitive states fails to provide us with. Just
because each representation is supposed to be a cognitive state through which we grasp hold of and communicate about real world events, there seems little room for the idea that we could view these events in different ways prior to assimilating them to social representations, in such a way that this prior difference in view could explain why one and the same event may be assimilated to different social representations. The problem is: if opting for one social representation rather than another explains how it is that we can think of the same event in different ways, then we cannot in turn explain the idea that we assimilate one and the same event to different social representations by appealing to the fact that we can think of the same event in different ways.

In addition, it is no longer clear, on an indirect-copy account of social representations, what the principles are which underlie the assimilation of two different sets of events to the same social representation. If the relationship between world events and representational, cognitive state is determined by thinking of the events in some specific way, an explanation is required of the fact that two different sets of events are assimilated to one and the same social representation by being thought of in the same way. We cannot explain their being thought of in the same way by claiming that they are viewed through the same social representation. For, once again, it is the very fact that they are assimilated to the same social representation which we are attempting to explain.
In sum, then, it can be seen that there are crucial difficulties with taking social representations to be representational, cognitive states. On such an imagistic view, we must separate out (a) the real world events which people communicate about from (b) the representational, cognitive states which they use in representing such events. But this means that there must be some account of the relationship between these real world events and the representational, cognitive states which represent them. This relationship cannot be explained by claiming that we refer directly to such worldly events through social representations which somehow mimic them - the claim that social representations are direct copies. If reference to real, worldly states were as direct as this, then whenever I think about such events in conflicting ways, I would have to be taken as expressing a contradiction to myself. In the example given about the street disturbance/anti-apartheid demonstration, for example, I would have been in the position of thinking: 'I both approve and do not approve of this episode of car-denting'.

It is because it ought to be possible to hold such conflicting views without explicitly affirming a contradiction that the idea was introduced of using different representations to think about the same set of events in different ways - the idea of social representations as indirect copies. This allows for the case where I unknowingly think conflicting thoughts about the same set of events. However the idea of social representations as indirect copies - the idea of using different social representations to represent the same set of events in different ways - introduces
two further difficulties: the social representations theorist is unable to explain how two different social representations represent the same set of events and is likewise unable to explain how one and the same representation can assimilate two separate sets of events.

§5 REPRESENTATIONS QUA IMAGES

It might be tempting to respond to these difficulties by claiming that, as representational, cognitive states, social representations can be both direct and indirect copies. We often think of pictures, for example, as directly copying the world. Equally, though, we often say that one and the same picture can represent a general theme which could have many different instances in the real world—we say that the picture represents Anger, or Piety, or Betrayal and so on. In this sense, we could say that pictures are also indirect copies—two pictures can be said, in different ways, to picture the same event thought of in different specific ways, while one and the same picture can be said to indirectly represent a number of different events thought of in the same specific way. And if this can be said of pictures, it might seem that it can be said of social representations too.

If pictures can have both direct and indirect aspects, this seems to imply that social representations, with their iconic elements, likewise can be viewed as copies of the world which are both direct and indirect. So it might seem that as long as social representations are viewed as images, which can be thought of as
both direct and indirect copies, the problems introduced earlier by supposing that a social representation must be either a direct copy or an indirect copy can be avoided.

As an example of how this idea of social representations as images might be used, consider Moscovici's notions of anchoring and objectification. Representation is, according to Moscovici, basically a system of classification in which novel experiences are 'anchored' to existing frames of thought in such a way as to make the 'unfamiliar' 'familiar'. The process of classification requires the thinker to locate an aspect of the world within a particular class of phenomena. One aspect of such classes is that they contain members which can be regarded as prototypical. If we regard representations as images, this means that such classes provide, among other things, a 'photo-kit' [19] which represents a 'test-case' for class membership, summing up the common features of the member cases. This 'photo-kit' is to be considered, on the cognitive social representations theory, as an 'iconic matrix' of observable (or at least decidable) features. This iconic matrix is inspected by the 'mind's eye' and the appropriate decision as to class membership or exclusion is then made for each given experience. Moscovici offers as an example of such a photo-kit the image of a Frenchman as an undersized, beret-wearing, French-loaf carrier.

The image which is the 'test case' or paradigm case underpinning the process of classification is separated out from the other cases in its class in virtue of its especial vividness. Because of
this vividness, the test-case image leaves a particularly deep impression on our memories and is therefore readily available to the thinker in the future as a 'model' against which future experiences can be gauged. This allows us to objectify our experiences. We deal with the concepts associated with experience by representing them to ourselves in image form. Even when expressed in words, such concepts are represented to us through non-verbal equivalents (except in certain cases of taboo). So in general, words are merged with the 'figurative nucleus' which forms an important part of the representing process. The eventual outcome of this objectification process is that the image we associate with a concept becomes, for us, reality. Images become elements of reality rather than merely elements of thought. In the end, these iconic 'replicas' come to constitute our environment.

Unfortunately, the appeal to direct/indirect images does not, in fact, help the cognitive social representations theorist to avoid the problems which were raised earlier. According to the idea of social representations as cognitive images, when we understand the unfamiliar by assimilating it to the familiar we represent an unfamiliar experience to ourselves as, in certain respects, like a picture which is already before our 'mind's eye'. But what still needs to be explained is how the thinker's image, on this view, can be thought of as a classificatory mechanism depicting events 'in some specific way'. We need to know, in other words, what it is about the image which guarantees that the thinker 'sees' it with the 'mind's eye' in the correct way.
Chapter 4

The thinker is to 'see' the image as an example of the given class, in that he or she sees it as depicting events in some specific way. But there seems to be no room, in the cognitive social representations theory of representing, for an account of how the thinker makes sense of the image as an image representative of some particular class rather than another. There is no room, that is, for an account of how the thinker sees the image as depicting events in one specific way rather than another.

In effect then, the same problem of interpretation which arose when we considered representations in general recurs when we consider representations as images. The problem is that there seems to be some process of interpretation required in order to make sense of an image as a classificatory mechanism - as depicting events in one specific way rather than another. But the theory itself offers no account of what this further interpretative process might be. It is, naturally, open to the theorist to claim that the thinker's view of the image through the 'mind's eye' is itself a matter of representation - the image is represented as being such and such or so and so. But what would then seem to be required is another image which will allow the thinker to correctly classify the first image. And the difficulty with this idea is that it leads to a damaging regress.

What is more, the appeal to images as the explanatory basis of the cognitive social representations theorist's account carries special difficulties of its own. The theorist claims that the paradigmatic images which underpin the representing aspect of the
classification procedure are particularly vivid. But this says nothing about why a given image should be regarded as a vivid example of one classificatory schema rather than another. Indeed the appeal to such a notion seems dangerously open to counter-example. If the paradigm or test-case image associated with a class is one readily remembered on account of its vividness, how does the theory prevent spurious occurrences of test-case memories? Vividness is presumably partly a contextual matter - memories of a few moments ago are sometimes more vivid than memories of years gone by just as a consequence of their relative recency. But even though social representations are to be taken as dynamic entities, there must be room within social representations for an account of the way the test-case images associated with a given class are somehow linked up to that class. There must, that is, be some account of the way the vividness of paradigmatic images differs from the vividness of images which arises out of mere recency. Otherwise classes will come to be associated with any old image which happens to be especially vivid in consequence of being associated with recent experience. The difficulty is that nothing the theorist says about the vividness of paradigmatic cases makes room for such a distinction.

It might be said that the image held before the thinker's 'mind's eye' is simply remembered by him or her to be that image which is associated with the particular class in question. But this merely pushes the problem one step back. For the cognitive social representations theorist offers no account of the interpretative
procedure through which the memory of that image comes to be associated with the relevant class.

There is, moreover, a worry about taking the notion of seeing images with the 'mind's eye' to be an explanation of what it is to represent something to oneself. To begin with, such an 'explanation' offers no account of how the 'mind's eye' 'sees' the image. An ordinary account of seeing will refer to the use we make of our eyes, to perspectives, to seeing conditions and to the possibility of being mistaken. None of this holds for seeing images with the 'mind's eye'. More importantly, perhaps, the cognitive social representations theory of representing offers no account of when two images are the same image. According to the theory, whenever the thinker assimilates experience to some given class, he or she likens it to the image associated with that class. But this leaves open the question of how, on two different occasions, the thinker can be certain that he or she has hold of the same image as on the previous occasion. There are checks which can be performed when one wishes to ascertain whether the object one sees now is the same object as the one one saw previously. But these will involve precisely the kinds of empirical investigation which are unavailable in the case of trying to establish whether two images are the same. The best the thinker can do in ensuring that two images are the same is to ensure that they represent the same thing. But this presupposes that we have a means of establishing the identity of things which is, in principle, independent of our conception of them via social representations.
And such a form of 'direct' conception of the environment is explicitly ruled out by cognitive social representations theory.

Even if these problems could be resolved, the cognitive social representations theory needs to explain how this process of representing through images can possibly constitute a social phenomenon. Moscovici's claim is that images are necessary for social communication. The process of representing which establishes an individual's social representations likewise establishes his social persona. A thinker will take himself to share the same world view as another just when he or she is convinced that that other shares his or her social representations. On the model of representing offered by Moscovici, this means, in part, that the thinker's communication and social involvement with others is dependent on a common stock of images. But this relies on the assumption that images could be even a part of the communicative process. And there is good reason to cast doubt on this assumption.

To see the nature of such a doubt, suppose we accept the claim, made by Moscovici, that most of our words and all of our ideas are linked up to images which play a central role in inter-communication. (Suppose that difficulties over how the thinker is to establish that the way he or she takes the image now is the same as the way he or she took it before have been resolved.) This still leaves the question of the role such images play among individuals. No individual has direct access to the images 'inside the head' of the other individual. All that each individual has to
go on, in interpreting the behaviour of others, is the outward consequences of the experiences, thoughts, conceptions and so on of those individuals. This seems to indicate that if a process is essentially social, then its explanation must rely, not on phenomena such as images which, if they exist at all are of necessity hidden, but on phenomena which are, themselves, essentially social. But this suggests that, far from being an essential part of social communication, the images conjured up through representing experience are exactly the kind of phenomena which are ruled out in providing an explanation of how such inter-communication functions.

It seems, then, that the cognitive social representations theorist ought not to appeal to the idea of social representations as cognitive images. Far from being an additional explanatory resource which allows the cognitive social representations theorist to avoid the difficulties which arise from treating representations as either direct or indirect copies, the theorist's appeal to images creates further problems of its own without contributing a solution to the original difficulties raised by consideration of social representations' direct or indirect interpretative functions.

**Recap**

It has been argued in the present section and in the preceding one that social representations theory mistakenly relies on the idea of mental entities which have representational features in providing an account of social life. The basic difficulty is that,
for it to function as a communicative medium, the social representation must be understood in the proper way. We said at the beginning of §4 that appeal to the representational aspect of social representations was the last hope the social representations proponent had of retaining a standard of correctness to which communal concept usage might be held responsible. With the difficulties subsequently outlined in the §4 and in the present section, this hope must be considered dashed.

In a way, there is a connection between the difficulties faced by social representations theory here and the more general difficulties raised in the first two sections. The general problem faced by social representations theory is that it offers no conception of correctness or incorrectness at the level of the community as a whole, beyond unavailing appeals to a transcendent world beyond representations. The theory's version of what, in Chapter Two, was termed the CT thesis allows no room for an accompanying version of RT correctness. The suggestion was that the representations theorist could reply with an appeal to the representational aspect of the theory. So long as the social representation is representational, this implies that we must regard it as having the property of being more or less correct, in the same way that a picture can be more or less accurate. But the problem faced by the social representations theorist who emphasizes representational functioning is that he gives no explanation of how to take the representation, or see the representational image, in the correct way. In fact, intending such and such by a given conceptual image or representation is
analogous to meaning such and such by a particular formulation of a rule for the application of that concept. This suggests that the social representations theorist's search for a set of images which stands behind communication and needs, of itself, no further interpretation may have been misguided from the start. For this process is analogous to the meaning theorist's misconceived programme for the substantive specification of linguistic meaning.

It seems, then, as though social representations theory faces a final dilemma. On the one hand, it could continue to emphasize the prescriptive aspects of representations, at the expense of explaining what counts as correct communal representation. On the other hand, it could provide an account of correctness at the communal level by appeal to the imagistic aspect of representations, at the expense of being committed to non-re-interpretable images which mimic the meaning-theorist's discredited 'knowledge that' format. On either arm of the dilemma, the resultant theory will prove unsuitable as a background for social scientific theorizing.

96 CONCLUSION

Moscovici claims that social representations theory offers a social scientific explanation of human life which takes account of the 'meaning-imbued' character of human activities. There turned out to be a number of difficulties with this claim. The only way in which the theory leaves room for the idea of a standard of correctness to which communal concept application might be held
responsible is by appeal to the 'real' world beyond communal representation. But the prescriptive aspect of the theory does not allow room for such a standard of correctness. In consequence, Moscovici's account of the relationship between individuals and the 'real' world contradicts what he says about the role of social representations in social life. A similar contradiction is generated by the account offered of scientific reasoning. Both 'taking hold of the unfamiliar' and scientific theorizing would allow us a form of direct access to the real world prohibited by the account of social representations on offer.

It was suggested that the representational aspect of social representations theory might offer a conception of communal concept application being responsible to some standard of correctness: representations and images, after all, can be misleading, incomplete or even downright incorrect. But the idea that representations or images could perform the communicative role assigned to them by Moscovici founders on the fact that, for such representations, there must be some further interpretative process standing behind our use of them. What all of this implies is that a social scientific explanation such as that offered by the social representations theorist needs to take on board both aspects of the criterialist framework. If it is to explain the way in which communal members create the meaning of their environment by constituting their own standards of correctness for concept application through communal application of such concepts, it must, at the same time, allow room for the idea that such communal usage is itself open to regulation. In addition, the difficulties
which are involved in the social representations perspective suggest that the assumption central to the view of social representations as cognitive phenomena - the assumption that social representations are representational, cognitive states - should be abandoned.

This demonstrates in a negative fashion that the adoption of RFA imposes certain constraints on the way correct social scientific theorizing can be done. In the next chapter, we move on to investigate a more positive characterization.
CHAPTER 5

§1 INTRODUCTION

In Chapter One and Two, it was argued that theoretical accounts of meaning such as those offered by semantic Realists and Anti-realists should be abandoned in favour of a rule-following approach. This led to the development of the RFA perspective. In Chapter Three, it was argued that RFA allows for a version of criterialism. The criterialist viewpoint was seen to offer a way in which social scientific practice could be described in terms of a theory/observation model. The only constraint was that where such theories deal with the meaning, to participants, of the activities observed, the explanation of meaning provided should make room for both CT and RT elements (the terms introduced in Chapter Two). The importance of this requirement was seen in Chapter Four, where a popular social scientific theory was criticized for lacking an RT element.

The aim of the present chapter is to draw some conclusions from all of this for the practice of social science. The social scientist's observations attempt to record the acts of his subjects: what the subjects did and what aspect of the world their acts were aimed at. The accompanying theory attempts to explain these acts, usually in terms of reasons or of causes. It follows from this that a successful explanation depends on a correct understanding of the meaning, for participants, of the relevant act and of the intended object [1] of that act. For in order to
produce explanation in terms of reasons or causes, the social scientist must ensure that he correctly understands the nature of the act to be explained.

There are, then, at least two basic aspects to social scientific practice: providing an appropriate description of subjects' acts and providing an explanation of those acts. [2] In Chapter 6, we will examine some of the implications which RFA carries for theoretical explanation in the social sciences. In the present chapter, we will concentrate on the implications for the former, descriptive, task.

Proper description
Just because the scientist is interested in 'participants' meanings' in this sense, it does not follow that he can discover such meaning merely by asking participants. The model of meaning in use, RFA, supposes that as far as linguistic expressions are concerned, their meaning may not be clear to the users of that expression. Their usage may not be 'perspicuous'. In such cases, explicatory work (such as re-presenting everyday explanations, examining exceptions, corrections, amendments and formulating language-games of various sorts) is required to prevent the adoption of faulty analyses of such terms. An analogous task may be required wherever the social scientist intends to describe the meaning of some social act. It may be necessary to study the way that act is explained or described by participants and to examine how such descriptions are used, accepted or challenged. The point of such a study is, then, to describe the meaning of a particular
act for the participants involved, a description which will centre on the way the community as a whole would ascribe meaning to this act through rule-following practices such as explanation and correction.

How should the scientist approach this descriptive task? Clearly, he must begin by recording participants' verbal (and nonverbal) behaviour in a wide range of contexts beyond that of simple question/answer formats (since simple verbal accounts given by participants themselves have already been pictured as potentially unreliable indicators of the meaning involved). So the question now becomes one of how (leaving aside the nonverbal) to deal with these verbal data. The general answer already given lies in treating such verbal behaviour as rule-following activity in which the meaning of a given act is constituted and regulated. But how, in practice is the social scientist to study such activities? Note that the question here is not one of discovering how such activities can be constitutive of, and regulative for, meaning—the answer to that question is a philosophical one (and was provided by the philosophical argument in Chapter Two which established the RFA as an account of meaning). The question is, rather: given that meaning is constituted and regulated by communal usage, what is it to provide a social scientific description of such usage? The general (empirical) answer to this question depends on isolating how explanations, corrections, exemptions, challenges to descriptions and the like function in their constitutive and regulatory roles. But this involves investigating matters such as finding out what participants regard
as acceptable explanation for some act or other, and discovering how participants create, maintain and challenge a particular description of that act. How is this done?

