

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

**A Thesis Submitted for the Degree of PhD at the University of Warwick**

<http://go.warwick.ac.uk/wrap/50789>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

REALISM AND EVIDENCE  
IN THE PHILOSOPHY OF MIND

LAURA JANE BENNETT

Submitted in fulfilment of the requirements for the  
Degree of Doctor of Philosophy

University of Warwick  
Department of Philosophy

June 1992

## Contents

Abstract

Introduction	1
Aims of the thesis	1
Structure of the thesis	4

### Section A

#### Chapter 1 - Motivation, Methodologies and the Veil-of-Perception Issue

1.1	Introduction	9
1.1.1	Sentences and synapses: motivation, evidence and methodologies	9
1.1.2	Questions and answers: methodology or metaphysics	14
1.2	Theories, explanation and evidence	16
1.2.1	What is a theory?	16
1.2.2	How does a theory explain?	17
1.2.3	The veil-of-perception issue	19
1.2.4	Hume and the veil-of-perception issue	20
1.2.5	Wider implications of the veil-of-perception issue	23
1.2.6	The veil-of-perception issue in twentieth century psychology	28
1.2.7	Gregory and the veil-of-perception issue	30
1.2.8	Neisser and the veil-of-perception issue	35

1.3	Summary of the major points of chapter 1 and their implications for the philosophy of mind in the twentieth century	39
-----	---	----

**Chapter 2 - Realism, Bootstraps and Bias: Problems of Evidence in the Philosophy of Science**

2.1	Introduction	44
2.2	Two kinds of realism	46
2.2.1	Metaphysical realism and the veil-of-perception issue	46
2.2.2	Scientific realism, the veil-of-perception issue and the problem of evidence	48
2.3	Scientific realism and the bootstrapping device	53
2.3.1	The Glymourian bootstrap method	55
2.4	Evaluation of the bootstrap strategy - part I	62
2.4.1	Achinstein's criticisms against Glymour and their implications	65
2.4.2	The outlook for philosophers of mind who bootstrap - part I	69
2.5	Evaluation of the bootstrap strategy - part II	73
2.5.1	Nordby's account of bootstrapping	74
2.5.2	Criminal elements in Co-evolutionary Philosophical Theories	85
2.6	The outlook for co-evolutionary theories	90

<b>Section B - Bootstrapping and Realism in the</b>	99
<b>Philosophy of Mind</b>	
<b>Introduction</b>	
B1 The story so far	99
B2 Application of the veil-of-perception issue to theories of the mind/brain	103
B3 An Introduction to the issue of representation in the mind/brain	109
B4 Structure of Section B	120
<b>Chapter 3 - <u>Metaphysical Elements in McGinn's</u></b>	<b>123</b>
<b><u>Account of Consciousness, Modularity</u></b>	
<b><u>and the Mind</u></b>	
3.1 Introduction	123
3.2 Metaphysics and McGinn	125
3.2.1 An outline of McGinn's theory of consciousness	126
3.2.2 The bootstrapping device in McGinn's argument	129
3.2.3 Putting the con into consciousness	132
3.2.4 Consciousness and the brain	135
3.2.5 Blindsight and blind judgement	141
3.2.6 McGinn and the veil-of-perception issue	147
3.2.7 Conclusion - McGinn the metaphysical realist	151

<b>Chapter 4 - <u>Taking Fodor to task: Churchland,</u></b>	<b>154</b>
<b><u>Mentalese and metaphysical realism</u></b>	
4.1 Introduction	154
4.1.1 What is mentalese?	156
4.2 Fodor's use of the bootstrap strategy	160
4.2.1 Fodor's support for the bootstrap strategy as a device for the development of psychological theories	160
4.2.2 The bootstrap line of argument within Fodor's Representational Theory of Mind	162
4.3 Fodor's attitude to realism	171
4.4 Churchland's criticisms of the language of thought	174
4.4.1 Introduction	174
4.4.2 Churchland's criticisms of Fodor's inference to $H_2$ from $H_1$ and $E_1$	177
4.4.3 Churchland's criticisms of Fodor's inference to $H_1$ from $E_2$ and $H_2$	193
4.4.4 Fodor unmasked - the metaphysics behind Mentalese	209
4.5 Fodor, the bootstrap strategy, and the problem of empirical underdetermination	218
4.5.1 Introduction	218
4.5.2 Empirical underdetermination in Fodor's representational theory of mind	219
4.5.3 Achinstein's criticisms of the bootstrap strategy as they apply to	221

	Fodor's representational theory of mind	
4.5.4	Bootstrapping in Fodor: the Nordby paradigm	222
4.6	Conclusions	225
<b>Chapter 5</b>	<b><u>Modularity and Metaphysics</u></b>	<b>232</b>
5.1	Introduction	232
5.2	What are modular theories of the mind?	234
5.3	Modular theories and the use of the bootstrap strategy	242
5.4	Modules: intelligent giants or ignorant dwarves? Gardner versus Fodor	250
5.4.1	Central cognitive processes: the status of the myth	250
5.4.2	Empirical underdetermination and the criteria for modularity	252
5.4.3	Memory and modularity: is Gardner's theory remotely tenable?	258
5.4.4	A modular account of general processes	262
5.4.5	Summary of the Gardner versus Fodor debate	264
5.5	Shanon versus Fodor - further failures to evaluate the evidence	268
5.5.1	Introduction	268
5.5.2	Remodelling modularity	270
5.5.3	Methodology under the microscope: Shanon on Fodor	275

5.5.4	Blurring the focus: Shanon's modules	280
5.5.5	Shanon, Fodor and the prospects for science	284
5.6	Some better criticisms of the Fodorean modularity debate	285
5.7	Fodor's modularity theory and the veil-of-perception issue	292
5.8	Conclusion	294
<b>Chapter 6 - <u>Metaphysical realism in co-evolutionary theories of the mind/brain</u></b>		298
6.0	Introduction	298
6.1	Marr and metaphysical realism	303
6.1.1	Marr's theory in outline	303
6.1.2	The bootstrap strategy within Marr's theory	309
6.1.3	Criticism of Marr's bootstrap strategy	320
6.1.4	Summary of the lessons to be learned from Marr's bootstrap method	342
6.2	Churchland's place in the debate	343
6.2.1	A brief outline of Churchland's co-evolutionary theory of the mind/brain	347
6.2.2	Churchland and the bootstrap method	354
6.2.3	Criticisms of Churchland's eliminative materialism	365
6.2.4	Churchland and the veil-of-perception issue	382



6.2.5	Analysis of Churchland's failure	386
6.3	Conclusion	388
<b>Section C</b>		
<b>Chapter 7</b>	<b><u>Evidence or Ignorance? Some final reflections</u></b>	<b>394</b>
7.1	Introduction	394
7.2	Summary and conclusions	394
7.2.1	Aims of this thesis	394
7.2.2	Structure of my argument	396
7.3	Conclusions	397
7.4	Future Work	401
	<b>Bibliography</b>	<b>403</b>

### Acknowledgements

I should like to thank all the relatives and friends who have managed to keep me relatively friendly whilst writing this thesis. I am particularly grateful to my parents for their unerring support during my years at Warwick, and to my husband, Steve Mason, for his tolerance, technical assistance and enduring good humour. Thanks are also due to my supervisor, Greg Hunt, who has been unstinting with both his time and encouragement. Without the help of all these people, and many more too numerous to mention by name, I doubt that this thesis would ever have been written.