**Discourse Analysis**

There is at least one approach to the study of social phenomena which provides for exactly this requirement. According to this perspective - Discourse Analysis [3] - the way people understand social phenomena is an outcome of constructive processes which are essentially communicative, being founded upon a manipulation, by the individuals involved, of readily identifiable discourse processes. These processes underlie individuals' abilities to present particular descriptions of an act or event as the 'natural' description. They also allow individuals to challenge competing descriptions, correct 'misleading' descriptions and so on. The importance of isolating such processes lies in the fact that many everyday verbal interchanges rely on a series of syntactic procedures which go unnoticed by participants and by the untrained observer. When a participant wishes to challenge a description, for example, he is likely to do so by by means of a complex 'topic changing' manoeuvre (see §2) rather than by overtly identifying the description and rejecting its application to the act or events in question.

So the discourse analyst's findings centre on the question of how individuals' verbal interactions are constructed and on the effects such constructions have. [4] And the general claim made out by means of such analyses is that a study of the discourse
Chapter 5

represented by individuals' verbal interactions allows the analyst to describe the meaning which those individuals attach to the events and activities discussed in those interactions.

The idea of the community constituting the meaning of acts presupposes that at some stage the meaning of such acts may be under debate. For this reason variability in expressions of what an act means is one of the central concerns of discourse analyses. The discourse analyst is interested in discovering the way in which explanations, say, of the same act systematically vary, and in establishing the function which such variability plays in the sense-making activities of the participants. Now participants are, themselves, sometimes aware of the potential tensions which arise out of the existence of such competing accounts. But the general finding of discourse analytic studies is that participants resolve such tensions by using linguistic resources in a strategic fashion when producing competing accounts. [5] It follows from this that discourse analysis pursues two different but related tasks.

(A) To identify competing accounts of the meaning of a given act in terms of differences in content within such accounts

(B) To identify the strategies by which such accounts are creatively introduced and defended in a wide range of different types of discursive settings (including formal and informal interview settings, everyday discussion settings, textual settings and so on)

The outcome of this is that the accounts collected by the social scientist are at one and the same time highly variable and yet highly regular. The variability arises out of the differences in meaning attributed to a set of actions and events by the varied
accounts. The order arises from the fact that, in establishing and defending such accounts, the social scientist's subjects regularly rely upon the same rhetorical strategies. [6] The former aspect will be dealt with in Chapter 6. Here, we will restrict our attention to the task of identifying ways in which such competing accounts can be established. Once this is done, we will be on the way to establishing how proper description in social scientific observation should be carried out.

Recap

We began by noting that RFA allows for the idea of theory and observation in the social sciences. A pre-requisite of proper observational description, however, is that the social scientist take account of the meaning of what is observed for the participants involved. The goal of the rest of this chapter is to provide a model for such meaningful description. Such a description must isolate the communal practices which establish this meaning through CT and RT. These practices are the explanations, corrections, debates, re-formulations and so on which constitute the verbal interactions relevant to the act in question. They are the means by which the community constitutes and regulates the meaning of that act. As such, they generate variability whenever there is dispute as to the exact meaning of a particular act. The means by which these practices, and the variability they generate, are studied is the discourse analytic technique. The resultant description of meaning produced by such analysis falls into two halves: the study of differences in content between competing meaning accounts and the study of how
these competing accounts come to be accepted or rejected. Here we will concentrate on the latter of these tasks.

§2 DISCOURSE ANALYSIS AND TOPIC

The aim of this section is to demonstrate how discourse analysis contributes to the social scientist's descriptive task. The data which follow, including the excerpt below, were taken from a discussion in which a group of craftsmen and craftswomen are assessing the work of applicants to a crafts guild. The assessment takes the form of a group examination of items of the applicant's work. During this examination, the assessors discuss among themselves the relative merits of the pieces presented and each assessor then gives an evaluation of the work. On the basis of these evaluations, a joint decision is then agreed as to whether the applicant ought to be accepted into the Guild once each member of the group of experienced practitioners has expressed a view in regard to the craft items on display.

The Guild group's discussions can be viewed as attempts to establish the precise meaning of what they are about. In a range of different cases, the group negotiate the extent to which some aspect or other of the situation is a relevant part of admitting the claimant or turning him/her down. These aspects include: features of the candidate's work, aspects of the candidate; elements of the decision procedure. In the present chapter, we will look only at aspects of the first sort.
Chapter 5

An indication of the complexity which underlies this apparently simple situation can be gained by examining the following segment of conversation:

1: Tape 1, *1 p. 2-3
Mary: His wife makes the cushions.
Liz: Oh! Oh dear! =
Mary: =(laughs) Another problem =(laughs)=
Liz: =(laughs)
David: I met him about two years ago actually, he, em, he's=
Liz: Oh really?=
David: =quite an interesting chap. He's em=
Liz: =Mm
David: One of the things he told me he'd done - he'd been - em, a road manager for a pop group. And=
Liz: =Really
David: I think he'd made so much money out of it that em went to this Parum=
Liz: =John Makepeace School
David: Yes. Yes.
Mary: =Really?
Liz: But is, is it?=
David: =It doesn't look comfortable=
Liz: =That one looks positively dangerous.
David: =Yeah.

How is the analyst to make sense of this? It is clear that at least three issues are under discussion here: (i) the fact that not all of the work is done by the applicant; (ii) the applicant's background; (iii) the comfort-value of the seating furniture. The problem is that identifying these elements seems to offer little to the analyst who attempts to make sense of the exchange in order to describe what is going on. Mary and Liz begin by introducing as salient the issue of whether the cushions are made by the applicant. David's reply seems to be the introduction of a quite separate topic: the applicant's background. And yet the episode ends with David and Liz jointly taking up the question of whether the furniture is comfortable. On the basis of this evidence, how ought one to expect Liz, Mary and David to respond to the
applicant's work? What, for them, is the nature of the object and how will their decision relate to it?

The answer to these questions depends on a detailed understanding of what went on in the conversational exchange. In this chapter, the focus of interest will be on demonstrating the way in which subjects construct the meaning of the decision process. They achieve this by selecting from a resource pool of complex conversational strategies in order to pick out certain aspect of the candidate's work as 'relevant' to their task of inclusion/exclusion. Here we will concentrate on analysing the way in which subjects' accounts are organised by strategies which manage the current issue under discussion and allow for the transition from one such issue to another.

The key analytic tool here will be an analysis of conversation topic. Previous studies of conversational topic [7] have demonstrated that conversations can be segmented according to the status of the conversational subject matter. At some points in a conversation, conversants will be in agreement as to what counts as the current conversational topic. At other points, specific conversational mechanisms will be brought into play in order to achieve an orderly transition from one topic to the next. An important feature of such studies is that conversational topic is seen to be an essentially flexible entity. While everyday conversations do exhibit topical coherence, this coherence cannot be rigidly made out in terms of the features of the conversational referent. Typically, the features of the conversational referent
or 'talking point' provide neither sufficient nor necessary grounds for sameness of conversational topic. At times, the same referent can induce two quite separate conversational topics, while at other times two separate referents can be taken by conversants as representing different aspects of a single conversational topic. [8]

The importance of this, in the present context, is that the notion of conversational topic flexibility suggests one way in which conversants can construct for themselves a version of the object under review. If a conversant is able to control conversational topic, and if conversational topic is inherently flexible, then he or she will be able to steer the conversation in such a way that the conversants' discussions centre on a 'version' of the object which has been specifically constructed to suit the requirements of the topic-controller. In the present instance, this allows conversants to influence the meaning ascribed to inclusion or exclusion by demonstrating that certain topics are, or are not, relevant aspects.

Conversations 'On Topic'

Perhaps the most obvious topic form a conversation can possess is that which Sacks termed 'on topic' talk. [9] In this form of talk, conversants are in agreement as to the nature of the conversational topic and generate exchanges which centre on that topic. The conversants' agreement upon a conversational topic allows them the possibility of constructing a single, unproblematically identifiable object of talk.
2. Tape 4, 'Cushions' *2 p.8-9

Stephen: It's a bit old hat really. Em. I don't think a lot of the stripy ones, especially that blue one. None of the stripes match up on that side.

Liz: Yes. it's terrible.

Stephen: That really annoys me on that.

Liz: Yes, I think it's very important.

Stephen: I thought I must be really fussy.

Jane: It is important.

Liz: Oh I think - I absolutely agree.


Mary: And he didn't do the upholstery. But all this is a restoration job.

Liz: What was there before he started on it though?

Mary: Well (laughs) maybe it just needed doing up.

Liz: He's painted it.

Stephen: Yes.

Liz: It's all hard to know, isn't it? Perhaps he's made some of the parts.

Mary: It may have been it may have been bits missing or knocked about. I don't know.

Liz: We can't judge it if we don't know, can we?

Mary: No.

Roger: We don't go in for that sort of thing anyway.

Michael: We don't want people that restore, do we?

Mary: Don't we? Oh, I see. So it’s immaterial.

Michael: Well that aspect of it. I mean if we=

Liz: If that’s a restoration job we aren’t even looking at it.

Michael: No.

Mary: So it's out.

In these examples, the subjects can be seen to be in unproblematic agreement over the conversational topic. In example 2, all three subjects agree in discussing the importance of the question whether the stripes on a cushion presented by a candidate ought to match up. This allows them to establish that, within this context, the matching of stripes will become a salient feature of the candidate's work. In this way, conversants create a particular context-local 'version' of the object being assessed. In example 3, at least four participants in the conversation agree in taking as topic the question of whether items of the candidate's work (furniture) are original or merely restoration. This offers them the chance unproblematically to agree on viewing 'restored-ness'
as an important feature of the candidate's work, thereby partially settling the issue of which aspects are the important features of the participant's work. Roger states that they 'don't go in for that sort of thing' and Liz ends by saying that they will not even look at restoration work, a view with which Michael agrees. So the result of the constructive efforts of the subjects is that even though the viewing area contains both new and restored furniture, the subjects are able to establish that the 'object' is made up of only the new pieces of work.

Of course the value, to conversants, of being able to construct an unproblematic picture of the object in question lies in the fact that they are then able to agree on the precise meaning which their subsequent positive or negative decision will have. It will be seen to follow directly from whatever was said about the candidate's work.

4: Tape 8. 'Furniture' *1 p.5
Lesley: It's nice to see something that isn't dripping with polyurethane.
Jack: Yes, it is, isn't it.
Jill: Isn't it just.

5: Tape 6(1). 'Pottery' *2 p.11
Paul: It's a girl that has done it. I liked the big bowl very much - liked the little decoration round the outside.
Liz: So do I. I think that's delightful.
Roger: It's lovely, isn't it.
Liz: It's so subtle.
Paul: It's so so - you hardly notice it.
Liz: Mm. And then when you do.
Paul: Yes, it's a delight, yes.

In these two examples, the individuals involved agree as to the topic they are discussing and the nature of the object has already been established. Thus in example 4, the three subjects each agree in discussing the fact that the piece of work presented for
inspection (a piece of wooden furniture) has not been coated with a polyurethane varnish. This allows the conversants to agree in affirming that the lack of such varnish is 'nice'. In example 5, Paul introduces as a topic the decoration of a bowl presented by a candidate. This then becomes an 'obvious' feature of the presented object, which allows Liz and Roger to agree in expressing positive evaluations ('I think that's delightful', 'It's lovely').

Of course, as the next example shows, an agreement over the nature of the object under discussion (established by means of acceptance of a particular topic) need not issue in agreement over evaluation:

6: Tape 4, 'Pottery' *3 p. 7
Stephen: I think it's functional and non-functional and I mean it looks very nice.
Roger: I haven't seen anything quite so nasty for a long time.
Stephen: No it isn't.

7: Tape 3(1), 'Furniture' *3 p. 18
Roger: Em. Who recently got in doing woodwork, for example? Well if you said, sort of, 'Pete, look at the turning' to Pete Porter to diagnose and comment and play with the turning; and you've got, as you have done, me on the actual finish of the woodwork and the joins, I think at the breakdown you would probably reject them. Because the technical merits aren't there. I mean, look at the finish on the seat, for example. That's horrid. Yeah.=
Jane: What's wrong with the turning?
Stephen: I can't see anything wrong= rog
Roger: Em.
Jane: No. I can't see what's wrong really.

In example 6, both Stephen and Roger discuss a tea pot presented by one of the candidates. But while both agree on topic and agree in terms of object involved, they disagree on their evaluations, Stephen saying that 'it looks very nice' while Roger expresses the opinion that it is 'nasty'. In example 7, Roger introduces the
topic of the wood-turning carried out by the applicant. This
thereby becomes established as a relevant and salient feature of
the object. But while Roger feels that the turning does not show
the appropriate 'technical merit', Jane and Stephen argue that
they cannot see anything wrong with it.

Moreover, as is seen in the following examples, matters become
more complicated once the question of the nature of the object
itself comes under scrutiny. In these examples, the subjects agree
in talking about a specific topic, but disagree as to whether the
topic genuinely captures aspects of the object which are relevant
or important.

8: Tape 6(2), 'Painting' *2 p.3
Michael P: She could have cleaned up before she painted it.
Liz: Could she have cleaned up?
Roger: Regrettable.
Michael: It's covered with dirty fingermarks. Well em.
Mary: That's not what she's being assessed on, is it?
Roni: Yes she is.

9: 3(1), 'Furniture' *3 p19
Roger: You wouldn't find it. It's just a sloppy thing like
that.
Jane: You mean to have the glue showing?
Roger: You wouldn't have the glue showing. It's not bad -
it's on the borderline, I would say.
Stephen: That's traditional isn't it, to leave it like that. So
you can't in fact criticise the underneath of it.
Roger: I don't think that's strictly relevant.

What these examples demonstrate is that agreement in topic can
still allow for disagreement over the nature of the object. In
example 8, Michael, Liz, Roger, Mary and Roni comment on the fact
that the applicant's work is dirty. But although they agree in
having this as a topic of conversation, they disagree as to the
relevance of this feature of the applicant's work. In example 9,
the agreed topic is the fact that the applicant has left signs of
glueing on his work (a chair) together with the question of whether this is traditional. But Stephen and Roger disagree as to the relative importance of this. Stephen argues that the tradition associated with leaving glue showing means that it cannot be a criticism. Roger responds by claiming that the issue of tradition (in the assessment of a piece of craftwork) is not 'strictly relevant'.

So far, then, what has been seen is that individuals are able selectively to draw upon conversational topics in order to establish the nature of the object under review. In some cases this allows for an unproblematic agreement in evaluations. But it also allows for the expression of evaluative disagreements. The selection of a given conversational topic can, however, be even more controversial, as was seen in examples 8 and 9. There, conversants were seen to disagree as to whether the current topic picked out an important or relevant feature of the object.

One noteworthy feature here is that up to now, when conversants have been attempting to establish the nature of the object by discussing whether a given conversational topic picks out relevant or important aspects of that object, or whether it should be evaluated positively or negatively, they have had to rely, in controversial cases, on outright denial of a co-conversant's position. Thus when Mary, in example 7, says 'That's not what she's being assessed on', Roni is forced to respond 'Yes she is', while Roger, in example 8, has to protest: 'I don't think that's strictly relevant'. This prompts the question of whether
individuals have other conversational devices through which the
danger of such potentially conversation-disruptive clashes can be
avoided. As the next section demonstrates, the Guild members do,
in fact, demonstrate the ability to establish their own versions
of the object, and their own evaluations of such an object, over
those offered by other conversants without recourse to open
disagreement.

The key to this ability lies in the way conversants manage changes
in the topic of conversations. If a feature of the applicant's
work is seen by a conversant as important (or relevant to the
establishing of an evaluation which the conversant holds), he or
she will introduce the relevant topic into the conversation. If
the feature is seen as unimportant (or relevant to the
establishing of an evaluation the conversant does not hold) the
conversant will attempt to steer the conversation away from the
respective topic. Moreover, because of the lack of necessary and
sufficient conditions for sameness of topic, these changes in
topic will not be regarded by co-conversants as untoward, provided
the topic changer follows established procedural practices. In
this way, manipulation of conversational topic offers the
conversant the chance to present a fresh 'version' of the object,
or to amend an existing one, without becoming involved in
potentially blameworthy conversation disruption. The next section
analyses some of the ways in which such topic change can be
effected by conversants by examining the kinds of 'procedural
practice' which are exhibited by the Guild assessors' discussions.
§3. TOPIC CHANGE

**Natural Topic Changes**

The aim here is to provide an analysis of the way conversants influence the nature of the conversational topic in order to establish the nature of the object under review. To understand this conversational ability, however, it is necessary first to consider the way in which conversational topic change can occur 'naturally'. Jefferson [10] notes that in everyday conversations, participants often do introduce a new topic which is unrelated to the preceding topic. Usually, such an introduction is *unmarked*, in that it possesses no special linguistic or verbal features. When this kind of unmarked shift from one topic to another - the 'bland introduction', as Jefferson puts it [11], of a new topic - occurs, conversants typically take themselves to be engaged in a cooperative switch of direction in the conversation. Such topic switches are common enough in the Guild-members discussions:

**10 Tape 4, 'Pottery' *3, p. 2**

Mary: Did you know her, Bert?
Bert: No. Only the foundation course, really.
Roger: The kid was at er Brighton, wasn't she?
Mary: Brighton and
Bert: Foundation at Camberwell.
Paul: Brighton and Camberwell.
Mary: Imagine this when it's full of soup. I can hardly lift it up now.
Liz: Yeah.

**11 Tape 6(1), 'Pottery' *3, p. 3**

Liz: It doesn't say why, how copper can be used.
Michael: Various grades of abrasive mixed with oil are fed to the wheel.
Mary: To the wheel?
Liz: I expect so, yes.
Roger: This is her third attempt to appear before us, isn't it.
Liz: Well the others have only been by 'phone - she hasn't been sort of serious
In example 10, the conversants exhibit no uneasiness in abruptly switching from Bert’s knowledge of the candidate to her educational background to the weight of one of the items of crockery offered for selection. Similarly, in example 11, Roger’s change of topic, from the question of how copper wheels can be used to etch glass to the fact that the candidate’s work represents a re-submission, passes unremarked and unchallenged. In these examples, none of the contestants is attempting to establish a particular ‘version’ of the object under review and, in consequence, each discussant is willing to accept the ‘natural’ switching from one topic to the next.

Clearly, however, such abrupt shifts in topic will not suit the needs of the conversant who, in the context of a Craft Guild meeting, wishes to challenge the current topic in use in order to construct an alternative, or even competing, version of the object under evaluation. Where the easy assumption of cooperation in topic choice cannot be made, the adoption of a policy of sudden topic switching would be likely to draw criticism. There are, however, alternative strategies. Jefferson notes [12] that conversants have available the use of at least four mechanisms whereby conversational topic may be altered. The first three of these can be thought of as forms of ‘marked topic jump’ while the fourth can be thought of as a form of ‘topic flow’. The importance of each of these, in the present context, is that they offer opportunities for the construction of particular ‘versions’ of the object.
Chapter 5

A: Topic Jump via Laughter

In general, the usefulness of marking topic jumps in effecting a topic change lies in the fact that such marking procedures allow the conversant to display to co-conversants acknowledgement of the fact that he or she is making a transition. By following such marking procedures, the conversant is able to exhibit that the topic change about to be introduced is to be taken as a 'legitimate' extension of the conversation. There are several ways in which a jump from one topic to another can be marked. The first of these involves the use of joking. Jefferson reports [13] that individuals who are the recipients of 'troubles-tellings' may attempt to prepare the conversational ground for a switch of topic away from the co-conversant's troubles. The technique involves, roughly, introducing laughter into a conversation and then treating that laughter as a sign that the conversation is apt for the introduction of a new topic.

The present data reveal that a similar process occurs in the negotiation of the nature of objects:

12 Tape 3(1), 'Furniture' "3, p.4
Jane: They comfortable?
Roger: Not bad actually. I feel a bit lost in it.
Jane: (laughs)
Mary: (laughs)
Roger: But otherwise it's all right.
Paul: What is that, yew?
Roger: Yew.
Mary: Yew.

13 Tape 4, 'Knitting' "1, p.10
Mary: I mean (laughs) if I did decide to have her in on the knitting I would definitely say no cushions. But I'm not - I can't make up my mind.
Liz: But she's not a cushion maker.
Mary: Yeah, I've seen that. (laughs)
Roger: (laughs) Nasty.
Liz: (laughs)
Mary: Most of what I've seen I really don't like much, but I
do find quite a few of those - presumably she's sold -
those photographs look good. Em, but we can't judge
things on photographs, can we, so I'll just have to
say, I think, no.

In example 12, which takes place in a discussion of wooden
furniture, neither Roger nor Mary sees anything untoward in Paul's
introduction of the question of which type of wood has been used,
despite the fact that the immediately preceding topic had been the
question of the comfort-value of the furniture. Here Roger's joke
'I feel a bit lost in it' allows Jane and Mary to produce
laughter, the effect of which is to play down the significance of
'comfort-value' as an aspect of the object. In example 13, Mary's
ironicism ('Yeah, I've seen that'), together with the laughter it
produces from Liz, Mary and Roger and Roger's joking accusation of
nastiness, allows the conversants to treat Liz's variant on the
topic of the applicant's cushions: 'But she's not a cushion maker'
as one which can be legitimately superseded by a fresh topic.

In both cases, the initial topic, by being seen to evoke
amusement, is taken to be non-serious, with subsequent effect on
the issue as to whether that topic is regarded as one which ought
to be continued into subsequent conversational exchanges. In this
way, the conversants are able to control the nature of the object
under evaluation. Roger's joke allows the conversation to move on
from the agreement over the comfort value of the chair. Although
he accepts that the chair is comfortable, he is able to minimize
the extent to which comfort value is seen as an important aspect
of the object. Mary's joke allows her to switch the topic away
from discussion of the fact that although she does not like the
cushions, the applicant does not claim to be a cushion maker. In
this way, she is able to ensure that the unpopular cushions remain
a significant aspect of the object.

**B: Topic Jump via Acknowledgement Token**

The introduction of laughter as a topic change marker, in the
construction of a particular 'version' of the object, relies, to a
certain extent, on complicity among conversants. This is less true
of the second form of marking, which relies on the use of
conversational features which normally mark the end of a
conversation - especially the insertion of a number of
acknowledgement tokens, such as 'yeah' or 'Mm'. These can be used
not only to bring an end to a conversation, but to bring an end to
segments of a conversation. Here such markings are used to bring
discussion of a particular conversational topic to a close.

**14 Tape 1. 'Pottery' #5, p. 8**

Liz: Paul, would you like to start?
Paul: Em, they're not very satisfactory, these. Em, it's
clumsy and I don't think I would say yes to it - I
know I wouldn't say yes to it. Some of these glazes
seem to be cracked around the inner side. That's
difficult on, that's difficult on a thing that size
but=

John =Can I have a look at that? Oh, it's a crazed glaze,
yes. That is something that some people worry about,
other people don't. I think by itself it wouldn't -
it's not terribly good because it will stain in use
eventually - but she's used an ash glaze by the look
of it and they nearly always do craze.

Paul: Oh. Er, I see, yes.
John It's perhaps not a suitable glaze for a domestic
casserole
Paul: Yes, em
John The glaze she uses inside is okay, I think
Paul: Mm.
John She's used different glazes on there and inside the
lid, which is odd.
Paul: Yes. Think that'll do Liz?
Liz: Thank-you. Michael?

**15 Tape 4. 'Pottery' #3, p. 18**

Liz: I think a lot of it is tremendous fun, but I think the
effect of all of it is very ugly.
Stephen: But you could say that with a lot of them. I mean if
you put just one piece, you know, I can sort of see it in the home, sort of person who would buy that. But if you put, say, the Millseys' work all together=

Liz: =Mm
Stephen: You know, it's really sort of - and, and Jolanada's as well. And Caroline's=
Liz: Mm.
Stephen: One piece on its own is much nicer, I think you have to see it all in you know the light of that=
Liz: Mm. I think - I'm thinking obviously of the context of a Guild exhibition on a stand.

In example 14, Paul introduces the issue of the cracking which has occurred in the pottery glazing produced by one of the candidates and uses this topic in characterising one aspect of the object in order to express a negative evaluation. John casts doubt on this evaluation by identifying the glaze as of a sort - ash glaze - which usually does craze. Paul response is to draw John's contribution, and thereby the topic, to a close by producing acknowledgement tokens ('I see, yes', 'Yes', 'Mm', 'Yes'). This allows Paul then to switch to Liz, who introduces Michael into the conversation. In this way, Paul is able to minimize the effect of finding that an aspect of the object he had introduced for the purposes of negative evaluation turns out to be associated with a positive evaluation.

In example 13, Liz expresses the view that the crockery offered by another candidate is 'ugly', when 'the effect of all of it' is considered. Stephen counters this claim by introducing the topic of the work of present members of the Guild taken 'all together'. This would, Stephen suggests, have the same effect. Liz responds by bringing the discussion of comparisons to a close through the use of acknowledgement tokens ('Mm'). The real topic, her conclusion implies, is how the exhibits would look on a stand at a
Guild exhibition. In this way, Liz avoids further discussion of a version of the object in which the applicant's own work is depicted as sharing a property in common with the work of other craftspersons who are already Guild members.

C: Topic Jump via Conversation Restart

A third form of topic jump marking is represented by what Jefferson terms conversation restart. Typically [14], a conversation restart marker will be of the form of a 'How are you' type of question. Clearly, there will be only limited, if any, scope for this kind of conversational topic introduction in the context of a discussion whose parameters are set by the evaluative task in hand. Nevertheless, the data reveal that certain kinds of utterances do play the role of allowing conversants to introduce a fresh topic. In Jefferson's data, conversation restarts are typically 'other-directed', in that the utterer typically asks questions about the co-conversant, rather than making statements which are self-referential. Jefferson's suggestion here is that by exhibiting 'interactional cohesiveness', the conversant is able to demonstrate that the topic rupture is, in some way, legitimate. In the present data, certain key utterances can be seen to have the same effect.

16 Tape 1. 'Painted Pebbles' #3, p. 9
Liz: Did Jane actually have any=
Mary: Yes. She said that she would vote against, em, if we would take it as valid, her vote, not having seen them.
Liz: Yes.
Michael: With all due respect, I think if she hasn't seen them=
David: =That's not fair.
Michael: I don't really feel she's entitled to=
Liz: =I think she is because she can imagine however well painted they are - which is a bit how I'm feeling.
Mary: So I did explain=
Liz: Yes. Anyway let's see how we get on. Em, Mary, what
Chapter 5

Mary: about yourself? Well, I'm just I'm gonna vote just in (pause) just.

Stephen: I think it's functional and non-functional and I mean it looks very nice.

Roger: I haven't seen anything quite so nasty for a long time.

Stephen: No it isn't. That is=

Roger: I think that you shouldn't prejudge it Stephen. Look - I know I shouldn't prejudge your answer, but how do you get a setting up grant from the - do you take someone else's work along? Em.

Stephen: The Craft Council? Well you don't know - it's very avant garde, isn't it?

In example 16, Liz is concerned to establish that Jane's viewpoint is an important feature of the applicant's work, despite the fact that the Jane has not actually seen the work in question. Michael interjects that Jane is not 'entitled' and Liz responds with an outright rejection of this: 'I think she is'. Further potential argument on this topic is then forestalled by Liz who suggests a new conversational topic (Anyway, let's see how we get on') and then introduces an opening for that new topic by asking Mary to express her opinion. Note that Liz introduces a new topic slot by asking one of the other conversants, Mary, about her opinion. This parallels the kind of 'other-directedness' found in Jefferson's 'How are you' questions. (Note also that Liz employs a single acknowledgement token as a further intimation of topic change.)

Example 17 represents the continuation of the segment of conversation which appeared in example 6. There, it will be remembered, the co-conversants found themselves on the brink of a potentially disruptive dispute. Example 17 demonstrates how such a dispute can be resolved through the use of conversational restart procedures. In the first portion of example 17, (the section which
appeared earlier as example 6), Roger's evaluation initially looks like producing conversational problems, since Stephen feels forced to issue an outright denial. At first, Roger's response is to accuse Stephen of prejudging. But then Roger amends this to a self-ascription of prejudging, which is immediately followed by a question which introduces a new topic, which Stephen takes up, and which is pursued for a number of subsequent conversational turns. Once again, it is noteworthy that the utterance Roger uses in order to introduce a new topic, is designed to show an interest in Stephen - it is 'other-directed' - rather than at Roger himself. In this way, Roger is able to circumvent a troublesome analysis of his own evaluation, by restarting the conversation.

So conversational restarts allow the conversant to steer the conversation onto new topics. And the value of this in the present context, to the conversant, is that it allows him or her to make a controversial statement (such as an evaluation or a claim about the nature of the topic) and then to change the topic of conversation before co-conversants can develop arguments in reply. It is also important to note that both examples of conversation restart feature another aspect of markedness typical of topic changes. In both cases, the individual who introduces the new topic produces expressions of hesitancy. In example 14, Mary says 'Well, I'm jus - I'm gonna vote just in'. This faltering expression is then succeeded by a lengthy pause, before she concludes with 'just'. In example 15, Michael cuts off his sentence: '..but how do you get a setting up grant from the..', and then reformulates it as: 'do you take someone else's work
along?' This is concluded by a hesitant 'Em'. Such indications of hesitancy are typical of segments of conversation in which a new topic is brought into view [15]

D: Topic Flow
The second overall stratagem which conversants have at their disposal in their attempts to construct versions of objects through the manipulation of conversational topic changes is a procedure which can be termed 'topic flow'. Jefferson [16] describes topic flow as having four main features: (a) discussion which centres on an initial topic throws up a comment which introduces an ancillary matter; (b) the co-conversant then works to stabilise the conversation around this ancillary aspect; (c) this aspect is then interpreted as being in common with some new topic, and the conversation correspondingly ' pivots'; (d) the conversation then proceeds to follow the new topic.

The data reveal that topic flow represents a further way in which the Guild assessors are able to construct versions of current object:

18 Tape 3 (1 & 2). 'Furniture' #5, p. 18-19
Roger: On the - I notice this - on this, you see, if anyone is normally cutting back, you would cut it back if you wanted to with a chisel. Well he - admittedly he says he's self taught, he is self taught, isn't he=
Mary: Mm.
Roger: And he's just sawn it straight off. well at least he shouldn't be doing that, er, so I'm not sure he knows how to make it better.
Michael: It looks as though you could help him by discussing things.
Jane: I mean= Liz: =I, yes Jane? Jane: I don't know, I mean how has he learnt how to do all these things? I mean has he been getting all the
information out of books, or is it trial and error?
Has he looked at the construction of the, you know,
the chairs that he actually used to copy, or or what?
You know ot's=

Liz: Well he doesn't say. All he does say is he's been
doing it for two years and I think for two years his
achievement is considerable. When you think of the
incredible messing around that most people show us=

Jane: Mm.
Liz: Perhaps after doing it for five years.
Jane: Mm
Michael: Yes.
Liz: Again I think that really speaks in his favour.

19 Tape 1. 'Painted Pebbles' "3, p. 4-5
Michael: Well I like the relationship to the pebble - that
she's selected it, as David says, very carefully.
David: Yes.
Liz: I, I always feel that it's not - although she's
selected it carefully I mean insofar as you can - but
it's not so much the shape of the cat - even so that
the perspective effects are so peculiar - that just
animals painted on pebbles I don't feel justify
themselves.

John: Well that's certainly what I felt.
Liz: But for this kind of thing these are about the best
I've seen.
David: Mm.
Liz: In fact I've been wondering all day what I would
accept that was painted on pebbles.
John: Mm.
Mary: Oh I don't know=
Liz: =I really felt that=
John: Pebbles are not for painting on.
Liz: It needs to be something abstract. Something of Escher
or Vasarelli that would geometrically relate to the
shape and=
David: Well don't you think this one does?
Liz: It does when you look at it two dimensionally, but not
as soon as you get perspective on it. For the third
dimension, that's where I feel they don't work, but
this is=
Michael: Well maybe we're not doing them justice in just having
them out on a white tissue on a table; I mean one
needs to see these like one would see Mary's batiks=
Liz: Exactly.
Michael: Framed on a wall.
Liz: That you need to see them from the angle at which they
were painted 'cause they don't work at any other
angle=
David: But they claim to be pebbles, after all,=
Liz: =Which=
David: =Don't they. They are=
Liz: =Which is why I think that it's inappropriate to paint
animals on pebbles. That's what I very much feel, but
I wonder what you think. I'm sorry - I've led the
discussion this time because I really have this thing
Mary: Would you like to hear Jane's views?
Liz: Yes.

In example 18, Roger introduces the topic of whether the applicant, a chair-maker, ought to have used a chisel or a saw in a 'cutting back'. Clearly, an important aspect of the object, according to Roger, is the way in which off-cuts of wood have been removed. During his discussion of this, he mentions the fact that the applicant is self-taught, with the apparent intention of relying on this fact to substantiate the claim that the applicant does not 'know how to make it better'. By introducing the 'chisel/saw' topic, Roger has been able to hint at a negative evaluation of the object. Michael and Jane, however, seize on the ancillary, 'self-taught' comment, and take it up as worthy of discussion. First, Michael expands on the applicant's self-taught status by deducing that the applicant could benefit from talking to Michael. Jane then wonders how he has acquired the knowledge which he does possess. Finally, Liz takes up Jane's question by replying that: 'Well he doesn't say'. In this way, Michael, Jane and Liz each contribute to the stabilising of the ancillary, 'self-taught' comment initially produced by Roger. Evidence of the self-taught status of the applicant in the craft items produced has become a salient feature of the object. At this point, however, the stabilised ancillary aspect then becomes a pivot. For having said that, in respect of the ancillary matter about sources of teaching, the applicant 'doesn't say', Liz goes on to introduce something which the applicant does say: 'All he does say is he's been doing it for two years'. Having managed the switch from the 'chisel/saw' topic to a new 'two years' topic, Liz is then able to
attach a positive evaluation to this revised version of the object as a piece of work created by someone with only two years' experience: 'and I think for two years his achievement is considerable'.

In example 19, the applicant's work takes the form of cats painted on pebbles. Michael begins by introducing the topic of the relationship between the painting and the pebble. Liz responds initially by abruptly 'jumping' topic to the subject matter of the painting, saying that animals painted on pebbles 'don't...justify themselves'. This conversational jump is, however, unaided by any of the usual jump markers. The first consequence of this is that Liz's comments display a large degree of hesitancy: repetition of 'I', reformulation after 'I always feel that it's not', the modification via 'I mean, insofar as you can' and the pauses after 'are so peculiar'. The second consequence is that Mary tries to make her comments somehow relevant to Michael's claim. The pebbles are, she says, 'the best I've seen'. But the real connection between the topic of animal subject matter and Michael's topic of the relationship to the pebble is only introduced in Liz's next turns. The reason she thinks that animals are an unsuitable subject matter for pebble painting is that she has been wondering what would be a good subject matter, and has decided that only abstract subject matters would 'relate to the shape'. Here, something like a topic flow has taken place, but the usual topic flow order of 'ancillary comment stabilisation - pivot - switch to new topic' has been disrupted.
Chapter 5

As a means of negotiating the nature of the object, however, Liz's awkward employment of an inverted topic flow form misfires. For the penalty which Liz pays for her awkwardness is that David is able to reintroduce his original topic, by asking 'don't you think that this one does' in reference to a particular painting's relationship to its pebble. This time, however, Liz has learned her topic flow lesson. She agrees with David that the painting/pebble relationship is acceptable - but only 'when you look at it two dimensionally'. Michael then picks up this issue of two versus three dimensionality, and comments that the pebbles ought to be treated more as two-dimensional works such as Mary's batiks. Here Liz makes an ancillary comment on the 'painting/pebble' relationship topic - she introduces the question of how the relationship is perceived when looked at two-dimensionally and how it is perceived when looked at three-dimensionally. Michael then stabilises this ancillary issue by means of his 'batik' comment.