## Abstract

This thesis evaluates a variety of important modern approaches to the study of the mind/brain in the light of recent developments in the debate about how evidence should be used to support a theory and its constituent hypotheses. Although all these approaches are ostensibly based upon the principles of scientific realism, this evaluation will demonstrate that all of them fall well short of these requirements. Consequently, the more modern, co-evolutionary theories of the mind/brain do not constitute the significant advance upon more traditional theories that their authors take them to be.

There are two fundamental elements within my discussion of the relationship between evidence and the constituent hypotheses of a theory. Firstly, I shall demonstrate that the traditional veil-of-perception issue has a wider relevance than that which has historically been attributed to it, since it is the paradigm case of an attempt to construct a two level theory on the basis of evidence that does not adequately support either hypothesis. This interpretation of the issue can be represented by constructing a semantically inconsistent tetrad. It is shown that similar tetrads can be constructed for each of the theories of the mind/brain discussed in this thesis.

Secondly, I shall argue that the theories discussed all employ a variety of the bootstrap strategy. This strategy is a relatively recent development in the philosophy of science, which suggests a way in which the same evidence can be used to generate both a general and a specific hypothesis within a theory without violating the constraints of scientific realism. However, I contend that recent use of this strategy in the investigation of mind is largely unsatisfactory as a result of a neglect of structural as well as more informal influences upon the kinds of evidence employed to support the hypotheses contained in the theories.

The thesis is divided into three major sections. The first (Section A) discusses the influence of the motivations of the individual theorists upon their arguments and provides a critical discussion of the issues of the veil-of-perception and bootstrapping. The second section (Section B) comprises a detailed examination of a range of modern theories of the mind/brain and critically analyses their success. The final section (Section C) draws together general conclusions and methodological consequences of the detailed analysis of the nature of realism and evidence in the philosophy of mind.

## Introduction

### Aims of the thesis

This thesis has four main objectives. My primary aim is to compare the more modern, co-evolutionary<sup>1</sup> approaches to the study of the mind/brain with more traditional theories based upon conceptual analysis to establish whether the former group really represent the major advance over the latter group that their authors take them to be.

This evaluative task necessarily raises the question of appropriate criteria for the comparison of such disparate theories. I shall therefore argue that it is imperative for theories of the mind/brain to provide us with testable explanations of how the processes and representations within the mind/brain actually give rise to our mental experiences. My second objective is to show that theories that fail to provide us with testable accounts suffer from the difficulties of metaphysical realism.

Moreover, I shall demonstrate that the veil-of-perception issue has a wider relevance than it has been assigned in the recent history of philosophy, since it constitutes the paradigm case of a theory that suffers from the problems relating to evidence and testability generally associated with metaphysical realism. Further, I shall illustrate that it is possible to isolate a semantically inconsistent

tetrad within the veil-of-perception issue which exemplifies the difficulties inherent in basing a two level theory upon evidence that does not adequately support either hypothesis. Similar tetrads may be isolated within all the theories of the mind/brain discussed in Section B, thereby proving that they all suffer from the problems of metaphysical realism.

My third and fourth objectives are concerned with an examination of the methodologies employed to construct the theories of the mind/brain described in Section B. I hope to show, firstly that, although the theorists discussed all rely upon different sources of evidence and postulate a variety of representational systems within the mind/brain, they all employ a version of the bootstrap strategy in order to derive both a general hypothesis and a specific hypothesis from the same initial body of evidence. I shall endeavour to render these bootstrapping arguments explicit before going on to examine the standard of the evidence that their authors cite in support of their hypotheses relative to their individual theories of the mind/brain.

Finally, I shall show that the arguments of Nordby (1989) concerning the difficulties faced by detectives when they bootstrap in order to solve a crime in the absence of any overall theory concerning the relative importance of diverse sources of evidence are readily extended to account for the particular difficulties encountered by co-

evolutionary theorists of the mind/brain who employ the bootstrap strategy.

It is essential to note from the outset that this thesis is by no means an attempt to evaluate the entire range of current theories about the nature of the mind/brain. I have chosen to concentrate upon a selection of views that I take to be representative of the mainstream of their various traditions within the philosophy of mind, in the hope that, if my arguments apply to these theories, then they may be accepted as having a wider relevance than they might have been had I selected only very obscure theories. Nonetheless, it is inevitable that I have had to omit any discussion of some influential contributors to the study of the mind/brain, including Gibson, Vygotsky and Stich.

I do not intend to champion the work of any one theorist working in this field. Indeed, it will become clear in the course of my arguments that I believe that all of them have made fundamental errors in the construction of their theories.

It should also be noted that I shall not be concerned with the philosophical minutiae of what constitutes a good or a bad explanation, nor indeed, with the details of what may be regarded as good or bad evidence. In what follows, I shall merely be concerned to show that theories cannot avoid the problems of metaphysical realism simply by using

the bootstrap strategy if their hypotheses are systematically protected from potential sources of empirical refutation or are backed by evidence that is itself either untestable, or equivocal between the preferred theory and a rival theory, or has already been disputed.

### Structure of the thesis

The first part of the thesis (Section A) contains a thorough discussion of the main criteria that must be fulfilled for an explanation to satisfy the demands of scientific realism. Chapter 1 begins with an examination of the way in which the traditional veil-of-perception issue can be reformulated in terms of questions about the structure of explanation and the criteria employed to decide which kinds of evidence are acceptable (see especially subsections 1.2.4, 1.2.5). It will then be demonstrated that it is these very kinds of questions which render the debate between metaphysical realism and scientific realism pertinent to the philosophy of science and in particular to the philosophy of psychology. The first chapter will also contain illustrations of the way in which the structure of the traditional veil-of-perception argument is still found in modern psychological theories of mind (see subsections 1.2.7 and 1.2.8).

In the second chapter, a paradigm methodology for coping with the difficulty of basing specific and general knowledge on the same epistemological foundations, bootstrapping, is discussed and re-evaluated. This evaluation includes a discussion of how Nordby's (1989) considerations concerning the use of the bootstrap methodology in detective work may apply to co-evolutionary theories in the philosophy of psychology, that is, to theories that evolve from a synthesis of philosophical thought with neuroscientific data.

Section B begins with a brief introduction to the theories of the mind/brain that are to be evaluated in this part of the thesis, including their views concerning the nature of representations within the mind/brain. I then proceed to discuss each theory in turn, starting with Colin McGinn's treatment of the question of consciousness, which demonstrates the paucity of the arguments employed in traditional conceptual analysis when dealing with the issues of evidence, metaphysics and science.

Further confirmation that arguments based upon conceptual analysis do not always yield the satisfactory and testable hypotheses that are required for a successful use of the bootstrap strategy is provided in my discussion of the Fodorean representational theory of mind in Chapter 4. I shall show that the evidence upon which Fodor relies in support of his language of thought hypothesis is



unsatisfactory, and worse still, that Fodor is guilty of systematically protecting his hypotheses from possible sources of empirical refutation, thereby rendering them untestable. Once more, the difficulties of metaphysical realism are revealed within the bootstrap strategy. I shall contrast Churchland's attitude towards the problems endemic in Fodor's theory with her neglect of the same issues in her own arguments (discussed in Chapter 6).

The very traditional view of man as a being with a mind and a body set apart from the external world is challenged by modular theories of mind, such as that of Jerry Fodor, which is discussed in Chapter 5. Fodor questions two key assumptions that are often made by traditional philosophers of mind. Firstly, he argues that the boundary between the cognitive self and what is cognized may not be exactly congruent with the division between man's body and the external world. Secondly, by dividing the activities of the mind/brain into perception and cognition, Fodor queries the notion that is traditionally thought to lie at the very heart of personal identity, namely the unity of consciousness. Fodor's version of the modularity thesis will be defended against criticisms as the most tenable of its kind. However, this does not mean that he can avoid the accusation that he too incorporates metaphysical assumptions into his theory. This issue will also be discussed in the course of Chapter 5.

In the final chapter of Section B, Chapter 6, more co-evolutionary theories of the mind/brain are examined. I begin with David Marr's three-tiered theory of information processing which strongly influenced his account of vision. I then move on to a consideration of Churchland's arguments for eliminative materialism. I shall demonstrate that, although the two theorists have completely different views of the way in which co-evolutionary methodologies should deal with evidence from neuroscientific sources, this has no effect upon their inability to avoid the use of untestable hypotheses and constraints in their bootstrap arguments.

The third section (Section C) evaluates the current status of the philosophy of mind in the light of what has gone before, and concludes that this branch of philosophy suffers, like many others, from an inability to square its theories with reality. The suggestion is made that current neurophilosophers have failed to see the extent of this problem (as have their critics) and that this failure has flawed their work, so that they have, as yet, made little significant advance upon traditional theories of mind, which rely to a greater extent on conceptual analysis. I shall also make some suggestions regarding the possibility of future successful use of the bootstrap strategy to construct theories of the mind/brain.

1. The term 'co-evolutionary' is used to describe research strategies that combine data and arguments from a variety of disciplines. Within the philosophy of the mind/brain,

it denotes those theories that synthesize philosophical, psychological and neuroscientific work. For a more detailed discussion of the term, see Chapter 6, subsection 6.0.

## Section A

### Chapter 1 - Motivation, methodologies and the veil-of-perception issue

#### 1.1 Introduction

##### 1.1.1 Sentences and synapses: motivation, evidence and methodologies

From sentence to synapse, twentieth century philosophy of mind has put forward a wide variety of hypotheses about the nature of representations and processes within the mind/brain<sup>1</sup>, many of which I shall discuss in the course of this thesis. It will be a central contention that these hypotheses and the problems that they occasion and are intended to solve, frequently result from their perpetrators' views on issues that are not, in themselves, part of philosophy. This is simply a more specific statement of the Popperian point that genuine philosophical problems frequently spring from non-philosophical roots, such as religious or scientific debates, for example. Popper<sup>2</sup>, believes that if we ignore the sources of such problems and simply study them in isolation, they will lose much of their importance, and our understanding of them will be greatly diminished. We must apply this lesson to the philosophy of mind.

Consider, by way of example, the two very different approaches of Patricia Smith Churchland and Descartes. The motivating factors that led Patricia Smith Churchland<sup>3</sup> to embark on a neurophilosophical account of mind<sup>4</sup> are very different from those that inspired Descartes to formulate his famous dualistic theory.

In the introduction to her (1986a), Churchland writes that

"[t]he guiding aim of the book is to paint in broad strokes the outline of a very general framework suited to the development of a unified theory of the mind-brain. Additionally, it aims to bestir a yen for the enrichment and excitement to be had by an interanimation of philosophy, psychology and neuroscience, or more generally, of top-down and bottom-up research" (Churchland, 1986a, pp. 3-4).

This passage makes it clear that Churchland sees the philosophical issues involved in the mind/brain relationship as closely connected with the realm of science. For Churchland, any satisfactory philosophical account of the representations and processes within the mind/brain must be formulated in the light of information yielded by neuroscientific research. Descartes, on the other hand, was greatly influenced by the religious

sentiments of his time, and hoped that his work would provide a rational justification of the belief that the soul can survive bodily death. In the *Dedication* of his *Meditations on First Philosophy*, Descartes wrote that he had

"always considered that the two questions respecting God and the Soul were the chief of those that ought to be demonstrated by philosophical rather than theological argument. For although it is quite enough for us faithful ones to accept by means of faith the fact that God exists, it certainly does not seem possible ever to persuade infidels of any religion, indeed, we may almost say, of any moral virtue, unless, to begin with, we prove these two facts by means of the natural reason" (Descartes, 1911a, p. 133).

In Descartes' view, the mind/body relationship is demonstrable by rational argument, despite the religious background to the debate. In other words, Descartes recommends that we need not take faith as our sole criterion for believing a theory to be true. Rather, we may be convinced by reason.

We must bear in mind these different motivations if we are to understand fully the issues at stake, because these motivations stimulate different assumptions and indicate different research programmes. These assumptions and research programmes in turn influence theories and arguments. However, this may become a problem if motivating factors make us seek a particular kind of answer which may result in our asking the wrong kind of questions and formulating the wrong hypotheses so that we are not only led in the wrong direction, but perhaps also prevented from seeing the situation from different or more appropriate angles. This seems to me to be an important point. The crux of this thesis is that the motivating forces of the philosophers of mind discussed in Section B may have led them into asking the wrong kinds of questions, specifically, questions which have caused them to make untestable metaphysical assumptions which impair the scientific credibility of their theories and the methodologies by which these have evolved.

I shall refer to this kind of error as 'bias' or 'evidential bias'. This term denotes the failure of philosophers to subject their hypotheses and methodologies to rigorous and impartial scrutiny. This scrutiny should include tests that are not specifically designed to confirm favoured hypotheses (perhaps by systematically excluding all potential sources of empirical evidence that might

falsify them), whilst rejecting rival views (by denying them potentially confirming sources of empirical evidence on a *a priori* grounds). Although all philosophers and scientists have a natural and ineliminable desire to see their preferred explanations triumph, this desire must not be allowed to bias the testing of their methodologies and theories. It is important to realize that my criticism is levelled at the faulty methodology that often arises from the desire to construct a successful theory, and not at the desire itself.

This is not precisely the same use of the term 'evidential bias' discussed by Paul M. Churchland in (P.M. Churchland, 1975a), although there are some important overlaps. Churchland is concerned to emphasise two ways in which our language is inherently biased. The first way, intensional bias, occurs because the meaning of our observation terms is necessarily embedded in theory and is determined very little by mere sensation. Secondly, our view of the world, and the theories presupposed by it exhibit extensional bias. That is to say, it is a contingent fact dependent upon our physical constitution that we happen to perceive and to classify the world in the way that we do. Other ways of perceiving the world, and the theories in which they are embedded may be just as valid, testable, internally simple and well-corroborated (P.M. Churchland, 1975a, p. 258). The situation then becomes one of



empirical underdetermination (see Chapter 2, subsection 2.2.2 for further discussion).

As the nature of our arguments about philosophical issues changes, so too does the kind of evidence that we require to be convinced of the validity of theories. Many different strategies may be employed in the evaluation of theories. For example, traditional philosophers of mind like Colin McGinn rely upon conceptual analysis<sup>5</sup>, whereas Patricia Smith Churchland, as we saw earlier, uses a multidisciplinary approach. In other words, as Popper rightly says, there is no such thing as the philosophical method. No method should be excluded from philosophy on a *priori* grounds.