Liz is, however, now able to seize on this stabilised ancillary issue of dimensionality and pivot the conversational topic to the notion of 'angle of viewing'. The pebbles, it turns out, 'don't work at any other angle'. It is this which explains, Liz concludes, why the applicant's work is inappropriate. It is interesting to note that Liz's topic change clumsiness has left David feeling dissatisfied ('But they claim to be pebbles, after all'). Liz therefore makes sure of her conversational achievement by moving into an other-directed conversation restart - 'I wonder what you think' - which allows Mary to introduce a fresh topic.
In example 18, then, Liz is able to steer the conversational topic in such a way that the kinds of problem outlined by Roger do not come to seem important aspects of the object. Instead, Roger's complaints are made to flow naturally into talk of an aspect of the object about which Liz can express a positive evaluation. In example 19, on the other hand, Liz's initial attempt to characterise the nature of the object as not involving the painting/pebble relationship fails. Her breach of topic change procedures leave her facing the original topic. Her second attempt to re-negotiate the character of the object is more successful, however. By locating a stabilised ancillary issue: 'two- versus three- dimensionality', she is able to switch the conversation to talk of 'angles of view'. This allows her to establish that the object is such that it has a deficiency in being able to be viewed from only one angle.

The value of topic flow is, then, that if the appropriate procedures are carried out, conversants are able, in an inconspicuous, stepwise fashion, to change the conversational topic in such a way that a particular depiction of the relevant object is achieved. Where such procedures are flouted, however, such constructive efforts are much more likely to draw criticism.

§4 CONCLUSION

There seems to be clear evidence that individuals do attempt to manipulate conversation topic when discussing an object, with the aim of establishing a particular 'version' of that object as the
relevant version. Sometimes, as in the early examples of section 2, this is a cooperative effort. On other occasions, individuals employ topic management strategies in order either to establish a particular object 'version' or to prevent its discussion. When carried out properly, these manipulative efforts pass, by and large, unchallenged. Moreover, the effect of such constructive efforts is subtle. On some occasions (e.g. example 16), the topic manager merely establishes the importance of an object 'version' to which he or she has already attached an evaluation. On other occasions, (e.g. examples 17 and 20), the topic management is intended to close discussion of aspect of the object to which an evaluation has been attached with which the topic manager disagrees.

This suggests that an analysis of discourse does represent a practical procedure for social scientific study. The social scientist can use discourse analysis to describe the way in which participants establish for themselves the meaning of a particular social action. Each of the above extracts represented an aspect of one or other of the group's decisions about including or excluding an applicant from the Guild. By manipulating the 'appropriate' characterization of the applicant's work, Guild members were able to establish the meaning of each of these social acts: whether one should consider neatness in deciding, whether material used is important and so on.

One test of the extent to which this discourse analytic approach is useful to the social scientist is whether such analyses can
make sense of problematic data. In this context, it is interesting
to go back to the puzzling data with which the present paper began
- the segment of conversation in example 1:

1: Tape 1, *1 p. 2-3
Mary: His wife makes the cushions.
Liz: Oh! Oh dear! =
Mary: =((laughs)) Another problem =(laughs)=
Liz: =((laughs))
David: I met him about two years ago actually, he, em, he's=
Liz: Oh really?= 
David: =quite an interesting chap. He's em=
Liz: =Mmm
David: One of the things he told me he'd done - he'd been -
em, a road manager for a pop group. And=
Liz: =Really
David: I think he'd made so much money out of it that em went
to this Parum=
Liz: John Makepeace School
David: Yes. Yes.
Mary: Really?
Liz: But is, is it?= 
David: =It doesn't look comfortable=
Liz: =That one looks positively dangerous.
David: Yeah.

From a discourse analytic perspective which concentrates on topic,
it is now relatively easy to describe what goes on in this
conversation. Mary introduces the topic of the applicant's wife
having made the cushions. Liz responds by formulating a joking
expression of concern ('Oh Oh dear'). Mary takes up the joke and
laughs. She then offers a reformulation of her original topic,
characterising the issue she introduced as 'another problem', but
signals that this has now become a relatively unimportant issue by
laughing again, as does Liz. This establishes that the fact that
the applicant has not made all of the exhibit is not to become a
salient feature of the object and leaves the floor open for the
possible introduction of a fresh topic. This opening is taken up
by David, who switches the conversation from talk of cushion
making to the fact that he had already met the applicant. This
new-topic introduction bears typical topic change hesitancy markers, including hesitancy exclamations ('Em'), repetitions ('he, em, he') and reformulations (he told me he'd done - he'd been, em'). [17] As David proceeds, Liz begins to insert acknowledgement markers which are intended to signal the closure of the topic ('Oh Really', 'Mm', 'Really'). Finally, Liz rounds of the topic by helping David to sum up his contribution ('John Makepeace School'). This allows her to begin a separate topic: 'But is, is it' (note the hesitancy marking here). David joins in the move to a new topic, introducing the issues of comfort-value, and Liz ends by taking this up.

So the importance of the discourse analytic approach is that it allows for description of what a given act means to participants even in cases where the meaning of that act is, at the outset, obscure. Discourse analyses alert the social scientist to the fact that holding and maintaining viewpoints on what a particular act means is a matter which is more complex than might at first be thought. Thus in a number of the present examples, the view taken on the object under review is revealed as much by the work done to suppress discussion of some aspect of the object as it is by outright evaluations of the object in terms of that aspect. Subtle differences are also signalled by evidence of topic management. Where 'bland' topic change occurs, the analyst can assume that differences will be muted, while evidence that specific topics are emerging (or submerging) as the result of an individual's conversational 'work' signify disagreement, even in cases where such disagreement is, initially, invisible to the naive onlooker.
or to the participants themselves. Indeed the present discourse analysis study provides the analyst with a practical ability to monitor agreements and disagreements of even the most obscure kind. If, for example, conversants agree in 'marking' a topic discussion with laughter tokens, the analyst can take this as a sign that, for the introducer of the laughter tokens, the topic introduced as important by the preceding speaker does not characterise an important aspect of the object.

This shows how the discourse analytic approach helps, even in the case of one tiny fragment of interaction, to generate a clear description of such interaction. Now at the beginning of this chapter, we noted that social science must have two features: it must provide a clear description of participants' social acts - one which reveals the meaning such acts have for those participants, and it must provide explanations of those acts. It now looks as though discourse analysis represents one way in which the first of these can be achieved. This leads to the question of how discourse analysis might help with the second aspect of social scientific study: the provision of explanations.

To start with, it is obvious that the discourse analytic approach relies upon the idea of purposive (if unconscious) manipulation of discourse strategies. It is already implied that individuals' contributions to conversation and debate are purposive not only in the sense that the content of their talk is uttered for a purpose, but that the extended conversational form of those contents may be chosen for that purpose too. Thus discourse analysis already
presupposes that there will be some form of explanation (whether in terms of reasons or causes) which lies behind the conversations (or other discourse) recorded.

In example 1, for example, David clearly wishes to establish that an aspect having negative evaluation potential which might have been seen to be relevant to the object - the question of who made the cushions - is of no importance. What is suggested, furthermore, is that Liz, at any rate, is willing to agree with this 'version' of the object. She is less willing, however, to see the applicant's craft-training background - which might be thought to have a positive evaluative status - as a salient feature of the object, and proceeds to close this topic off. [18] However she does finally agree with David in taking up the issue of whether the chair looks comfortable - an issue that might be expected to have relatively little evaluative impact. What the analyst might expect, then, on the basis of this one small segment, is that David at least is likely to express an approval of the object and that, so far, there is little reason to expect Liz to express opposition. (In fact, at a later stage of the conversation Liz and David agree that the applicant should be selected.)

What is needed now is an assessment of the further contribution which discourse analysis can make to discovering the explanation which lies behind the events described by means of such analysis. It is this assessment which is pursued in the following chapter.
CHAPTER 6

§1 INTRODUCTION

The previous chapter introduced the notion that social science must not only describe the meaning of social activity but must also attempt to provide explanations for such activity. To provide such an explanation is to give a reason or cause for the participant's acting in the way he did. Of course the types of thing or event which are traditionally taken to act as causes or reasons in social science is enormously varied. They span the range from individual psychological attributes to macro-sociological constructs such as ideology. But even supposing the theorist is able to isolate one specific cause or reason in his theorizing, a difficulty remains. For an understanding of these reasons and causes often depends on understanding the meaning which they have for the participants whose activities are being explained. If an individual acts as he does because he belongs to a particular group, for example, this is often because he has a particular conception of what such group-membership means.

This suggests that if the social scientist is to produce a genuinely social analysis, he may have to cope with the problem of meaning not only at the level of description but at the level of explanation too. [1] For example, suppose the social scientist wishes to assess a specific explanation: that subjects' acts vary in accordance with the adoption of different social roles at different times. [2] In part, this explanation will depend on how
the participant himself thinks of the role he is adopting on a given occasion. This, in turn, will depend on the meaning which this role has for him and his fellow participants. It follows that in order to produce a proper explanation, the social scientist would, in such a case, have to explain the meaning which this role possessed for those participants.

It is in this respect that the discourse analytic technique introduced in the previous chapter can contribute to social scientific explanation. Discourse analyses allow the social scientist to identify competing accounts among participants of the meaning of some aspect of their social world and to identify the strategies by which these accounts are introduced and defended by those participants. In the case of the social role theorist, then, the discourse analyst would be able to demonstrate what a particular role means to the relevant participants by isolating competing accounts of that role and showing how these accounts are challenged and maintained.

To see how such analyses function within the explanatory framework, the following sections look at how a group of participants' activities can be understood in terms of the notion of 'self-presentation'. [3] The idea behind such explanation is that many actions can be seen as attempts, by the actors involved, to present themselves and their beliefs, opinions and the like in a positive light by emphasising the rationality and clarity of those beliefs and opinions. In the examples which follow, the emphasis will be on how competing presentational accounts can be
identified by discourse analytic means. (For the sake of brevity, we will ignore the discourse strategies, of the sort discussed in the previous chapter, which participants employ in order to maintain or challenge these accounts.)

62 SCIENTISTS' CONFERENCE TALK

In this section we examine the way a particular group of scientists attending a major psychology conference deal with the conception of scientific models. In the discussions generated by these scientists, tensions can be seen to arise among the various accounts given the nature of such models. First, the relationship between such models and the experimental evidence with which they are associated becomes a topic for debate among the gathered scientists. Some of the discussants depict major scientific models as deriving in a constructive manner from the testing of small-scale hypotheses. This is termed the 'bottom-up' approach. Others give an account in which the scientific endeavour begins with an a priori construction of large-scale models from which small-scale hypotheses are then derived and tested. This is termed the 'top-down' approach. In the ensuing discussions among the conference delegates this distinction becomes a major source of disagreement.

The analytic portion of this section is concerned with the way variation in accounts is created. In particular, it concentrates on the existence of a particular combination of patterns of discourse within the top-down/bottom-up debate: the 'empiricist repertoire' and the 'contingent repertoire'. 'Empiricist
repertoire' is a term used by Gilbert and Mulkay to refer to a pattern of discourse in which an account of actions and beliefs is organized in such a way that its author's interpretative activities are seen as irrelevant. The account is presented as deriving neutrally from the facts in such a way that the author's personal involvement becomes less visible. [5] 'Contingent repertoire' is a term used by Gilbert and Mulkay to refer to a pattern of discourse in which the accounts of others can be depicted as being influenced by factors outside the empirical realm such as personal bias, political motivation and so on. Accounts of scientists' actions and beliefs which employ this repertoire depict them as dependent upon contingent circumstances arising within these personal and social dimensions. [6]

In the analytic section, then, an attempt is made to discover whether both of these repertoires appear in the scientists' discussion of the bottom-up and top-down approaches. These repertoires, the prediction is, will be found to be variably drawn upon for different interpretative purposes. In particular, an examination is made of the extent to which scientists give accounts of their own actions and beliefs using the empiricist repertoire while accounting for the errors of opponents through the use of the contingent repertoire. [7] (It should be noted that terms such as 'error', 'Correct' and so on will be used solely to reflect the views of the current speaker throughout the rest of this text.) In this 'accounting for error', not only is a speaker's own position taken to be synonymous with the correct view as derived from the facts, but the view of opponents is
characterized as mistaken - a characterization which relies upon explaining the opponents's actions and beliefs in contingent terms.

In this way, discourse analyses of the data will be seen to contribute to an explanation of what went on at the conference in terms of 'self-presentational' strategies.

There is one further complexity with the present data when compared with that collected by Gilbert and Mulkay. They used interviews with individual scientist in which the scientist was communicating only with the researcher and, in addition, was assured anonymity. In this situation, scientists can produce highly negative descriptions of others' behaviour without fear of comeback. In contrast, conferences are highly public arenas where talk is designed to be heard by scientist peers. This faces scientists who account for error with a problem. For if they produce a contingent version of other scientist' behaviour - citing incompetence or bias, say - this could be heard as an accusation and, as a consequence, perhaps involve the speaker in heated dispute. At the very least, such dispute would interrupt the orderly production of the speaker's version. To examine this phenomenon, the analysis will document both the existence of error accounting and the particular forms in which it occurs.

The data were generated at a psychology conference concerned with 'fundamental theoretical issues'. The participants were mainly invited according to the convener's criteria of having made an
important contribution to the development of psychology. The conference was split roughly in half into formal presentations and discussion periods, all of which took place in a single lecture hall with no parallel sessions. All the sessions were tape-recorded by the organizers and a verbatim transcript of the discussions was produced (150,000 words). All names were changed to ensure the anonymity of the participants. The entire transcript was read and all accounts of the nature of scientific models were extracted and coded in terms of their relevance to the major themes under investigation. After this initial coding, each extract was then compared with the others in an attempt to discern similarities and differences in the nature of the explanations employed. There was, of course, no attempt to draw conclusions about whether any of the aspects of models under discussion have any 'real' or 'objective' importance, nor was there any attempt to defend a particular viewpoint. The purpose of this section and the next is solely to examine the organization of the discourse involved in the giving of accounts of such aspects and viewpoints. This methodological procedure is described in more detail by Potter, Mulkay and Wetherell. [8]

In the analysis that follows, examples from a wide variety of instances of scientists are seen to give an account of their own viewpoint through means of the empiricist repertoire. They are also seen to provide versions of their opponents' actions and beliefs through the use of the contingent repertoire. This latter interpretative task is undertaken by the scientists in question in order to explain why their opponents fall into error given the
claim, made by each scientist in turn, that the correct view can be derived from the empirical facts. [9]

**The Top-Down Proponents Speak**

We can begin by examining examples in which proponents of the top-down approach explain and defend their conception of how scientific models should be related to empirical data.

The following excerpt occurs at the beginning of an extended account in which Rees [10] describes the correct view of scientific models as being one which favours the top-down approach over the bottom up.

1. It strikes me that we are not going to get very far in a bottom-up approach. We can get so far along the line and then we hit logical difficulties. (163)

The speaker's rejection of the bottom-up approach is depicted as arising out of the 'logical difficulties' inherent in such a view. This theme is returned to by Reese at a later point in the same speech:

2. It seems to me that there are two bottom-up approaches ... one lateral coalescence ... And two would be expanding a single model upwards ... the lateral interbreeding model seems to me to be not possible ... because each miniature model has its associated ... notion of what is an appropriate fact, notion of what to measure, notion of type of empirical data ... they have associated metatheoretical features which don't allow this interbreeding. (163)

Here two types of bottom-up approach are offered for inspection - a 'lateral coalescence' model and a model which 'expands a single (small-scale) model upwards'. The adoption of the former is
depicted as being 'not possible' because of the objective features of small-scale models: they are 'associated' with particular types of data and measurement in such a way that they themselves 'do not allow interbreeding'. The thrust of this account then is that the rejection of the bottom-up approach to scientific models is forced upon the speaker by empirical and logical constraints. This is, of course, a classic empiricist mode of argument.

As for the latter 'upwards expansion' type of bottom-up approach, its objective difficulties can be seen, says Reese, by looking at a specific example of bottom-up theorizing: the development of action and expectancy theory.

3. (in such a model) ... just before you get to action tendency, there is another arrow that comes in and says 'social pressures'. Now this great big variable has taken over and it has in a sense invalidated the original model .... (that is) ... the generalizable instance of the problem of trying to move upwards. (164-65)

It is a generalizable feature of the latter kind of bottom-up approach, says Reese, that in the end, some particular addition in the upward movement 'invalidates' the original model (thereby undermining the overall bottom-up approach). Once again, the top-down proponent's rejection of the bottom-up approach to scientific models is portrayed, in empiricist terms, as being a reaction imposed upon the speaker by the nature of the models themselves.

Objective problems associated with the bottom-up approach also appear in the accounts given by two other conference speakers, Thomas and Charter:

4. (in psychology) ... anything you can think up to get that subject to do, by definition the subject, because he is a human being like you, can do ...
There is therefore an infinite possibility of data gathering. (Thomas, 175)

5. ... we might save a lot of time by trying to work out ways of deciding between models at the conceptual framework level. And some way of deciding between them through reason, I suppose, not through testing them out in the sense of going on and on trying out mini-models because that takes years, obviously. You have got to find some way of deciding between them intellectually. (Chartier, 168)

These accounts depict the bottom-up approaches suffering from obvious yet objectively determinable shortcomings: data-gathering possibilities are infinite, and so take up too much time. The choice between top-down and bottom-up approaches must therefore, Thomas and Carter suggest, be decided in favour of the former, given its practical, time-saving aspects. It is, says Carter, 'obviously' in the nature of the bottom-up approach that it requires a wasteful testing effort which 'goes on and on' and which could 'take years'. The top-down approach on the other hand, he says, allows for decisions to be based on 'reason' and by 'intellectual' methods.