### **1.1.2 Questions and answers: methodology or metaphysics?**

As I noted in the Introduction, one of my main concerns is to argue that the veil-of-perception issue should be revived in the form of general question about the kind of relationship between hypotheses and evidence that is required for theories to be testable. This question must be confronted by all philosophers who are concerned to develop methodologies that conform to the demands of scientific realism (see Chapter 2, subsection 2.2.2). Consequently, the veil-of-perception issue has a wider relevance than it has generally been accorded in the

history of philosophy. It should no longer be regarded primarily as a summary of the problems that face attempts to explain how we perceive the external world. Such explanations encounter difficulties precisely because they do not take adequate account of the complex relationship between hypotheses, evidence and testability. I shall say more about this relationship in the remainder of this chapter and in Chapter 2.

In the light of these considerations, I shall then proceed in Section B to an evaluation both of theories of mind that have, by and large been constructed using conceptual analysis (such as McGinn's theory of the hidden structure of consciousness in Chapter 3, and Fodor's Representational Theory of Mind in Chapter 4), and of theories that adopt a more innovative approach (the modularity theories in Chapter 5, and the theories of David Marr and Patricia Smith Churchland in Chapter 6). I shall be particularly concerned with the issue of whether the kinds of questions that these philosophers ask predetermine the kinds of explanations that they obtain. If they do, then these philosophers may be vulnerable to the charge of systematically ignoring, or more charitably, overlooking sources of evidence that call the validity of their hypotheses into question. This raises the more general question of whether scientific realists<sup>6</sup> can ever be wholly

free from metaphysical assumptions within their theories.

## 1.2 Theories, explanation and evidence

### 1.2.1 What is a theory?

There has been much dispute over the precise definition of a theory, and over whether or not a distinction can be drawn between theoretical vocabulary and observational vocabulary and between scientific theories and metaphysical theories on the other<sup>7</sup>. Broadly speaking, theories are formulated to explain or predict the nature or occurrence of events. They do this by virtue of expressing law-like statements which link types of events. Philosophers do not all view theories and their explanatory roles in the same way. For example, some instrumentalists<sup>8</sup> deny that theories present us with a true or false account of events. For this group of philosophers, theories do not purport to describe what is going on in the world; their role is merely to predict what will happen at an observational level. An adherent of realism, on the other hand, will claim that theories can be true or false, even though it may never be possible for us to know with absolute certainty that any one theory is true or false<sup>9</sup>. The realist believes that the instrumentalist view of theories

precludes explanation, and that this is unacceptable, since

"[t]here is no doubt that rightly or wrongly we want not only to be able to predict, we want to be able to explain" (Newton-Smith, 1981, p. 31).

The theories of the mind/brain discussed in Section B are (with the exception of Howard Gardner's modular theory) undoubtedly realist, since they all make claims about the ontology of the mind/brain, which they take to be true or false even if they cannot be verified, in a manner that instrumentalist theories do not. However, realism does not necessarily result in an explanatory theory. The difficulties faced by realism are outlined in Chapter 2 (subsections 2.2.1, 2.2.2).

### **1.2.2 How does a theory explain?**

A theory, being a set of one or more hypotheses, cannot be regarded as genuinely explanatory unless it provides us with additional useful information. One objection to theories that fail to provide such information has been dubbed the 'virtus dormitiva' objection, after its most famous example. This occurs in Molière's *Le Malade Imaginaire*, in which a doctor informs his patient that opium sends people to sleep because of its dormitive

powers. Clearly, since the very term 'dormitive powers' simply means 'to send to sleep', no further information is given to the patient, and no explanation of the effects of opium have been provided.<sup>10</sup>

Theories consist of hypotheses, which are backed up by evidence. The relationship between these hypotheses and the bodies of evidence that are supposed to support them forms a central issue in this thesis. For a theory to be genuinely explanatory, its hypotheses must be testable independently not only of the arguments and evidence which were employed in their formulation, but also of those arguments or events that the hypotheses are currently being used to explain.<sup>11</sup> As we shall see in a moment, the traditional veil-of-perception issue delineates a situation in which a theory may be said to lack explanatory value, precisely because its hypotheses cannot be independently tested.

So philosophers and scientists alike should aim to formulate theories that are genuinely explanatory, in the sense that they have greater content than their explananda. Moreover, their hypotheses must be readily and systematically testable in the ways just described.

### 1.2.3 The veil-of-perception issue

We have just noted the necessity of formulating theories that attempt to describe the world accurately and coherently so that our experiences are made more intelligible. It has also been emphasized that satisfactory theories are ones in which the hypotheses can be tested. For a test to have any worth, it must be possible to at least say what kind of result would falsify a hypothesis: in other words, tests that guarantee *a priori* that the hypothesis under test will be valid are not good tests.<sup>12</sup>

The veil-of-perception issue is pertinent to this discussion since it delineates a situation in which a theory fails to provide a satisfactory explanation precisely because its hypotheses cannot be tested in the manner just described.

The issue of the veil-of-perception has had a thorough airing in the history of philosophy. It is mentioned in Plato's Republic, Book vii, as a thought experiment. In it, a group of men are chained up in a cave with their backs to the light, so that all they can see are the shadows of objects passing between the light and the cave entrance as they appear upon the cave wall in front of them. They

"hold the shadows of those ... articles to be the only realities" (Plato, 1892, p.236).

The Platonic account goes on to imagine that one of the men is then released so that he can see the original objects which pass before the entrance to the cave.<sup>13</sup> The version of the veil-of-perception issue that we wish to consider here, however, caters for no such eventuality, because, of course, the very way that the situation is set up prevents any possibility of testability.

#### 1.2.4 Hume and the veil-of-perception issue

The issue is admirably restated by Hume in his *Treatise of Human Nature*, Book I, Part IV, Section ii. He writes that since

"no beings are ever present to the mind but perceptions; it follows that we may observe a conjunction or a relation of cause and effect between different perceptions, but can never observe it between perceptions and objects. 'Tis impossible, therefore, that from the existence or any of the qualities of the former, we can ever form any conclusion concerning the existence of the latter, or ever satisfy our reason in this particular" (Hume, 1978, p.212).

The traditional veil-of-perception issue arises if a theory of how we perceive reality rests upon the assumption that the following four statements are all true:

- i) there is an external reality over and above our perceptions
- ii) this reality stands in relation to our perceptions (visual, aural, tactile etc.) as cause does to effect
- iii) we can only ever have direct knowledge of our perceptions; our knowledge of external reality comes to us indirectly, mediated by these perceptions
- iv) nonetheless, our perceptions normally furnish us with a generally veridical representation of the world.

These statements fail to provide a suitable basis for a satisfactory explanation of how our perceptions yield knowledge of the world, because they form what I shall call a *semantically inconsistent tetrad*. This semantic inconsistency arises as a result of conflict between interpretations of the claims in the light of the particular evidence or assumptions upon which they are based. I shall illustrate that versions of this semantically inconsistent tetrad may be isolated within the



theories of the mind/brain examined in Section B. The difficulties identified by Hume constitute a fundamental flaw which is rooted deep in philosophical tradition.

There are two levels of inconsistency within the tetrad. Firstly, Hume argued that, since we only ever have direct knowledge of our perceptions, which we take to be representations of an outer reality (assumption iii), we can have no grounds for asserting that there is an external reality which causes these perceptions (statements i and ii). We may term this the strong or *ontological* version of the argument.

Secondly, even if we simply assume that such an external world exists, we have no reason to believe claim (iv), that is to say, that our perceptions provide us with a wholly accurate picture of it. This constitutes the weaker, *epistemological* version of the veil-of-perception issue.