In each of these extracts, the proponents of the top-down approach can be seen to employ the empiricist repertoire in arguing that the bottom-up approach should be ruled out as an appropriate conception of the link between scientific models and empirical findings. Bottom-up approaches are characterized as doomed to inconsistency and as wasteful of time and effort because of the infinite potential in data gathering. The use by the top-down proponents of the empiricist repertoire allows them to minimize their own constructive role in the argument for the top-down approach. They are able, instead, to depict it as an approach...
which follows naturally from the nature of the facts, independently of any interpretation carried out by its proponents.

The Top-Down Proponents Account for Errors

It was pointed out earlier that the empiricist repertoire is not the only interpretative framework which scientists use when giving accounts of actions and beliefs. They also draw upon a contingent repertoire. One of the occasions on which they are most likely to use this alternative repertoire is when they are describing scientific errors. [11] As was noted, the use of the contingent repertoire for such accounts resolves the interpretative problem of how errors can be made when belief is taken to be a direct consequence of experimental findings. For in the contingent repertoire the direct connection between experimental findings and belief is disrupted.

In this section, such accounting for error is shown to form part of the discourse of those who argue that scientific models should be seen as top-down phenomena. The following excerpts are taken from an extended speech given by Thomas in defence of the top-down approach to scientific models against the bottom-up proponent's conception of experimentation without top-down guidance.

6. If you look back through the journals ... you will find them absolutely full of fascinating experimental work which has absolutely no effect whatsoever on anything. People did it for a few years and then they got bored and they stopped doing it. And then they did something else and they did that for a few years and got bored with it.

(175)

7. ... unless I devoted the time to analysing the task to see what I hoped to learn from it about mental
function in general I would be wasting my time. It seems to me that at the moment many experimenters do not take this important step. They rush straight into the laboratory.

8. ...it is accepted by the journals that any task will do. If I went out and did an experiment on Pooh sticks you know, I strongly suspect I could get it published.

9. ... it seems to me that we can look at, for example, Watson's 4 Card task ... a large amount of public money has been spent on financing this research. At least fifty or sixty papers have been published and there has been at least one international congress look at this particular task

In these passages Thomas explains why people still cling to an approach to scientific models which he has earlier characterized as possessing determinable objective flaws. They are, according to extract 7, unwilling to think about whether their experiments have any genuine value - they prefer to 'rush straight to the laboratory'. They are pictured, in Extract 6, as being willing to continue shallow, ineffectual bottom-up experimentation because they tend to be people who are easily bored. And such erroneous activities, according to Extracts 8 and 9, are encouraged by the lack of diligence of the editors of 'the journals' and by the fact that 'a large amount of public money' is available as finance. In this way, by appeal to contingent personal and sociological factors, Thomas is able to give an account of why his opponents choose their erroneous course.

There is another notable feature of these extracts. Despite the fact that they are produced in the course of disagreements with other contributors (who in the course of the conference have espoused bottom-up positions), it is not these contributors who
are identified for criticism. In fact, no specific scientists are
identified. Instead reference is made using such vague descriptors
as 'people' (Extract 6) and 'many experimenters' (Extract 7) or,
alternatively, direct reference to scientists is avoided
altogether (Extracts 8 and 9).

On other occasions, top-down proponents account for the errors of
adherents to the bottom-up approach using other facets of the
contingent repertoire.

10. I still want to ... develop broad conceptual
frameworks ... why aren't they there? A) Because
it's jolly difficult ... and B) because they have
certain logical and practical problems. This leads
some of us to say 'well', therefore, 'let's not
bother'.
(Reese, 250)

11. Now I think that there are many possible
explanations of this. (lack of interest in a top-
don oriented conception). One is in fact that the
theory is excessively complicated and I think in
two volumes this perhaps puts people off. I think
that's one explanation.
(Harvey, 139)

In these passages, the top-down proponents account for their
opponents' acceptance of the erroneous bottom-up approach by
drawing a particular picture of them. They are depicted here as
people who prefer to avoid complex issues. When faced with
difficulties or problems they prefer, according to Extract 10, to
say 'well ... let's not bother'. Similarly, in Extract 11, they
are pictured as being 'put off' by complicated theories and by
books in two volumes.

Just as with Extracts 6 - 9, the error accounts do not focus on
specified individuals. Extract 10 utilizes the very oblique
construction 'some of us', which leaves vague the exact reference
of 'us'. It could mean the speakers in the debate, the entire conference membership or everyone in the discipline. Extract 11, like Extract 6, draws on the general, unspecified term 'people' when talking of those who find the position too complicated.

By employing the contingent repertoire in this way, the top-down proponents formulate particular versions of their opponents' actions and beliefs in order to account for their acceptance of an approach depicted as being determinably erroneous. The bottom-up adherents are characterized as easily bored and as unwilling properly to analyse the value of their experiments. This is compounded by their preference for avoiding difficult subject areas. And the effects of these personal weaknesses are made worse by contingent sociological factors such as the distorting effect of public finance and the fact that editors of the journals lack diligence.

These kinds of accounts are potentially troublesome. The characteristics they attribute to scientists and and scientific collectivities are extremely blameworthy. [12] And in this situation failure to respond immediately to what amount to accusations of deviance from acceptable scientific conduct might well be taken as a tacit admission of guilt. [13] The problem for the speakers is that making such an immediate response may well prevent them completing their general point; they risk getting bogged down in a detailed dispute about the conduct of individuals. Given this, the highly non-specific way in which the error accounts are directed can be seen as a deliberate design
feature. [14] The accounts are constructed in such a way as to be very difficult to dispute, and to minimize the overt blaming of colleagues.

In general, in debating the nature of scientific models, just as the proponents of the top-down approach rely on the empiricist repertoire in giving accounts of their own actions and beliefs, so they can be seen to rely on the references to personal characteristics and contingent factors in the social world which constitute the contingent repertoire, when they characterize the errors of their opponents. However, this repertoire is skilfully employed to minimize the disruption caused by the potentially negative attributions. Such selective employment of repertoires in the achievement of interpretative goals is not, however, restricted to the proponents of the top-down method, as the next section shows.

83 THE BOTTOM-UP PROPONENTS SPEAK

The top-down proponents claim that their approach allows for a rational a priori decision procedure in the choice of scientific models. One concern of the bottom-up proponent is to show that this feature does not characterize the correct notion of scientific models. For example, the version of events offered by Norton here depicts the bottom-up approach - the approach which refrains from a priori speculation about criteria as models - as being the one operative in fields of psychology where 'rapid advances' are being made.
12. ... there are fields of psychology in which advances are being made rather rapidly, and they are fields in which people do not actually ask themselves questions about what ought to be the criteria of models. (169)

This version is developed by Norton later in the same speech - in which he talks of learning theory, one particular area representative of the bottom-up approach, in the following terms.

13. ... there has been a steady stream of progress in the field of learning since approximately 1897 ... We know a great deal ... of factual information ... and we have a number of bodies of data where models are relatively successful. (169)

Here the bottom-up approach is represented as involved in a 'steady stream of progress'. Norton pictures this as dealing with quantities of data derived from information grounded in fact. He explicitly draws the implication that this factual process allows for the production of models which have been proven to be fruitful. So the success of the bottom-up approach, in Norton's account, lies in its ability to interact with the world of facts through its ability to generate useful bodies of data which lead to advances and progress in the field.

This picture of the bottom-up approach as one whose merits lie in its relationship to the factual world is further stressed by Norton at a later point in the same discussion. At that point, a dispute is under way as to the relative merits of two rival bottom-up approaches, during which Norton states:

14. ... the point is that in each case the rival theories are being specified in more details and (are) making more specific predictions. (194)

The picture drawn in this account is of a bottom-up approach which allows for a continuously tightening specification of details and
which generates predictions which become more and more specific. So once again the choice of the bottom-up approach is pictured as resting solely upon that approach's objective merits which, this time, include the ability to generate predictive theories.

The reference which Norton makes to objective features such as 'progress', 'rapid advance', 'factual information' and so on is echoed in the accounts of the bottom-up approach supplied by Coleman. In describing one example of the bottom-up approach, for example, Coleman says:

15. I don't think you have to ask these higher order questions at all. I mean one very, one extremely useful set of experiments that I think - a body of data that has been used ever since it was produced - is the work of Furster and Skinner on schedules of reinforcement. It has been extremely interesting within the field to have this body of information and very valuable. (178)

the bottom-up approach is represented in this account as an approach which generates important experiments and data. Thus Coleman, like Norton, bases his account of the importance of the bottom-up approach firmly on the notion of the objective features possessed by that approach.

In their emphasis on the data-collection, hypothesis-formation, prediction-making, scientific progress and so on, the proponents of the bottom-up approach can be seen to be offering accounts of their own preferred methods of explanation which parallel those of the top-down proponents. Those who support the bottom-up approach attempt to characterize that support as a matter of falling in line with the dictates of empirical reality. What is more, this parallel is continued in the error accounts which the bottom-up
proponents produce through the employment of the contingent repertoire, as the next section will demonstrate.

The Bottom-Up Proponents Account for Errors

One position regularly identified as bottom-up by its proponents at the conference was behaviourism. [15] In the following extracts, a particular version of the top-down proponents' actions and beliefs is depicted as a consequence of contingent factors.

16. (Top-down proponents) push me into this role of being a kind of psychological black, and it seems to me, almost the victim of prejudice. People will come and criticise without taking the trouble to find out what is actually happening; there is no sympathy for the point of view at all. At best one can expect to be patronized. (Coleman, 10)

17. (When Darwin's book was published) ... There was this fear around ... to Bishop Wilberforce, Darwin's book The Descent of Man almost ... might represent ... the new fall of man. And I think that we are having a rerun of these fears. And that the attitude towards behaviourism is precisely the same kind of attitude that was evoked by Darwinism. The notion that ... it is terribly dangerous, it's subversive, if we are to believe in it then it will ... somehow dehumanize man, lower the status of man. (Coleman, 309)

In Coleman's version of events (Extract 16) bottom-up proponents are the victims of prejudice. Their conception of scientific models is rejected without any attempt at rational assessment. Similarly, in Extract 17, Coleman depicts the opponents of the bottom-up approach to scientific models as being beset by irrational fears. Just as there was an attempt to characterize Darwin as dehumanizing man, [16] Coleman's account runs, so there is a contemporary attitude that one particular example of the
bottom-up approach - behaviourism - is similarly dangerous, subversive and dehumanizing.

Just as in earlier cases, these error accounts are applied in a highly unspecific fashion. In Extract 16 the source of criticism without understanding is identified only by the epithet 'people'. While in Extract 17 actors are not even identified in the vaguest terms. Instead, the account formulates only an 'attitude'. Again, the failure to attribute faults to specific conference participants, or even to any specific scientists, cuts down the probability of potentially damaging disputes about the substance of the accusations.

Behaviourists (or at least scientists adopting this label on this occasion) are not the only ones to criticize opponents of bottom-up procedures. In the following extract, Steele responds to an account of the bottom-up approach given by Squire which depicts it as essentially reducing to triviality (square brackets indicate material omitted).

18. I don't think I have ever had time to sit down and rationalise in any great detail [] my actual practice as a scientist and attempt to create some doctrine out of it. If I had, however, I [] would never allow that doctrine in any way to dictate the course of my work. [] To justify your own doctrine you appealed to the fact that it was a good representation of the actual practice of scientists who have deemed it to be successful. Now I would claim the right to use that same basic criterion and [] say [] that in my opinion you cannot simply dismiss a whole area of psychology for doctrinaire reasons [] until you can show that your particular doctrine has [] proved more fertile and productive scientifically, which is not clear to me yet. (209)
Having denied that he himself is constrained by any narrow system of thought, Steele sets out a distinction between those who do and those who do not subscribe to doctrines. His dismissal of the top-down procedure does not constitute a doctrine, he says, but his top-down opponent does subscribe to a doctrine - a doctrine which causes him to attack approaches which are not at some stage top-down influenced - even though this doctrine has yet to be proved more scientifically productive. In this way, the top-down proponent's choice of model is depicted as a results, not of a factual analysis based on empirical ground, but of the top-down proponent's overly rigid set of beliefs.

The error account in Extract 18 is different from those examined thus far. Instead of Steele directing his account at unspecified scientists or groups of scientists, it is explicitly directed at an individual: Squire. One likely suggestion about the motivation [17] for avoiding such focussed accounts is that they would lead to disruptive conflict. Thus what happens after Extract 18 provides an interesting test of this suggestion.

The first thing to note is that Squire does indeed respond to Steele, just as might be expected. However, the detail of this response is very interesting.

19. I think it isn't right that there, you can as it were, proceed in a kind of, how can I put it, dramaturgical naivety in the presentation that you have just done. [Steele laughs] I mean, if you are doing anything in the practical world of experimenting or whatever you are not doing that innocently.  

(210)
In this extract Squire responds to Steele with another contingent account. He addresses the rhetoric of Steele's presentation, suggesting that it is less innocent and naive than it might sound. This initiates a lengthy and somewhat heated dispute which is ended by the chairperson after the following exchange.

20. Steele: We may be doing all sorts of very untidy and inelegant things, but I am still quite sure that given any strike we could explain it a lot better at the moment than anyone else. And I would like to see anybody do better than us.

Squire: Now, well you see there we are arguing with each other, I think I could explain it a damn sight better than you could. [laughter] But, err ...

Chairperson: Ladies and Gentlemen ...

(210)

The fact that the exchange ends with the participants simply and boldly claiming they could better explain a particular phenomenon than the other shows the potential dangers of using specific error accounts in this kind of situation. The participants risk becoming involved in a distracting and potentially heated digression.

The Organization of Model Talk

The prediction made at the start of §2 was that, in their discussion of models, scientists would generate a variety of accounts of actions and beliefs and that, furthermore, such variability in accounts would represent a resource drawn upon the scientists involved in achieving their interpretative aims. This prediction has been shown to be substantially correct. Both top-down and bottom-up proponents selectively draw from a wide range of potential accounts in making sense of their own actions and beliefs and those of others.
The findings of the current analysis also mesh, in their broad outline and fine detail, with the conclusions of Gilbert and Mulkay concerning error accounting. The motivation for this kind of accounting lies in the problem of erroneous beliefs (the beliefs of other scientists) which scientists are presented with when they draw upon the empiricist repertoire. If beliefs typically arise straightforwardly from experimental evidence, why are some scientists mistaken? The difficulty is resolved on these occasions by the deployment of the contingent repertoire. This introduces distorting factors like bias, incompetence and institutional pressures to account for why scientists have got it all wrong. In the analysis above a variety of factors of this kind can be seen to be drawn on in accounting for the prevalence of erroneous scientific models: incompetence, careless journal editing, ill-judged financial support, an inability to cope with complex models and issues, prejudice and doctrinaire support of a position. It would be of interest in future work to discover the extent to which the identified factors represent systematic elements relied upon in the design of such accounts, and the extent to which such elements can be determinately associated with specific aspects of the interactional work undertaken.

It was noted earlier that there is an important difference between the present data and that of Gilbert and Mulkay. Their analysis identified instances of error accounting in transcripts of interviews which are essentially private; while the present extracts come from a highly public conference situation where talk is intended to be overheard by peers. The danger in this situation
is that error accounting may embroil the speaker in disputes which may become heated and may distract from the speaker's general point. In this situation, then, it was suggested that two constraints may be operating simultaneously. [19] On the one hand, scientists need to account for erroneous belief, and thus maintain the coherence of their own positions. On the other, there is a motivation to avoid potentially disruptive blamings.

The pattern of error accounts examined strongly supports the suggestion that such multiple constraints are at work. In eight out of nine of the instances discussed above the source of the personal/social characteristics used to explain errors is highly inexplicit. For instance, Reese (Extract 10) pictures an unspecified opponent saying, of some difficult task, 'let's not bother', rather than accusing an identified individual of what is outright fecklessness. Or take Coleman's error account (Extract 17), in which he describes anti-behaviourists as having an 'attitude' or set of 'fears', but refrains from explicitly accusing named individuals of stupidity or lack of rationality. In general, specific groups or individuals are not named; instead broad, vague descriptors such as 'people' or 'many experimenters' are deployed.

This pattern is in marked contrast to those accounts discussed in Gilbert and Mulkay's analysis. In twelve out of the sixteen extracts reproduced, one or more specific, identified individuals or groups are named. [20] The following is typical of their data: 'Gowan in fact never understood that hypothesis. This was very,
very obvious to anyone who talked to him. He had such a dislike of it that he never bothered to think through what the consequences would be.' [21] The criticism here is strong and specific. As such, the potential for contesting it is much greater than, say: 'many experimentalists never understood that hypothesis', which is the type of formulation found in the conference data.

What the evidence points to, then, is that the contingent repertoire itself represents a variable resource, allowing individuals to 'fine tune' their employment of it for the specific occasion at hand. Part of the skill participants display when they produce error accounts, for example, lies in their ability to produce direct or indirect formulations appropriate to their current situation. Thus, although the Gilbert and Mulkay extract could be directly contested in a number of ways - for instance, by specifying information about Gowan which suggests that he did understand the hypothesis - a much less direct formulation of the type seen in the conference data would be much more difficult to contest. Consequently, indirect formulations are the ideal choice for a scientist wishing to provide error accounts in the course of actual discussion with the individual or individuals to whom the error might well be taken to be attributable. [22]

The relative difficulty of contesting indirect error accounts can be seen if we consider the ways indirect formulations such as 'many experimentalists never understood that hypothesis' might be countered. One response would be to specify a particular individual, or set of individuals, who do 'understand that
hypothesis'. But this could simply be dismissed as an appeal to individuals outwith the group of 'many experimentalists' referred to. An alternative response would be to request that the indirect error account be made more specific, with the intention of rebutting the account in its more determinate form. However, there are two potential problems with such a manoeuvre. First, it could easily sidetrack the discussion away from the issues currently of interest - a point which may be particularly relevant when there is an audience present who may not be much concerned with the detailed rights and wrongs of error accounts. Secondly, such a request for specific information is likely to cue the previous speaker that some sort of rebuttal is imminent, and thus allow him or her to modify his or her discourse accordingly, perhaps softening its blaming potential. Atkinson and Drew have shown the prevalence of this phenomenon in court-room cross-examinations. [23] Thus it seems likely that an important motivation of indirect error accounting is to avoid problems of this kind.

We saw above that, although rare, specific and directed error accounts do sometimes occur. In Extract 18 just such an account is deployed in Steele's discourse. But what happens next provides further support for the present analytic claims, because the speakers (Steele and Squire) do indeed become embroiled in a progressively more heated dispute, speaking increasingly loudly and cutting each other off before they have completed turns. This dispute is terminated by the intervention of the chairperson at a point of apparently total impasse. This illustrates well the
potential dangers of using specific and directed accounts of error in this context.

At this stage, two broader points seem important. First, how should these data about the organization of a psychologists' conference be taken? One approach would be to suggest that these analyses elucidate the sort of talk that takes place in the context of conferences. However, the suggestion here is rather stronger than this; it is that the 'softening the blame' feature of error accounting is in fact constitutive, in part, of the social setting we know as a scientific conference. that is, instead of the context 'scientific conference' being an exogenous framework for certain kinds of interaction which, endogenously, generate the structured and recognizable nature of this context. This is not to say that there will be no features of the 'scientific conference context' which are independent of forms of talk; the coming together of groups of people, for example, seems to be one such independent issue. But such issues are, typically, not sufficient to explain the organization of discourse which occurs at conferences. [24]

Second, this analysis has illustrated the potential of combining discourse and conversation analytic work. Research on the principles at work in ordinary conversations can throw light on the organization of scientific discourse. One of the important aspects of the conversation analytic perspective is the potential it offers in transferring an understanding of the detailed
elements of language use from the sphere of everyday conversation to other, more institutionalized, settings.

§ 4 CONCLUSION

We began this chapter by pointing out that in his work the scientist must provide explanation as well as description. As far as the social scientist is concerned, however, there is a problem with this requirement which stems from the terms in which the explanation is offered - the reasons or causes which are said to explain the actions observed. In many cases, the participants whose acts are to be explained have their own understanding of the matters which are said, by the social analyst, to behave as reasons or causes for the participants' acts. As an example of this, we looked at data which might be explained in terms of 'self-presentation' strategies. It was seen that the participants studied did present themselves and others in terms of the 'rationality' of their beliefs and actions. But they also ascribed a particular meaning to the terms, such as 'reductionist', 'empirical', 'research', through which these self-presentations were achieved. Taken in one way, 'empirical research' was seen to exemplify a rational approach to scientific study. Taken in another, it was seen to exemplify an irrational perspective.

What this implies is that to offer an explanatory theory may demand a prior understanding of how the terms of that explanation are understood by the participants involved. Wherever this is the case, discourse analysis can contribute, not only to establishing
a clear description of what went on in a particular social setting, but to an explanation of what went on too.

**Attitude Theory**

In this respect, discourse analysis can be favourably compared with more traditional approaches to social scientific study such as attitude theory. Many studies in social psychology represent social processes as the outcome of an interaction between objective causes and societal effects. Human behaviour is regarded as an effect produced by phenomena which are themselves accessible to the analyst only in terms of their physical, causal properties. Perhaps the best known study area within this tradition is the field of attitude research. The basic model employed by the attitude theorist [25] is one in which a mental state - the attitude - causes the subject to behave (or be pre-disposed to behave) in a specific way in response to a given stimulus. The mechanism visualised is one in which an object is picked out from the environment and given a positive or negative evaluation which then affects the subject's behaviour with regard to that object. A central assumption is that the assessment of subjects' attitudes, by means of questionnaire, survey and so on, allows for the explanation and prediction of behaviour associated with the attitudinal object.

There have been a number of criticisms of attitude theory in respect of its predictive success [26] to which attitude theorists have responded [27]. However the preceding arguments demonstrate that there there is a basic problem inherent within the
explanatory framework of the theory itself [28] - the theory's basic idea of a simple attitudinal object is unsupportable. Individuals typically respond to objects in terms of how they interpret them. And as attitude theorists have themselves come to accept [29], the way an individual interprets an object depends on which aspects of the object the individual takes to be salient or relevant to the current situation. This allows the subject to construct for him/herself the nature of the attitudinal object which he or she is trying to evaluate. However if subjects do in this way construct the nature of object which is the focus of the attitudes under investigation, this means that there can never be an unproblematically neutral description of the relevant attitudinal object. Indeed the whole distinction between attitudinal object and evaluation of that object becomes blurred [30]. It is in precisely these cases that the sort of analyses developed in Chapter 5 will prove profitable.

Part of the difficulty for attitude theory lies in the notion of salience. According to the theory of reasoned action [31], for example, an attitude can be regarded as the causal product of a set of evaluative beliefs. Each evaluative belief in the set comprises an expectancy as to whether such and such an action will produce such and such an effect and an evaluation of that effect. According to Eiser [32], however, there is a problem with taking attitudes to be causal products of evaluative belief sets. This problem stems from the notion of salience of beliefs. The theory of reasoned action assumes that there will be a set of beliefs over whose salience there is broad consensus. But Eiser argues
that, in practice, individuals exhibit a wide degree of difference in regard to the beliefs items which they regard as personally salient. The conclusion to be drawn from this is that the question of which belief items are salient, in the study of a given attitude, admits of a much wider degree of variability than the theory of reasoned action allows for. In consequence, there seems no reason to accept as a universal rule the idea that salient beliefs causally determine attitudes; in many cases, it makes as much sense to suppose that an individual’s attitude is causally determining his or perception of the salience of a given belief item.

This presents the social analyst with a challenge. The concept of attitude can no longer be explicated by appeal to the relationship between antecedent salient belief and consequential attitude. The issue of which beliefs are salient in respect of a given attitude is one in which there is a wide degree of variability among individual interpretations of such salience. The question is, then, one of how this interpretative variability intrinsic to attitudinal phenomena is to be understood. The answer offered here is that such interpretative variability can best be dealt with through the adoption of the discourse analytical approach.

A second problem with attitude theory stems from the implicit assumption that there is a straightforward way of reducing talk about complex issues to the simple question-and-answer format of the kind used in questionnaires and surveys. If discourse analytic studies of real life discourse are correct in viewing everyday
talk as essentially variable then, contrary to the predictions of attitude theory, ordinary discourse cannot be reduced to a simpler form in which such variability is excluded. In this respect, discourse analysts find much more sympathy with the social representations approach (with its emphasis on social construction) than with attitude research. As we saw, discourse analytic studies of everyday discourse find that ordinary talk relies on a number of different interpretative repertoires: sets of concepts, conversational terms, figures of speech and so on which constitute a resource from which subjects flexibly draw throughout the term of a conversation for the purpose of meeting specific conversational goals. [33] The problem for attitude theory lies in the fact that, in pursuing their immediate conversational goals, subjects often draw from a particular interpretative repertoire in such a variety of different ways (on a given occasion or across a number of occasions) that what is said often comes to seem, to the analyst, confusing or even inconsistent. In consequence, there seems no obvious way in which to draw simple attitude pronouncements from the welter of conflicting real-life data.

A discourse analytic approach which studies, for example, the management of change in conversational subject matter will provides an understanding of how individuals construct a version of the object about which they exhibit beliefs and evaluations without the analyst having to appeal to the idea of attitudinal objects as unproblematic entities. Moreover, such an approach will be able to cope directly with the kind of complex data represented
by everyday talk without appeal to elicited expressions of attitude which are usually over-simplifications of real discourse.
CONCLUSION

Discourse and Psychology

The preceding chapters have argued for a Wittgensteinian perspective on understanding language use. It has also been suggested that such a perspective imposes constraints on the kinds of explanation which empirical social science can generate. One way to meet these constraints, it was suggested, is to adopt the discourse analytic approach.

This approach offers a valuable perspective for areas of psychology other than those dealing solely with self-presentation effects or with conversational mechanisms. It is clear, for example, that the kind of debate centred on the notion of person/situation interaction [1] might benefit from an analysis of the way persons in situations construct or negotiate the meaning which those situations have. Similarly, social constructionists [2] would find the ability to understand naturally occurring speech data an important tool in understanding how, for example, emotional responses are socially created.

Just as importantly, the discourse analytic perspective offers new light on how cognitive processes might be studied. It has recently been argued, for example, that an understanding of the cognitive structures of speech perception may depend on a prior grasp of how human beings cooperate to produce joint communicative effects. [3] This, in turn, suggests that the discourse analyst may prove to be
a useful ally in other areas of cognitive research. An understanding of how people creatively deploy linguistic strategies may prove important in the further development of expert systems, for example. These systems are 'artificial intelligence' programmes designed to mimic the behaviour of a human expert in a specific area of knowledge. Typically, what separates out these programmes from other areas of artificial intelligence is the attempt to combine generation of a human-like set of responses in a given problem area with the ability to generate a human-like rationale for such responses upon request. An expert system will usually be formed out of two basic components: a knowledge base and a 'shell' composed of an inference engine and a user interface. The knowledge base represents the store of task-specific data with which the human expert works, while the shell comprises (a) a set of rules and procedures for manipulating the base data in order to generate answers to questions - the inference engine and (b) a set of procedures for communicating with external users by passing on such answers, requesting more information and so on - the user interface.

The field of application of such a systems approach is already extensive. In general, the most suitable areas in which expert systems can be put to work are those in which the subject area is large and where knowledge claims are best understood as probabilistic. At present, expert systems research has already encompassed such diverse problem areas as medical diagnoses, oil prospecting, organic chemistry, computer fault-finding, foreign
Conclusion

language translation and income tax assessment. In each area, the system attempts to replicate a human expert's fund of knowledge and to emulate the kind of rational, human decision-making which such knowledge allows those experts to produce.

What most expert systems have in common, then, is the fundamentally dualistic form represented by the combination of knowledge base with inference engine. In turn, this means that the task of the expert systems designer can be thought of as falling into two halves: the generation of an explicit characterization of what it is that human experts know (sometimes termed 'knowledge engineering') and the provision of a set of programmes which will act as an inference engine for suitable encodings of that knowledge. What is also common, however, in many areas of application of the expert systems approach is the complaint amongst programmers that knowledge engineering accounts of what experts know often lag behind developments in programming technology. What is needed, it seems, are accompanying developments in the field of knowledge engineering. The general difficulty which faces the knowledge engineer however is that, in general, human beings prove to be poor at grasping how it is that they do what they do. So in order for an expert system to demonstrate that a given conclusion has been reached by recognizably human reasoning processes, the very system designed to mimic human experts often has to display explicit grasp of what, in its human expert counterparts, is at best implicit knowledge - 'rules of thumb', hunches and so on.
Conclusion

It is at this point that the discourse analytic perspective becomes of interest to the expert systems designer. Traditionally, knowledge engineers rely on structured data collection methods such as question and answer sessions or 'talk-through' sessions in which a human expert gives verbal reports while performing tasks. The overall aim is to extract from such data a series of explicit representations of what it is that the expert knows. These are not, however, the only ways in which the crucial task of gathering explicit representations of the knowledge elements relied upon by experts can be carried out. There is a wide range of 'everyday' or naturally occurring speech situations in which experts can be seen to make explicit reference to those items of knowledge which make up the 'lore' or 'rules of thumb' which expert systems designers seek to capture. In everyday situations such as conversations, discussions, arguments, justifications, criticisms and the like, speakers can be seen to give explicit formulation to the ways in which they represent to themselves, and others, a given problem or event. Discourse analysis records and collates such naturally occurring segments of discourse and examines both the kinds of representation in which such discourse consists and the uses to which such representations are put. So to the knowledge engineer, discourse analyses of everyday speech situations offer a radical alternative to more traditional attempts to gather explicit representations of knowledge states from human experts.

What all of this suggests is that the discourse analytic approach may be of use to the social scientist even outwith the bounds of strictly social-psychological research. Many human activities,
Conclusion

even those studied at the level of 'cognitive psychology', can only be understood and explained when the meaning which those activities have for participants is understood. It is the goal of the discourse analyst to place the concept of meaning, understood in terms of the Wittgensteinian perspective outlined earlier, once more at the heart of the study of human life.
NOTES TO CHAPTER 1


[5] These two perspectives on how to provide a philosophical account of language use are often associated with two opposing views on the relationship between language and thought. The first of these views conceives of internal acts of judgement taking place in a 'realm of the mental' which issue in exterior acts of communication. The second rejects this 'mentalistic' model and, instead, pictures social communication as prior to inner thought. Historically, the former mentalistic view has been attached to those theories of language and meaning which aim at producing a theory of meaning by appeal to a single, unifying concept. The latter 'social communication' view is historically associated with those accounts of language influenced by the later work of Wittgenstein. These associations are not essential however. For example, even if the mentalistic view on the relationship between thought and language is rejected in favour of the social communication view, this still leaves scope for the distinction in approach between those who aim to devise a theory of meaning and those, like Wittgenstein, who reject such a procedure, emphasising instead the variety of linguistic acts.


[7] This holism is not, of course, the only difference between contemporary Anti-realism and verificationism. As far as the concept of the theory of meaning's network of sentences is concerned, this will carry further implications. Should the interrelationships between sentences be cast in a 'probabilistic' form (see Dummett (1981):'Original Sinn')? How do we express the relationship between a sentence and other sentences which, if true, defeat the first sentence? Is this a relationship of meaning (Taylor (1987), p.284; p.272; see also Dummett (1976), p.97)? Note in particular the conception of molecular explanation in such a framework as a 'partial ordering'. This raises the question of whether holism is a problem for the theory of meaning approach. (Weir (1985); Blackburn (1984), chapter 5; Baker and Hacker (1984))
In addition, the adoption of molecularism or 'holistic molecularism' still leaves room for opinion on the linguistic significance of language 'atoms' - individual expressions. In particular, the theorist's use of molecularist principles may be specifically designed to leave room for some notion of reference associated with individual expressions. The importance of this lies in the fact that issues of reference are sometimes seen as playing a role in demarcating the two main Theory of Meaning schools introduced later. (See Dummett (1981): 'Realism', p. 400; 'Context Principle', pp.370; Blackburn (1984): chapter 6; Rasmussen (1985))

[8] Now it might be thought that such revision would seem gratuitous - that a language theory/language practice clash ought always to be resolved in favour of the latter (a philosopher's version of falsificationism). But this might not always be the case. Suppose, for example, that the meaning theory issues in certain proposals about tensed statements. These proposals could involve rejecting claims about the metaphysics of time which might otherwise seem independently appealing. The view that truth value attaches to individual statements (to specific utterances of a sentence with assertive force, rather than to sentence types) carries with it, for example, the intuitive idea that the truth thereby ascribed is timeless, if the statement is not overtly context dependent. If the theorist proposes to reject this view, then he must do without that particular metaphysical picture and must regard truth as in some sense relative. Of course opting for one particular metaphysical picture in place of another need not itself be revisionary of natural language, but it will have revisionist results if the new picture requires the abandonment of certain logical laws which are prima facie intuitively acceptable and which are embedded in our normal language practice. See Dummett, (1973/81): 'Thoughts', p. 397. Of course whether such laws can ever be regarded as so 'embedded', and whether a particular view on truth would require the requisite abandonment, are themselves contentious issues. See Dummett (1976), p.103


[10] This concern with truth shows that the debate between the Realist and the Anti-realist has metaphysical implications, but this must be characterized carefully. The Anti-realist is happy to accept that when a statement is true, there is something in virtue of which it is true. Of course, from within this philosophy of language perspective, the idea of 'that which makes a sentence true' must be understood in terms of the meaning theory. It has already been seen that such a theory demonstrates how a sentence is determined as true by showing the role played, in establishing the truth value of the sentence, by the sentence's component parts. It is this constructive aspect of the theory, it will be remembered, which allows the theory to explain our capacity, as language users, to deal with an indefinitely extendible number of new sentences as they arise in a wide variety of
novel situations. Both the Realist and the Anti-Realist not only agree in accepting that, when a sentence is true, there is something in virtue of which it is true, they further agree that our understanding of the way the relevant sentence is determined as true (by that 'something in virtue of which...') is reflected in the language theory's analysis of the sentence.

This is, however, only a limited agreement—only which refers solely to a general conception of the way in which each statement is related to what it is that determines that statement as true, when it is true. This agreement says nothing about the range of cases in which, for any given sentence, 'that which makes the sentence true' obtains. And as noted above, it is here that the Realist and the Anti-Realist diverge. The Anti-Realist insists that, whatever the central concept, specification of that range of cases must somehow involve our recognition capacities. The Realist denies this. It is this feature of the Anti-realism, position which is often taken to generate a rejection of the principle of bivalence. The Anti-realism takes 'true' to mean that any true statement is 'true in virtue of something' yet insists that truth be epistemically restricted. From this it seems to follow that if we cannot discover the truth value of a given statement, we should reject the idea that that statement must be either true or false. Whether this does engender the logical revisionism associated with abandoning bivalence is controversial, however. See the debate between Weir and Tennant: Weir (1983, 1985, 1986), Tennant (1981, 1984, 1985). See also McGinn, (1980, 1981, 1982) and Williamson (1988).

For Wright's criticism of revisionist tendency, see Wright (1967): chapters 10 & 11. A re-statement of anti-realism in terms of the rejection of bivalence is offered in Luntley (1986). For versions of the realism/anti-realism debate beyond bivalence, see Putnam's 'Models and reality' paper in Putnam (1983) (also Koethe (1978) and Hansen (1985)). For criticism on whether realism/anti-realism arguments are best cast in terms of truth at all, see Harré (1986), chapter 1. See also Horwich (1982) and Winkler (1985).