Hume rightly points out that we are generally unaware of the veil-of-perception issue, since we are psychologically incapable of making a distinction between our perceptions and the world beyond them. This fact is important for our day-to-day existence, but it does not justify our ignoring the problems of explanation raised by the issue. It merely emphasises the truth of Descartes's remark that

"it is sometimes requisite in common life to follow opinions which one knows to be most uncertain exactly as though they were indisputable" (Descartes, 1911b, p. 100).

#### 1.2.5 Wider implications of the veil-of-perception issue

Those philosophers who have discussed the veil-of-perception issue have, I believe, failed to see its implications for the philosophy of science and, particularly, for the philosophy of mind<sup>14</sup>. Jonathan Bennett is a good example of a philosopher who understands the problem very well within its historical context, but who does not succeed in relating it to more general questions about the nature of evidence required for the hypotheses of a theory to be genuinely testable.

Jonathan Bennett (1971, pp. 63-70) discusses the veil-of-perception issue as it arises in Locke's *An Essay Concerning Human Understanding* (1690). Some places in Locke's *Essay* do create the impression that Locke may have fallen into precisely those problems that are delineated by Hume's discussion of the veil-of-perception issue. For example, Locke wrote that

"[ideas] are only designed to be pictures and representations in the mind of things that do exist, by

ideas of those qualities that are discoverable in them."  
(Locke, 1981, p.238)<sup>15</sup>

Later on, in Book IV, chapter iv of the *Essay*, Locke poses exactly the problem highlighted by Hume (see section 1.2.4 above). Locke comments that

"[i]t is evident the mind knows not things immediately, but only by the intervention of the ideas it has of them" (ibid, p.348).

He then goes on to ask

"[h]ow shall the mind, when it perceives nothing but its own ideas, know that they agree with things themselves?" (ibid, p. 348).

Bennett quite rightly says that

"[w]hat is wrong with the question is, precisely, that nothing could count as a legitimate argument for an affirmative answer to it." (Bennett, 1971, p.67)

Indeed, Bennett hits the nail on the head when he comments that the difficulty in postulating a causal connection between appearances and reality is that

"[t]o know that there was [such a causal connection], however, we should need independent access to empirical facts about the objective realm" (ibid, p. 70).

But Bennett's discussion of the issue becomes sterile, since he restricts his criticism of the veil-of-perception issue to the relationship between sensory states and the external world. However, many of his comments are pertinent to the wider issue, since Bennett rightly points out that empirical arguments can only provide us with limited, specific conclusions about the external world, based upon specific information from the senses, since

"[a]ny such argument turns on the fulcrum of an unquestioned acceptance of the existence of an objective world about which we know a good deal. If we stand back and try to focus on the relation between sensory states as a whole and the objective realm as a whole, asking *en bloc* whether the former are ever reliable guides to the latter, empirical arguments cannot get a grip; and the 'cannot' is a logical one" (ibid, pp. 67-8).

This passage demonstrates that no amount of empirical evidence can provide an exit from the difficulties inherent

in the veil-of-perception issue, because a reliable relationship between perception and the external world is assumed *a priori* and is therefore not empirically testable. However, there are no grounds for inferring that Bennett has realized that the traditional veil-of-perception issue results from an unsatisfactory relationship between hypotheses and the nature of the evidence available to support them - a relationship that can and does exist in theories that are not concerned with our perception of the external world. In other words, Bennett (like Locke before him) commits the sin noted by Popper (see above, subsection 1.1.1) of neglecting the broader issues at stake.<sup>16</sup>

The broader issues at stake here are clear. We need to discover if there are any strategies that will allow us to make justifiable and testable inferences from the realm of experiences to the existence of those structures and objects that both lie beyond and cause our experiences. As Raymond Tallis comments

"[e]mpirical observations may generate laws that correlate one type of experience with another; but can they take us "beneath experience" to its basis? It seems unlikely that experience can take us outside of the closed circle of experience to reveal that upon which experience -

experience in general, rather than particular experience - is based" (Tallis, 1991, p.93).<sup>17</sup>

Clark Glymour has embarked upon a search for just such an epistemological breakthrough. Glymour writes that

"whether the foundation of all empirical belief was thought to rest on beliefs about phenomenal appearances or on beliefs about observable properties of observable things, the structure of the problem was the same: what relations between statements about phenomena, or observation statements, on the one hand, and statements about material objects or, respectively, about unobservable things or unobservable properties, on the other hand, permit statements of the former kind to confirm statements of the latter kind?" (Glymour, 1980, pp.10-11).

In chapter 2, Glymour's attempt to solve this problem by employing the bootstrap strategy to ground both our knowledge of the general and of specific facts upon the same evidence will be discussed and evaluated in detail (subsections 2.3.4 - 2.4.1). Briefly, Glymour's method is to use one piece of evidence,  $E_1$  in conjunction with a hypothesis  $H_1$  to confirm a second hypothesis  $H_2$  relative to

a theory T (where T is simply a theory which consists of hypotheses  $H_1, H_2, H_3 \dots H_n$ ). Once  $H_2$  is thus confirmed, Glymour argues that we can bootstrap from  $H_2$  to confirm  $H_1$  relative to T using a second piece of evidence,  $E_2$ . Given that  $E_1$  is used to confirm  $H_2$ , which in turn confirms  $H_1$ ,  $E_1$  grounds our knowledge of both hypotheses, one of which may be specific and the other general. Chapter 2 (subsections 2.4.2 -2.6) will also describe the variety of bootstrapping found in the theories examined in Section B.

The need to solve the problems of evidence associated with the veil-of-perception issue is still a very pressing one, since failure to deal with these difficulties is endemic in twentieth century theories of mind.

### 1.2.6 The veil-of-perception issue in twentieth century psychology

This form of argument about the nature of perception can still be found in twentieth century psychology in various guises. Consequently, the metaphysical issues at stake are still very much live questions and should not be dismissed as pertinent only to seventeenth century philosophy<sup>18</sup>.

I will discuss two examples out of a number in modern psychology which stand out as being related to the veil-of-perception argument under discussion. These examples are

the views of Richard Gregory (1970) and those of Ulrich Neisser (1976). Gregory and Neisser both put forward top-down theories of perception, ie: theories that start from the assumption that what we see is predominantly determined by high-level constraints which control our interpretation of visual information.<sup>19</sup> These constraints take the form of theoretical constructions about the categories of the world in Gregory's theory, and anticipatory schemata on Neisser's view. Both of these theories are concerned with the nature of our perceptual processes rather than with the nature of the external world.

Gregory and Neisser's theories are very different from each other. Gregory's theory is a good example of the belief that perception consists chiefly in the testing of hypotheses formed about the nature of the external world as a result of previous experience. In fact, Gregory holds that we infer from our perceptions to the objects that cause them on the basis of these very hypotheses. Feature detectors within our perceptual apparatus collate the information in our retinal images and match it against stored information about objects that we have previously encountered. Perception is an active process of problem-solving, and Gregory thinks that vision provides us with an indirect characterization of the external world.



Neisser, on the other hand, rejects Gregory's model of perception as the internal processing of perceptual images according to categories. Instead, he favours a perceptual cycle model. On this view, we construct schemata which then direct our visual attention towards one object or another. This theory is equally top-down, but does not regard vision as a process of matching an inner representation with a real object.

I intend to show in the next two subsections that both these theories exhibit the difficulties highlighted in Hume's analysis of the veil-of-perception issue in subsection 1.2.4. My aim is not to criticise the views of Gregory and Neisser per se, but merely to illustrate the point that the veil-of-perception issue, and hence the questions of evidence and method, are still importantly relevant in twentieth century psychological debate.

### 1.2.7 Gregory and the veil-of-perception issue

In his (1970), Richard Gregory describes perception as the interpretation of our sensations by reference to the theoretical constructions we have of the external world and of external objects. These theoretical constructions are not wholly *a priori*, but are at least partially influenced by our previous experiences and interpretations. Thus Gregory writes that

"[g]iven the slenderest clues to the nature of surrounding objects we identify them and act not so much according to what is directly sensed, *but to what is believed*" (Gregory, 1970, p.11).

Gregory's view rests upon the acceptance of all four of the following statements:

- a) there is an external world over and above our perceptions
- b) this world stands in relation to our perceptions (visual, aural, tactile, etc.) as cause does to effect<sup>20</sup>
- c) we can never have direct knowledge of the external world; such knowledge comes to us indirectly, mediated by our perceptions, which are in turn shaped by our preconceived ideas about this world
- d) yet our perceptions provide us with essentially veridical representations of what lies beyond our senses.

Once more, we are presented with a semantically inconsistent tetrad of statements. The logical conclusion of Gregory's claim that we have no way of perceiving the

external world directly (statement c) is that we can have no grounds for asserting that there is such a world which causes our perceptions (statements a and b).<sup>21</sup> This bears a striking resemblance to the ontological version of the veil-of-perception issue delineated in the discussion of Hume (section 1.2.4 above).

Moreover, even if we do not need to have direct perception of the external world in order to believe that it exists, we still have no reason to believe that our perceptions are essentially accurate (statement d). Hence Gregory still falls victim to the epistemological version of the veil-of-perception issue (cf: section 1.2.4 above).

Two related points should be noted at this juncture. The first deals with the paucity of evidence for Gregory's claims, whilst the second considers how perceptual errors might be possible within his theory.

Firstly, I am not trying to make the positivist claim that direct perception is needed in order for us to have any justifiable belief in the existence of an external world. I am simply trying to demonstrate that the way that Gregory sets up the argument is inadequate because he provides us with no *independent evidence* either for his assertion that the external world exists (statement a), or for his claim that our perceptions are basically accurate (statement d).

Gregory writes that

"[w]e are forced ... to suppose that perception involves betting on the most probable interpretation of sensory data, in terms of the world of objects" (ibid, p. 29).

This means that we may realise that an error has been made on specific occasions. There are two ways in which this may occur. Our perceptions may conflict with our expectations of what should be seen, forcing us to look more closely and to decide whether our expectation is correct or should be revised in the light of our perceptions. Alternatively, our perceptions may conflict with each other, revealing an inconsistency that must be resolved. Indeed, Gregory comments that

"[v]ision demands that every received pattern be interpreted according to a theoretical construction of the world of external objects, and the same is true of all indirect measures in science. Both are used to suggest and test between alternative hypotheses" (ibid, p. 98).

However, this cannot count as independent evidence for the general accuracy of our perceptions, because we can only allow for the possibility of errors in individual instances

by assuming that our perceptions are, on the whole, accurate.<sup>22</sup> This means that we are simply assuming the very point that is at issue. Gregory's argument provides us with a first-class illustration of the difficulties that psychological theories may encounter when relying upon the same source of evidence to support both specific and general conclusions.

This brings me to my second point. Gregory's account of perception restricts the concept of error to a deviation from the norm, when two perceptions are inconsistent, or when our perceptions conflict with our theoretical construction of the external world. The validity of postulating the existence of such a theoretical construction in the first place and subsequent reliance on it is never questioned.

In Gregory's view, it is the very possibility of error that enables us to see how perception works. Yet his theory will only permit us to have the restricted concept of error outlined above. We can never be completely mistaken. The concept of wholesale error is missing from Gregory's agenda (see also my discussion of realism in Chapter 2, subsection 2.2).

So, Gregory's theory suffers from the difficulties associated with the veil-of-perception issue because the

premises upon which the theory rests are not supported by appropriate evidence. Nor does Gregory provide us with any suggestions as to how we might obtain any independent evidence for these premises. We therefore have no justification for regarding the inferences that the theory makes from our perceptions to the nature of the external world as accurate.

#### **1.2.8 Neisser and the veil-of-perception issue**

It is clear from our discussion of Gregory that psychological theories may encounter problems if they rely upon the same evidence when they are addressing individual instances as they used when building the general outline of their theory<sup>23</sup>. Ulrich Neisser's theory of perception is vulnerable to the same problems, although he denies that perception is carried out by means of inferences about the world from retinal images. That is to say, perception is not just phenomenalism. Neisser writes that

"[perception] need not be organized in terms of momentary retinal 'snapshots' at all, and the similarity or dissimilarity between perceived objects and their projected images is irrelevant to it" (Neisser, 1976, p. 16).

Nevertheless, Neisser's rejection of the retinal image theory of perception does not preclude his adherence to a theory of perception that is at least partially top-down. His theory rests upon the concept of what he calls the perceptual cycle. There are no retinal images inside the head; rather

"the cognitive structures crucial for vision are the anticipatory schemata that prepare the perceiver to accept certain kinds of information rather than others and thus control the activity of looking" (ibid, p. 20).

Perception is only possible in virtue of these anticipatory schemata: the precise schemata that each person has determine what he sees, the nature of the errors he is likely to make, and his interpretations of the world around him. We are able to amend our schemata in the light of what we actually perceive, but our expectations play a large and decisive role in determining which of two conflicting schemata should be retained and which rejected. Neisser comments that

"[i]f the environment is rich enough to support more than one alternative view (and it usually is), expectations can have cumulative effects on what is perceived that are virtually irreversible

until the environment itself changes" (ibid, p. 44).

Neisser believes that his theory has distinct advantages over those theories that involve inference from retinal images or perceptual data<sup>24</sup>. This is because he allows for the possibility that our perceptual schemata may be modified by experience, so that the environment does influence what we see. He writes that

"[w]hen a perceptual cycle is carried out normally, schemata quickly attune themselves to the information actually available. Perception is veridical" (ibid, p. 43).

However, it is not at all clear that Neisser can justifiably claim these advantages for his theory. Neisser takes the following four statements to be true:

I) there is an external reality over and above our perceptual schemata

II) this reality can have a causal effect on our perceptual schemata (visual, aural, tactile, etc.)

III) it is impossible for us to perceive this external reality except by virtue of our perceptual



schemata, which determine what we anticipate perceiving and how we set about exploring the external world

IV) our perceptual schemata provide us with a veridical view of this reality (assuming that we are mobile organisms in a mobile world).

Once again we are confronted with a modern psychological theory which can be summarized in the form of a semantically inconsistent tetrad. If we can only ever gain a purchase on the nature of external reality through our perceptual schemata (statement III), then we have no independent grounds for asserting that such an external reality exists over and above these schemata (statement I), let alone that this reality can have a causal effect upon our schemata (claim II). Neisser cannot therefore meet the challenge presented by the ontological form of the veil-of-perception issue any better than Gregory.

Furthermore, the assumption that such an external reality does exist does not relieve Neisser of the obligation to demonstrate that his theory does not fall foul of the epistemological version of the veil-of-perception problem. Neisser declares that

"[w]hat is seen depends on how the observer allocates his attention: ie., on the anticipations he develops and the perceptual explorations he carries out" (ibid, p. 39).