[11] It might be thought that the Verificationist Anti-realist could do away with the notion of truth altogether, but this cannot be correct. Statements of the conditional form whose components are in the future tense have complex correctness conditions. The conditions for asserting the antecedent are not the conditions which must obtain as presented by the antecedent. This means that assertion may have correctness conditions which differ from the correctness conditions associated with the asserting of that assertion. See Dummett (1973/81): 'Assertion'; Dummett (1975), pp 83-88 & 116-118. For criticism, see Appiah (1984, 1986, chapter 8).

Note that we have to settle two issues: (a) how strong should evidence be before it is counted as warrant - this will be settled by convention; (b) what is it that decides, of some evidence, that it is evidence, of a given strength, for a given statement - this is settled by Theory of Meaning. (See Dummett (1976), 86) The question is: how far can we maintain this distinction when we are dealing with defeasible evidence? (see Taylor, (1982))

Within each topic area, it will be important to keep clear the distinction between a dispute about what objects exist and a dispute about the objectivity of truth. Sometimes an account of correctness conditions will entail a denial of the existence of a certain type of object. One implication of Frege's Context Principle (which stipulates that meaning must be thought of in sentential terms), for example, is that it is possible to give a contextual definition of the senses of certain sentences. In such a definition, a sentence which makes explicit mention of an object or objects of a given type is translated into another sentence in which there is no mention of objects of that type. This is made possible by the fact that the meaning theory attaches sense only to whole sentence units. Now when such a contextual definition is given, the implicit claim is that there is no need to regard the expendable expressions from the first sentence, which made mention of objects of a given type, as standing for 'real' objects. In effect, talk of such objects has been reduced to another form of talk. See Dummett (1981): 'Realism', p.452; Dummett (1982); Luntley, (1985); Almeder (1986).

Such reduction, in which talk of a given type of object is held to be dispensable, cannot be completely identified with Anti-realism, however. It is possible for the Realist to be a reductionist. Dummett has captured part of this notion in his use of the terms 'sophisticated realism' and 'naive realism'. (Dummett (1976, 1982)). The former refers to those Realists who engage in some forms of reduction, while the latter refers to those Realists who do not. For the latter, the truth conditions of a given sentence may well not be reducible to any more explanatory formula. In such a case, the 'naive' Realist holds the sentence to be 'barely true' so that an expression of its truth conditions will be of a trivial form. Note that this must still, however, remain consistent with the Theorist's aim of providing a structural explanation, which raises the question of whether the Context Principle is in tension with such 'structural' goals. See Dummett (1981): 'Realism', p400; (1973/1981): 'Abstract Objects', p.499

Because the Realism/Anti-realism debate centres on this question of whether truth is a transcendent matter, there must be some scope for the consideration of epistemology. Just because the notion of 'correct judgement' is prerequisite, the theory of meaning cannot regard itself as 'epistemologically pure'. For example, it has already been pointed out that a theory of meaning will want to explain the
contribution each expression makes to the sense of the sentence as a whole. Now in part this will involve establishing the referential role of that expression. But sometimes, as pointed out above, there will be debate as to whether a given expression plays a genuinely referring role. This debate must involve an epistemological component.

The meaning theorist wants to give an account of what a grasp of correctness conditions consists in. And at some stage, the Realist is likely to appeal to the notion of recognition of the obtaining of truth conditions. Where such an appeal is made, the Realist will be expected to give an epistemological account of what such recognition consists in while the Anti-realist may well seek to question this epistemological account. But this assumes that the proper way to set the agenda for a study of meaning and understanding is to ask what a speaker must know if he is to be credited with understanding an expression. The critic may well object, however, that, though important, epistemological considerations will fall short of the task of explicating such understanding. (Dummett himself may accept some such conclusion - see Dummett (1981): p. xiii; also Sainsbury, (1980); Taylor (1987); Cockburn (1985)). It may be denied, for example, that the distinction between theoretical and practical knowledge is exhaustive of the options required. Note especially that whatever the epistemological account offered, it must be given against a background account of rationality. As long as it is accepted that to understand the use of a linguistic expression demands an understanding of the point of its use - and this is implicit in the meaning-theorist's talk of a theory of force - there must be some explanation of what the human experience of 'seeing the point' is. What is it, in other words, that makes human existence so essentially rational? Now once considerations of rationality are admitted, it becomes important to assess the extent to which such matters are, or are not, a part of the theory of meaning. (See Chapter 3)

[15] See Dummett (1978): chapter 14; McCarty (1983); McDowell (1976); Tait (1983). The Anti-realist position is often generally described as the claim that the assumption of bivalence is required for Realism, but can be demonstrated to be illicit. But it is a moot point whether the acceptance of bivalence marks out the Realist. Even if bivalence is taken to be an important element in the debate between the Realist and the Anti-realist, the real distinction between Realism and Anti-realism may be more complex than a mere affirmation or rejection of the principle of bivalence. As far as sentences are concerned, Dummett suggests - see Dummett (1982; 1981) chapter 20 - that it may be better to characterise Realism as the adoption of a truth conditions account of meaning together with the adoption of a non-epistemically constrained account of truth which results in the adoption of a two-valued semantics where a unary conception might do just as well. At the level of sentence structure, a further requirement for Realism may be a commitment to a particular view on the way individual
expressions contribute, referentially, to the truth value of the statements they appear in.

[16] Note the link between the equivalence principle and Davidson's view on meaning as definable in terms of T-sentences. See Davidson (1984); McDowell (1976) in McDowell & Pettit; Platts (1979): chapter 2; Soames (1984); Wright (1986).

[17] Of course, opponents will argue that, given the caveats mentioned earlier about the suitability of 'knowledge' as an explanatory term in a theory of understanding, and given the lacunae in respect of the Anti-realist's epistemological concerns, this bold claim about practical capacities ought to be, at the least, amended. Understanding, if captured by knowledge at all, must be something in between practical and theoretical knowledge. (See Taylor (1987), p.281; Dummett (1981): chapter 5, p81; Dummett (1976): p.69) What counts is the idea of explaining understanding, as Wittgenstein did, via practice. (See Dummett (1979); Dummett (1981): 'Realism', p.445) This raises the question of how far this requires amendment to the Theory of Meaning approach.

[18] The more usual term is 'truth value link'. The phrase 'temporal responsibility' has been adopted to allow for the distinction, drawn later, between Dummett's position and that of Wright. The truth value link is not, of course, the only challenge. See Appiah's criticisms, for example, in respect of negation: Appiah (1986): chapter 7. Also Price (1983); Shearmur (1986).


[21] This, as will be seen, is the crucial move. But note that it relies upon the formulation of linguistic ability in terms of propositional knowledge.

[22] The use of 'effective' here is particularly important. For the same lack of effectiveness charge is sometimes levelled against the Anti-realist (by the Strict Finitist). See Wright (1987): chapter 4.


[29] For the sake of argument, it is assumed that nothing hangs on Peacocke's reformulation of the Dummett challenge in terms of thought rather than language.


[33] Peacocke (1986)b: p. 16-17

[34] In this regard, note the ambiguity of Peacocke's expression of the contraposition of C:

"A canonical commitment of the content that p is something which, if a thinker finds it fails to hold, constitutively requires him to judge that not-p..." (Peacocke (1986)b: p. 47)

Here all of the work is done by 'a thinker finds (the commitment) fails to hold'. If this is read as a truth about the commitment, then this is a straightforward contraposition of C as it stands. If, on the other hand, it is read as a truth about the thinker, then it collapses the notion expressed by disjunct (b) into that expressed by disjunct (a), and the fulfilment of commitments becomes a recognizable affair.


NOTES TO CHAPTER 2


[3] Now it might seem as though this formulation cannot be non-substantive because the statement of the rule relies on a homophonic translation of 'green'. But it is arguable that this need not be the case, so long as the theoretician is also willing to give some account of what it is to apply a term to green things. Those influenced by anti-realism, for example, might appeal to manifestable perceptual abilities. However the formulation on offer also threatens to be non-substantial in a quite different way. Whatever the notion chosen to fill out the homophonic translation the difference between the non-Wittgensteinian and the Wittgensteinian will consist in the fact that for the latter, the specification of whatever it is that makes homophonic translations useful
cannot count as an authoritative interpretation of the rule for the use of the relevant expression. Whether the compositionality approach is sensible depends on point of view. For criticism, see Schiffer (1987): chapter 7.


[5] See the Wittgenstein works noted in the 'References' section. For a sample of the vast secondary literature, see: Holtzman & Leich (1981); Hintikka (1983); Mounce (1986); Kripke (1982); Wright (1980; 1984); Baker and Hacker (1983; 1984; 1985); Krebbs (1986); Pitcher (1966).


[7] Note the connection between this position and Kripke's criticism of dispositions, which involves the denial that a contingent fact can have normative outcome. (See notes 12 and 14 below.)


[9] If the whole community is affected by hallucinogens, say, we might want to say that for a period it adopts a different practice just because there is nothing to rule out the idea that they are all affected by the drug in the same way.

[10] Kripke (1982): p.86ff. Note that this account has attracted many criticisms. For a sample, see: Goldfarb, (1985); Coates (1986); Tait (1986); Wright (1984; 1986). The primary aim here is not so much to add to that criticism but to demonstrate the importance of RT. Note the similarity of Kripke's position to that outlined in Wright (1980).

[11] Although he uses the jargon, Kripke's account of 'assertibility conditions' is unlike that of the meaning theorist.

[12] Kripke's explicit use of 'inclinations' - e.g. (1982): p. 90 - is bound to remind the reader of dispositional accounts which he previously criticized. The difference, for Kripke, is that the dispositionalist tries to obtain a normative conclusion out of dispositional premises, whereas he, Kripke, is happy that inclinations, and our agreement in such inclinations, be regarded as mere contingencies. For defences of the dispositionalist view, see Duncan (1986); Mellor (1974); Sober in Hookway (Ed.) (1978).


[14] This, remember, is what separates Kripke out from the dispositionalist. (For an account of the inductivism Kripke denies himself, see Lowe (1987).)

[16] Holtzman and Leich (1981): Introduction. The sole exception to the invincibility of the community is in the case of judgements involving expressive uses of words which the community might mistake for referential uses.


[18] Wright in Travis (1986). Note that Wright uses this idea to make out a conception of 'Objectivity of Judgement'. (See also Wright (1987): Introduction, p. 6)

[19] Or at least this is the criticism levelled by Wittgenstein on some interpretations. For debate on whether this is a valid interpretation, see Morrisk (1967); Blackburn (1984); Wright (1984; 1986).

[20] In part, this will require an account of the holistic nature of language use. It is clear that an account of communal usage of 'green', say, cannot be restricted solely to applications of 'green'. For it seems reasonable to suppose that if someone could accord with us in our applications of 'green', but showed no signs of understanding the relationship such applications bore to other colour-word uses, (e.g. that being green all over rules out being red all over) we would be suspicious of the claim that such a person nevertheless fully understood the meaning of green. We would expect, in other words, a limited form of holism to be accommodated within the account of correctness of expression use offered by CT. Moreover, any specification of the way the community uses 'green' must take into account the fact that such uses typically take place in a specific context against a background ascription of human rationality and intentionality. What this means is that a hearer typically assumes that the relevant speaker sees specific contexts in the same way and sees the same point in the utterance as he, the hearer, does. Now in a 'modest' theoretical account, none of this should matter much. Specification of that which a competent speaker knows - specification, in the present case, of what counts as communal usage of 'green' - need only be accompanied by some formulation to the effect that the context of utterance and background of speaker's aims, intentions desires and the like is taken to be normal. A substantive account must, however, do better.

[21] It might be argued that in such a theory the requisite interpretation is provided by a mere description of the communal rule. The whole point of the appeal to CT, for example, is that there can be no authoritative interpretation of a language rule which can be stated independently of communal usage. There can be no substantive explanation of what counts as applying green which does not resolve into saying that to apply 'green' correctly is to do what the community does with 'green'. This suggests that one ought, after all, to be able to state a competent speaker's knowledge in terms of such communal usage: the competent
speaker knows that the rule is to be applied in the way the community applies it. But this expression of the competent speaker’s knowledge will not, by itself, suffice. For now what is required, in a substantive account, is an explanation of what it is to apply the rule in the way that the community applies it. On the account as it is now being developed, in other words, a substantive statement of that which the competent speaker knows must include an explanation of what it is for the community to follow the rule so described. But any attempt at such a substantive statement would rely upon the formulation of precisely the kind of authoritative interpretation which the inclusion of CT was supposed to prohibit. According to CT, the analyst’s account of language must explain what it is for the community to be rule-guided without appeal to an interpretation of the rule which stands independent of communal usage. But such an interpretation must be possible if the competent speaker’s knowledge is substantively expressible in terms of that which he knows. By modus tolens, then, it follows that so long as CT is retained, the model which aims to explain language understanding substantively in terms of ‘knowledge that’ must be rejected.


[24] Naturally, it might be argued that in taking obverse and reverse as an analogy for opposites in meaning, Baker and Hacker illegitimately rely on a particular aspect of the three-dimensional metaphor. Clearly, then, this analogy must be spelled out.

[25] Baker and Hacker (1984). The mistake which, according to Wittgenstein, Russell made. Such internal relatedness, as is seen later, is also a feature of other aspects of the rule-following notion, for rule concepts in general constitute a family of related notions. Thus the rule for the use of an expression is also internally related to whatever counts as ‘doing the same as a previous correct application’, ‘following the practice of using the expression correctly’, or ‘acting in accord with the rule’. For the reference to Russell’s account of desire and Wittgenstein’s reaction to it, see Wittgenstein (1980): p. 9. (See also Russell (1940): chapter 15.)

[26] Namely: as signifying an internal relation. As to whether it is legitimate to take ‘determines’ in this fashion, see later.


[28] Baker and Hacker’s account of the role of consensual agreement is convoluted. They deny that consensual agreement represents a direct criterion of correctness, and say that the entire community can, at times, go wrong in applying its own rules (see, e.g., Baker and Hacker (1985): p 172). But
for there to be any such thing as rule-following, they say, there must be a general 'agreement in action' which forms a 'framework' in which the rule operates. Whether this idea coheres with the general Baker and Hacker perspective will be seen later.


[33] There are two ways to take this. One might claim that the distinction between a rule application and a rule formulation rests upon a prior distinction between different language contexts, a distinction which separates out instances of normative activity from instances of non-normative activity. On the other hand, one might claim that normative activity and rule formulation are internally related, such that one only understands the former when one understands the latter.

[34] This, of course, is the claim explicitly rejected by Kripke.

[35] Wittgenstein (1974): §242; (1978): p. 343; (1980): § 347-350. It is important to be clear that although the range of judgements in which we agree is a contingent matter, it is not a contingent matter that we, as rule-followers, have such agreement. There is no possibility of explaining language use in terms of substantive rule-formulations. So rule-following is only possible where such agreement in judgement exists. In this respect, agreement in judgements is a necessary property of language using communities.

[36] Wittgenstein (1974): §133. This need not be the only method though. Wittgenstein opposed the method of logical analysis as a form of determining the meaning of a statement. But this leaves open the possibility of using logical analysis to generate a meaning.

Notes to Chapter 3

[1] For comment in the secondary literature and elsewhere on this notion, see: Albrinton (1966); Baker (1974); Garver (1962); Hacker (1986): chapter 11; Kenny (1967); Lycan (1971); Macolm (1959); McDowell (1982); Rorty (1973); Scriven (1959); Strawson (1959): chapter 3; Wellman (1961; 1962); Wright (1987): chapters 1, 7 & 8.

302
[2] Attempts have been made to make the distinction between context types more directly explicit: see Cohen (1985; 1986); Pollock (1984; 1987); Horstein (1986).


[7] Questioned, that is, in a sceptical light

[8] Moreover, if background doubts raise danger of disruption, the question of how to make out defeasibility becomes a tricky one.


[10] And by 'apply' we mean: make a judgement

[11] At least as far as thought is founded on concept. Of course the question of which grounds are appropriate will vary from case to case. There is no suggestion here that criterialism represents a general solution to all sceptical challenge. See Bogen (1974); Hanfling (1984a; 1984b; 1985; 1987). For the background debate in epistemology, see Bogdan (1985); Harman (1980); Williams (1960); Fogelin (1976) chapter 12.


[15] Of course he might claim that this is supplied by an analytic identity. But this presupposes that the content of all theories might be fully explicated by means of restatement in observational terms. And this seems too tall an order. Pearce (1981); Laudan (1986a; 1986b).


[17] The use throughout the present paper of 'ethnomethodologist' and 'strong programmer' is probably an invalid application of the economy rule for categorisations. If the reader so desires, he or she can read these category terms as referring to groups of individuals brought together in common reference in order to fulfill the goals and interests of the present author. For the ethnomethodological perspective, see Sharrock and Anderson (1984; 1985; 1986).

[18] Now one objection to this might be that one could retain C and T but reject the idea that a statement's truth (or falsity) need be regarded as a determinate issue. The idea here would be that M need not conflict with C or T because
truth, like meaning, is fluid, labile entity. Suppose one said, for example:

"It depends on how talk is managed in practical situations. On some occasions, to suit one set of purposes, one could take a given statement as meaning one thing, and so regard it as true, whilst on other occasions one could take it as meaning quite the opposite, and so regard it as false, by appealing to different features of the statement and its context of utterance. In such cases, we might want to claim that the statement was both true and false."

Does this offer a picture in which M becomes acceptable? Couldn't we say that on this picture, M both means what it says and represents a rejection of the determinacy of meaning talk? Couldn't we say that, paradoxically, M is neither false nor true? Well, first it's important to note that such talk of 'occasioned purposes' and 'management of talk' might well, in a way, presuppose both C and T in their original, determinate forms. It depends on the extent to which we can make sense of indeterminate commitments. So even if this picture makes M coherent, we might still be forced to conclude that M is false (given that holding to M rules out holding to the original versions of C and T).