This puts Neisser's theory on a par with that of Gregory. He may define error in terms of a conflict between what we expect to see and what we actually see, but he cannot justifiably assert with any confidence which of our perceptions are correct and which are incorrect (claim IV). Any of our perceptual schemata may be completely mistaken, but this possibility is systematically ignored by Neisser. Like Gregory, he can only account for the possibility of specific errors (in terms of perceptions which conflict with the anticipations embodied in our perceptual schemata); wholesale error lies completely outside the scope of his arguments.

### 1.3 Summary of the major points of chapter 1 and their implications for the philosophy of mind in the twentieth century

The theories of Richard Gregory and Ulrich Neisser both illustrate my point that, even in the twentieth century, psychologists and philosophers frequently fail to recognise the need for independent evidential paths in

support of the hypotheses that constitute their theories (see my discussion of bootstrapping in Chapter 2, subsection 2.3.1). It is for precisely these reasons that neither Gregory nor Neisser can escape the same kinds of difficulties as those that Hume identified within the veil-of-perception issue almost three centuries ago.

In section B, I shall examine the views of Colin McGinn, Jerry Fodor, the modularity theorists, David Marr and Patricia Smith Churchland to see whether they exhibit the same kinds of problems concerning the nature of the evidence that they evince in support of their theories. It is clear from my discussion of Gregory and Neisser that, if the philosophers under scrutiny in section B do suffer from these problems, they are not unique in their plight. However, failure to provide independent evidential paths for the hypotheses within their theories would indicate that the arguments of Marr and Churchland do not represent the significant advance over traditional theories of mind that their proponents take them to be.

Consideration of the veil-of-perception issue, properly understood, forces us to examine the testability of our theories, that is, whether their hypotheses are sufficiently supported by independently obtained sources of evidence. It therefore raises the issue of realism.

Is our claim to have accurate knowledge of the existence of an external world which exists over and above our perceptions a justifiable statement which is open to revision in the light of fresh evidence, or is it merely an assumption upon which all scientific activity is based, but which is itself untestable? This question will be discussed at length in chapter 2. If it turns out that the philosophers examined in section B take the second view, then their theories might be beset by the sort of metaphysical presuppositions that they explicitly seek to avoid.

1. None of the theorists discussed in this thesis are dualists, but since some of them (for example: McGinn, Fodor) place a greater emphasis upon traditional concepts of the mental than do others (such as Patricia Smith Churchland), I have elected to use the neutral term 'mind/brain' throughout.

2. See, for example, Popper, 1969, chapter 2.

3. For the sake of brevity I shall refer to Patricia Smith Churchland as 'Churchland' throughout. Where Paul M. Churchland is discussed, I shall state his name in full.

4. Neurophilosophy attempts to cast light on the mind/brain relationship by revising traditional philosophical methods to include a consideration of the methods and results of neuroscientific research. It is a term used by Churchland to refer to her own version of co-evolutionary philosophy of the mind/brain. For further discussion, please see Section B, Chapter 6, subsections 6.2 - 6.2.5. I shall use the generic term 'neuroscientific' to refer to any research into the structure and function of the brain. For further discussion of terminology, see P.S. Churchland (1986a), pp. 153-4.
5. See chapter 3 for a brief discussion of the problems associated with McGinn's approach to the mind/brain question.
6. Scientific realism and metaphysical realism will be discussed in Chapter 2, subsections 2.2.1 and 2.2.2.
7. See, for example, W. H. Newton-Smith (1981), Chapter II.
8. Newton-Smith (ibid, p. 30) refers to this class of instrumentalists as "semantical instrumentalists".
9. See also my discussion of realism below, Chapter 2, subsection 2.2
10. A prominent example of a theory of the mind/brain that falls foul of this objection is Fodor's Representational Theory of Mind. See Chapter 4, subsection 4.4.2, below. The 'virtus dormitiva' objection is also discussed in Dennett (1981), pp.56-9.
11. See, for example, Salmon (1989).
12. For an excellent discussion of the fundamentals, see Giere (1983), Popper (1969, eg: p. 249).
13. Thus, the Platonic version of the veil-of-perception issue would, in more general terms, describe a situation in which hypotheses could not in point of fact be tested, but in which the logical possibility of eventual testability would remain open.
14. The views of David Papineau (1987), who believes that the issue fails to cast any light on representational theories of the mind/brain, will be discussed in context in the Introduction to Section B, subsection B2.
15. Note that no part of my argument depends crucially on whether Locke actually held such a view.
16. However, it should be said in Bennett's defence that he is engaged in a critique of the theories of Locke, Berkeley and Hume, and not in a discussion of the criteria for satisfactory explanation required within the philosophy of science.

17. This is simply another, more general way of stating the relationship between Bennett's conclusions about the relationship between perception and the external world, cited above.

18. This point notwithstanding, there are still some important lessons to be learned both by examining criticisms of philosophers of the seventeenth century who may have held this view, such as John Locke, and by drawing comparisons between him and modern philosophers of mind - see later on, chapter 3, subsection 3.2.6.

19. For further discussion of the differences between top-down and bottom theories of the mind/brain, see Chapter 6, particularly subsections 6.1.1 and 6.2.1.

20. This statement is true by virtue of the fact that Gregory would certainly want to maintain that if the external world did not exist, then we would definitely be aware of that, even though he also holds that our perceptions are to some extent influenced by our beliefs about what the external world is actually like. This conflict is at the very centre of the inconsistency of this tetrad. Of course, Gregory assumes the truth of current physics and physiology in his analysis.

21. I am not suggesting that Gregory would endorse this conclusion, merely that it is a logical consequence of the flawed relationship between his hypotheses and the evidence that he cites in support of them.

22. cf: the quotations from Bennett and Tallis, cited above, subsection 1.2.5.

23. Philosophers sometimes rely on the methodology known as bootstrapping as a way of solving precisely this problem. See chapter 2 for an analysis of its contribution to this field.

24. I think that Neisser would regard both John Locke and Richard Gregory as holders of such theories: Locke because he is traditionally thought of as propounding a representative theory of perception, and Gregory because he upholds Helmholtz's view that perception consists of making unconscious inferences from sensations to causes.

## Chapter 2 - Realism, Bootstraps and Bias: Problems of Evidence in the Philosophy of Science

### 2.1 Introduction

In the course of chapter 1, it was emphasised that one of the main aims of philosophy (and, indeed, of science) is to formulate theories that are genuinely explanatory. However, we have seen that philosophers are liable to encounter difficulties if the hypotheses that make up their theories are not sufficiently supported by independent evidence. It was also stressed that these difficulties become strikingly apparent when we attempt to go beyond direct observation and draw conclusions about the nature of a world beyond our perceptions. Moreover, it was demonstrated that the veil-of-perception issue as it is delineated in Hume (1978) may be regarded as an illustration of precisely this scenario. Consequently, the four assumptions which form the backbone of theories which fall into the trap of the veil-of-perception issue were isolated. I was therefore able to show that the veil-of-perception issue occurs as the result of the presence of a semantically inconsistent tetrad of assumptions.

Finally, since the perceptual theories of both Gregory (1970) and Neisser (1976) are founded upon variations of

this tetrad, I concluded that the problem of the relationship between evidence and hypothesis is still a live issue for twentieth century philosophy of mind. The problem raises the entire issue of realism, because it is important to determine what attitude philosophers should take to the hypothesis that there is a world beyond our perceptions, and what role they think such a hypothesis should play in the formulation of their theories.