Anyway, the implication that the occasioned nature of meaning shows that statements might at one and the same time be recognisably neither false nor true is misleading. If a statement is true under one set of circumstances and false under another, and if those circumstances are not yet fulfilled (or if the evidence as to which of them pertains is not yet to hand) then it might make sense to reject the claim that the statement is either true or false (if one has no sympathy with the realists). One could say: I don't know yet. But this does not mean that one is warranted in claiming, under any circumstances, that the evidence shows that statement in question is recognisably neither true nor false.

What this means is that an attempt to fill out the indeterminacy of meaning by talk of occasioned practices and so on may well colour the way we understand C and T. But it does not represent a background against which M can be made coherent.


[21] Moreover, there is a further embarrassing question (Dummett (1978): chapter 11; Wright (1980): chapters 18 & 19; Putnam (1979): chapter 9). Is the application of a convention itself a conventional matter? It seems likely that the Strong Programmer will answer in the affirmative. For if the application of conventions is not a conventional matter, then the strong programmer will be required to explain what it is that distinguishes cases of contingent concept application from cases of convention application in which the concepts involved non-contingently entail the methods of their own application. If, on the other hand, the application of conventions is another case of contingent concept application, then the appeal to convention goes no way towards explaining the patterned nature of communal concept use.

Take the case of an analyst's explanation of an item of participant's concept application in which the participant concurs with communal usage. The analyst's explanation is supposed to take the form of the identification of a communal participants' convention grounded in participants' interests (call it convention C) which explains the way the concept is used by demonstrating how such a convention would serve communal interests. But if the application of convention C is, itself, a conventional matter, then it is clear that before talk of C can be taken as an explanation of the participant's concept application, the analyst has to explain why convention C should have been regarded as applicable to the specific act of concept application in question. Now it might be thought that this matter can be resolved by invoking a second convention (call it convention C'). The suggestion might be that appeal to C', and the interests from which it issues, explains why participants should apply convention C to occurrences of concept application such as the one in question. But this cannot be regarded as satisfactory. For the issue then would be: how is convention C' to be applied? If its application is likewise a matter of convention, then the strong programmer's account cannot be taken as an explanation of the patterned nature of communal concept application until the question of why C' should be deemed to apply to C is resolved. The problem is, then, that, short of an appeal to an infinite regress of conventions, the strong programmer offers no explanation of why, given the supposedly arbitrary and free nature of the judgements each individual brings to concept application, there should be any such thing as agreement in usage. According to the strong programme, the patterned nature of social life is mysteriously inexplicable.

[23] Pettit (1986): p. 41. Of course, it is a moot point whether this provides the kind of regularising/normalizing distinction for which Pettit seeks. (Pettit himself accepts that regularising explanations could be a priori. He merely points out that we have no explanation as to how such a priority could be explained.


[26] This leaves undiscussed the debate over whether reasons are causes. For this debate, see: Davidson (1980); MacIntyre in Laslett & Runciman (1972); Humphreys (1986); Swain (1978). As to whether reason could, anyway, be accommodated within an explanatory framework, see: Harré (1972); Harré & Secord (1972); Laudan (1986); Pratt (1978); Ryan (1970; 1973); Winch (1958; 1970). For the question of whether this explanatory framework would be a scientific framework, see Hempel (1968) and Davidson's response in Davidson, (1980). (For Hempel's use of 'reduction sentence' in Hempel (1968), see Carnap (1950; 1956)). Whether the resultant science, if achieved, would be a psychological science is also a matter of debate: see Coulter (1982); Heil (1981; 1985; 1986); Dennett (1982); Dupre (1983); Kim (1979); Loewer (1987); Stromberg (1981).

Notes to Chapter 3: Appendix


[6] Or, at least, this is one version of an ethnomethodologist's counter criticism as interpreted by strong programmers - see Barnes (1981). See also Manicas and Rosenberg (1985).

[7] This need not be taken as criticism, the ethnomethodologist says, and especially not personal criticism. See e.g. Woolgar's 'Critique and Criticism: Two Readings of Ethnomethodology'.

Indeed at least one "critic", Woolgar, claims that this is the thrust of his complaint against strong programmers (see 'Critique and Criticism...'). Woolgar's complaint, he says, is that the appeal to interests is under suspicion not because it gives an explanatory account but because it seems to purport to give the explanatory account of participants' activities. The strong programmer's application of concepts involves judgement, it is agreed. So although the strong programmer's decision is to present his or her text in such a way that the discussion of interests is in terms of causal entities, things might have been otherwise. Interests might have been presented as nothing but artefacts of participants' accounts, for example. But the strong programmer's presentation of the text hides this potential for interpretative variability. Instead, the reader is more or less forced to construe interest-talk in terms of stable, 'real-world' entities. This shows that something is wrong, says Woolgar. Even if the strong programmer does admit to the theoretically constructed nature of his interest artefacts, this does not 'come across' in the text. The criticism is, then, not that there is anything necessarily wrong with the appeal to interests, but that the way this appeal is presented seems to rule out other alternatives, and thereby wrongly delimits the range of potential variability in textual interpretation. What this shows, Woolgar concludes is that greater attention ought to be paid to the phenomena under-pinning this diversity of interpretative resource.

See e.g. Barnes (1981): p. 30

'Decision' here need not necessarily refer to conscious decision. The relation of practice to rule-following is an issue which strong programmers themselves have discussed, cf. Bloor (1981): p. 121. Most of the discussion seems, though, to centre on the idea that routine practice is an area of life in which decisions are made unconsciously. To this extent, such arguments seem to make the same assumption about the universality of contingent judgement in concept application as is made by Barnes. See, e.g. Barnes' claim (Barnes (1981): p. 31) that even in cases of routine concept application it is "formally appropriate" to describe concept applications in terms of decision-making. Of course some of Barnes's 'left-wing' opponents accept this idea of non-determinateness but deny that it constitutes a problem to be resolved: See Woolgar, S. 'Irony in the social study of science', in Knorr-Cetina & Mulkay (1983); Mulkay, M. (1985). The phrase 'left-wing' is taken from Manicas and Rosenberg, op. cit.

see references at note (1)


Notes to Chapter Four

[1] For the sake of clarity, the references to Moscovici's own work have, in the main, been given a separate format. Each of the four main texts referenced have been allocated a code letter in the following way: Moscovici (1983) = A, Moscovici (1984a) = B, Moscovici (1984b) = C, Moscovici (1985) = D. This is followed by a page number and the whole is enclosed in square brackets.

[2] Although such a formulation might require complex empirical investigation.

[3] There is also, of course, a substantial body of empirical work on social representations, much of it in French.


[8] There is no space to explore in this chapter the problematic and as yet unexplicated relationship between Moscovici's work on social representations and his work on minority influence.


[14] Such symbols nevertheless, according to Moscovici, correspond to 'objects in the outside world': Moscovici (1984): p.5

[15] Indeed this follows directly from Moscovici's claim that all thought is necessarily social - (1983): p. 135 - and that all
action, to be understood, must be translated into social representations - (1984a): p.951

[16] On occasion, Moscovici claims that the nature of a group’s social representations is partly a matter of the culture of that group (Moscovici (1983): p.134) and this line of thinking has been developed by Farr (1984, 1986). However given (a) the variety of technical uses of the notion of culture and (b) the complex issues raised by theoretical studies of culture and (c) the multifarious everyday uses of this concept, it seems clear that 'culture' does not denote a precise, well defined phenomenon. It is unlikely, therefore, that an appeal to the idea of social representation-as-cultural-product will, of itself, resolve the problematic issues raised here. It seems, rather, that a range of further problems - specific to the theory of social representations - may well present itself, such as the difficulties involved in the analytic disentanglement of 'culture' from 'consensual universe', the clear separation of 'culture' from 'social representation' and the model implied by the suggestion that social representations are cultural products.

[17] Moscovici suggests that social psychology should deal with the consensual universe, which is the realm dealt with by social representations. Since science is the study of reified universes, this precludes taking social psychology (or similar enterprises such as 'mass psychology' or 'political economy') in any unproblematic way as a science.


Notes to Chapter 5

[1] Where 'object' means 'the object as thought of by the participant'. For the question of how the meaningfulness of human activity intrudes on simple notions such as 'attitudinal objects', say, see Tajfel (1981; 1982).

[2] It is, as we saw in Chapter 3, a contentious point whether we need always separate out description from explanation. (See: Harré (1979); Harré and Secord (1972).)

[3] For the various ways in which the discourse analytic approach is currently developing, see: Antaki (1985); Levinson (1983); Jefferson (1985); Atkinson & Drew (1979); Atkinson & Heritage (1984); Atkinson (1984); Brown & Yule (1983); Drew (1978; 1986); Edwards & Middleton (1986a; 1986b; 1987); Pomerantz (1978a; 1978b); Schenkel (1978); Watson (1978); Litton & Potter (1985); Potter (1983); Potter and Litton (1985a; 1985b); Potter and Wetherell (1987); Potter, Stringer and Wetherell (1984); Stubbs (1983); Thompson (1983); Wetherell (1986); Winant (1983); Wowk (1984); Henriques et al. (1984); Holtgraves (1986. For the contrast between DA and its more conservative cousin, attribution theory, see Hewstone (1981); Milech & Nesdale (1984); Semin (1980).
[4] Of course many discourse analysts are equally interested in written discourse.

[5] There is no suggestion that participants need be conscious of these attempts to organize discourse.

[6] This has produced a theoretical interest among social scientists over the extent to which rhetoric is a conversational feature. (See Billig (1984; 1987); Atkinson (1984).)


[8] The fact that there are no necessary and sufficient conditions for topic identity is neatly brought out by the example quoted from a Sacks lecture in Levinson (1983), p. 292:

A: I have a fourteen year old son
B: Well that's all right
A: I also have a dog
B: Oh I'm sorry

Most people would be willing to accept that dogs and fourteen year old sons are quite different topics. But in the quoted example they form a single topic, the nature of which becomes obvious only once the specific context is made clear: A is attempting to rent a room from B. Once this is stated, it becomes clear that sons and dogs form a single topic which might be termed 'potential rental disqualifiers'.


[13] Ibid. p. 218


Notes to Chapter 6

[1] This may explain the possibility of the kind of explanation/description overlap mentioned in note 2 of the previous chapter.


[8] See references at note [3]

[9] This need not, of course, simply be an assumption that any specific utterance will be performing a single task. In particular, the notion that a scientist is defending his or her own view can sometimes be taken to entail that an opponent's view is being, at least implicitly, criticized. The suggestion here is only that participants' accounts of actions and beliefs fall into two broad groupings, one grouping representing participants' explanations of why his or her own view is preferable, the other representing accounts of why opponents make the mistakes they do.

[10] All names are, of course, pseudonyms. Numbers added in parenthesis to the quotations refer to pages of the conference transcript.


[12] For detailed analyses of the organization of blamings in everyday and courtroom discourse, see J.M. Atkinson and P. Drew (1979): in particular Chapters 4 and 5; Drew (1978); R. Watson (1978); M.T. Wowk (1984). Atkinson & Drew, op. cit. Chapter 5, concentrate on the manner in which council can develop blamings over a number of turns, and looks at some of the techniques which witnesses can use to minimize the effect of the blaming - for example, downgrading the severity of the imputed damage, or providing an account providing strong casual constraints on activity. The other studies concentrate on the way blamings can be deflected with different kinds of categorizations, which, in turn, imply different, generally less culpable motivations.


[14] This pattern is comparable to that found by A. Pomerantz, (1978). Pomerantz found that in everyday sequences where issues of blame were relevant, the first turn typically avoided direct reference to the blameworthy actor or agent.
This allowed blame to be allocated to self in a subsequent turn.

[15] It is not claimed that this identification of bottom-up with behaviourism is anything more than an occasioned phenomenon. In the data we have examined terms of this kind have a flexible, shifting character. For example, just prior to Extract 18, Steele spends some time in presenting a redefinition of the terms in which the top-down versus bottom-up debate should be couched. Other research on scientific categories has shown their flexible deployment on different occasions: Gilbert & Mulkay (1984): chapter 7; J. Potter (1983), Chapter 6; S. Yearley (1984).

[16] The use of 'great names' with whom 'all scientists agree' is a common rhetorical tack. Darwin was the popular choice at this conference. See Potter (1984).

[17] The term 'motivation' is not used here to refer to inner psychological dispositions but to characterize pragmatic functions and constraints on discourse.


[19] The notion of discursive organization being a consequence of multiple constraints is an important one in conversation analysis. It has been used, for example, to explain the problems which arise in the receipt of compliments in both everyday and scientific discourse; A. Pomerantz, (1978); M. Mulkay, (1985).

[20] The difference here is not only large but statistically significant ($x^2 = 9.42, p < 0.01$).


[22] The issue of the hearability of blamings is addressed in the conversation analytic literature; see note 3.


[24] This perspective on discourse is outlined in more detail in Heritage, op. cit. See also Mulkay (1985).

[25] For comment on attitude theory: McGuire (1985; 1986); McCauley (1980); Jowell & Witherspoon (1985); Cvetcovich & Baumgardner (1973); Smith (1982; 1987); Grube et al. (1986); Ajzen & Fishbein (1977); Bagozzi (1986); Bem (1981); Brewer (1979; 1985); Cialdini et al. (1981); Davidson & Morrison (1983).


Notes to Conclusion


## REFERENCES

**Philosophy References**

<table>
<thead>
<tr>
<th>Author</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baker, G.P.</td>
<td>Scepticism, Rules and Language</td>
</tr>
<tr>
<td>Bernstein, M.</td>
<td>'Socialization and autonomy', Mind, 92, (1983), 120-123.</td>
</tr>
</tbody>
</table>
References


Cohen, S........... 'Knowledge, Context and Social Standards', Synthese, 73, #1, (1987), pp 3-26


Craig, E............ 'Meaning, use and privacy', Mind, 91, (1982), 541-564.


Dennett, D.C......... 'How to study consciousness empirically or nothing comes to mind', Synthese, 53, (1982), 159-180.

315


Dummett, M. .......... 'Realism', Synthese, 52, #1, (1982), pp 55-112


316
References


Hacker, P.M.S......... 'Events, ontology and grammar', Philosophy, 57, (1982), 477-486.


References

Hanfling, O............ 'Was Wittgenstein a sceptic?', Philosophical Investigations, 8, (1985), 1-16.


Harré, R. & .......... The Explanation of Social Behaviour, Secord, P.F.


Hintikka, M.B. &.... Essays on Davidson, Oxford: Clarendon Vermazen, B.

Hollis, M. &.......... Rationality and Relativism, Oxford: Basil Lukes, S.

Holtzman, S.......... Wittgenstein: To Follow a Rule, London: Leich, C.

References


Laudan et al........... 'Scientific change, philosophical models and historical research', Synthese, 69, (1986b), 141-223.


References


McDowell, J. ............ 'Values and secondary qualities', in T. Honderich (Ed.), *Morality and Objectivity*,
References


Pettit, P.........


McGinn, C............  'Realist Semantics and Content Ascription', Synthese, 52, #1, (1982), pp 113-134.


References


Anderson, B.

Anderson, B.

Anderson, B.


Smith, M.A............. 'Peacocke on red and red''', Synthese, 68, (1986), 559-576.


Tennant, N. ............ 'Is this a proof I see before me?', Analysis, 41, (1981), 115-119.

Tennant, N. ............ 'Were those disproofs I saw before me?', Analysis, 44, (1984), 97-105.

Tennant, N. ............ 'Weir and those disproofs I saw before me', Analysis, 45, (1985), 208-212.


324
References


[1929-1930]

[1929-1932]

[1930-1932]

[1931-1934]

[1932-1935]

[1933-1935]

[1937-1944]

[1943-1949]

[1945-1948]

[1946-1947]

[1947-1948]

[1949-1951]


Wright, C. 'Rule following and constructivism', in Travis op. cit., (1986a).


Wright, C. 'How can the theory of meaning be a philosophical project?', Mind & Language, 1, (1986c), 31-44.


Wright, C. 'Scientific realism' in MacDonald and Wright (Eds.), op. cit., (1987), (1987b).

Psychology References


Argyle, M. 'Sequences in social behaviour as a function of the situation', in Ginsburg,
References


Barnes, B. *Social life as bootstrapped induction*, *Sociology*, 17, (1983), 524-545.

References


<table>
<thead>
<tr>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
</table>
References


<table>
<thead>
<tr>
<th>Reference</th>
<th>Citation</th>
</tr>
</thead>
</table>
References


References

Knorr-Cetina, K.D. & Mulkay, M.


Kress, G & Hodge, R.


Kugel, P.

'Thinking may be more than computing', Cognition, 22, 137-198, (1986).

Lalljee, M., Brown, L.B., Ginsburg, G.P.


Lemert, C.C.


Levinson, S.C.


Litton, I. & Potter, J.


Manicas, P.T. & Rosenberg, A.


Manstead, A.S.R. & Tetlock, P.E.


Marsh, P., Rosser, E. & Harré, R.


Maynard, D.


McCauley, C. et al.


McGuire, W.J.


McGuire, W.J.

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wattenmaker, W.D.</td>
<td></td>
</tr>
<tr>
<td>Nesdale, A.R.</td>
<td></td>
</tr>
<tr>
<td>Moos, R.H.............</td>
<td>&quot;Sources of variance in responses to questionnaires and in behaviour&quot;, in Social Encounters, (1978), Argyle, M. (Ed.).</td>
</tr>
<tr>
<td>Moscovici, S..........</td>
<td>&quot;Society and theory in social psychology&quot;, in Israel &amp; Tajfel, op. cit., (1972), 17-68.</td>
</tr>
</tbody>
</table>


Mulkay, M. &............ Gilbert, G.N. 'Accounting for Error', Sociology, 16 (1982), 165-83.


References


Semin, G. 'The individual, the social and the social individual', British Journal of Social Psychology, 25, (1986), 177-180.