The following two sub-sections will discuss two different kinds of realism, metaphysical realism and scientific realism. The scientific realist's claim that he is better able to deal with the problems raised by the veil-of-perception issue using a version of the bootstrapping methodology introduced in chapter 1 (subsection 1.2.5) will then be evaluated in a more detailed discussion of the methodology itself and of the forms that it may take (subsections 2.3 - 2.6). I shall be arguing that use of the bootstrap methodology does not necessarily prevent those philosophers who would claim to be scientific realists from postulating theories which turn out to be laden with (often implicit, rather than explicit) untestable metaphysical assumptions. Consequently, they will failed to avoid the difficulties highlighted by the veil-of-perception issue of grounding their theories in testable hypotheses. I shall contend that, although a form of bootstrapping is used by the philosophers discussed in



Section B, co-evolutionary theories<sup>1</sup> within the philosophy of mind are particularly unsuitable candidates for bootstrapping.

## 2.2 Two kinds of realism

### 2.2.1 Metaphysical realism and the veil-of-perception issue

Adherents of metaphysical realism take it to be a prerequisite of scientific activity that the world exists more or less as we perceive it. They presuppose a correspondence theory of truth such that our perceptions are only true if they are by and large accurate representations of the way that the world really is<sup>2</sup>. But this relationship cannot be independently tested, since our only knowledge of the external world is derived from our perceptions, and this begs the question of the accuracy of these perceptions. So metaphysical realism precludes the very possibility of there being any scientific evidence for the existence of such a world. As a result, theories based upon metaphysical realism cannot be regarded as satisfactory explanations. As Ellis comments

"the postulated causes of the phenomena must be supposed to exist if the theory is to be accepted as doing what it purports to do; and normally we should expect to be able to find

independent confirmation of their existence from various sources" (Ellis, 1985, p. 57).

Even if we could grant that such a transcendental world existed, thus dismissing one of Hume's concerns in his formulation of the problem of the veil-of-perception (which I termed the "strong" version of the veil-of-perception problem in subsection 1.2.4), the metaphysical realists would still have difficulty in satisfying us that our perceptions and representations of this world are accurate. In other words, they would still be faced with the traditional sceptical challenge. As Ellis writes

"[w]e can investigate nature and develop a theoretical understanding of the world, but we cannot compare what we think we know with the truth to see how well we are doing. We cannot even be assured that science has made progress toward its goal of discovering the true nature of reality" (ibid, p.69).

As Ellis points out, this is not a merely contingent, state-of-the-art difficulty, but a necessary consequence of the metaphysical relationship between the world and our scientific theories about it. He continues

"[i]f even the perfection of human knowledge by

human standards does not necessarily lead to truth, then the truth is essentially unknowable" (ibid, pp. 71-2).

The very manner in which metaphysical realists frame their questions about the world precludes discoverable empirical facts from constituting independent evidence in favour of their theories.

### 2.2.2 Scientific realism, the veil-of-perception issue and the problem of evidence

The second strand of realism under discussion is generally termed internal or scientific realism. Internal realists are just as committed as metaphysical realists to the existence of a materialist<sup>3</sup> ontology, but they do not seek to justify the existence of an external world by reference to a correspondence theory of truth.

This is not to say that scientific realists do not think that they have any evidence for the existence of an external world. It is simply to say that the way that they set up the problem both allows for and demands evidence that is not available to the metaphysical realist. Scientific realists justify their acceptance of a materialist ontology using the argument from best explanation. This argument is in itself a scientific

hypothesis which states that the best explanation of why scientific theories are successful is that the world really is as they postulate it. Hence the internal realists derive increased support for the hypothesis that the world exists through the existence of successful scientific hypotheses. If a theory fails to make successful predictions or to explain phenomena it is intended to explain, then the theory should be discarded, or at least revised.<sup>4</sup> As we shall see later on in this chapter, this is the basis of the bootstrapping methodology. Here, the scientific realist puts forward the relatively unconfirmed hypothesis that the world exists beyond our perceptions (together with some tentative pieces of evidence) as a foundation upon which to build other, more specific scientific hypotheses. If these hypotheses are then well supported by independently obtained evidence, then they, in turn, provide further confirmation of the original hypothesis that the world exists beyond our perceptions.

Scientific realists regard their individual theories and the relationship of these theories with the world as empirically testable. Yet it is sometimes argued<sup>5</sup> that all scientific theories are empirically underdetermined, since no matter how many of their consequences we examine, it will always be possible to construct rival theories which contradict them and yet succeed in explaining exactly the

same phenomena. If this argument is valid, then the internal realist is no less vulnerable than the metaphysical realist to the difficulties encompassed by the veil-of-perception issue. If scientific theory A' is as good at explaining our experiences as scientific theory A and is backed by equally strong evidence, then the internal realist has no reason for believing that A provides a more accurate representation of the world than does A'. Consequently, he falls victim to the weak version of the problems outlined in my discussion of the veil-of-perception issue (in subsection 1.2.4).

Internal realists claim to have a reply to the argument from empirical underdetermination. They maintain that such an argument can never be justified, since it is impossible to predict future theoretical developments which may yet enable us to distinguish between two theories that are currently empirically equivalent. Such evidence may not come from any direct consequences of either theory. Ellis writes in support of this claim that

"the point is a Duhemian one. Theories do not normally occur in isolation, and evidence for or against a theory can come from unexpected quarters" (ibid p.65).

Evidence for or against the truth of a theory may include

"values such as ontological simplicity, coherence, and explanatory power" (P. M. Churchland, 1985, p.42).

As Paul Churchland rightly points out, these criteria are values and not straightforwardly observable kinds of evidence. Nor are the criteria mentioned here an exhaustive list of the values that it is possible to employ in the assessment of a theory. The drawbacks of Paul Churchland's view will be analyzed later on in this chapter in the course of the discussion of Clark Glymour, who holds similar opinions.

Hence, the internal realists' criterion for the truth is more than one of correspondence to a transcendental world, and a rather more pragmatic one at that. Whilst they would not wish to claim that the true theory is simply equivalent to the best theory we have, they do not insist upon the separation of truth from epistemic values, such as rationality, as the metaphysical realists do. However, this can lead to parochial entrenchment within one's own theoretical preferences. What it is rational for one person to believe may not seem rational to another, or indeed, to a Martian!<sup>6</sup>

However, the success of the scientific realist in escaping the difficulties of metaphysical realism is

debatable since evaluative criteria like ontological simplicity and coherence may themselves incorporate implicit and untestable metaphysical assumptions. Clark Glymour writes that

"scientists holding contrasting theories are bound to see the same evidence as having a different bearing, even if they are in full agreement as to the evidence itself" (Glymour, 1980, p.121).

Indeed, as the discussion of Patricia Smith Churchland and Descartes in chapter 1 showed, the questions that philosophers ask are very much influenced by their preconceptions of the issues that they want to resolve. Moreover, if these questions are also phrased in such a manner that they filter out responses and evidence that would count against their theories, then they are as guilty of making unjustifiable assumptions as the metaphysical realists that they are so eager to criticize. Once more, philosophers are confronted with the problem of evidential bias as defined in Chapter 1, subsection 1.1.1. In Section B, the consequences of this argument for the mind/brain question will be discussed.

### 2.3 Scientific realism and the bootstrapping device

It has already been explained (chapter 1, subsection 1.2.5, chapter 2, subsection 2.2.2) that the bootstrapping methodology has been regarded by scientific realists as a tool for deriving both specific and general views from the same initial evidence.

This option is not open to the metaphysical realist, who assumes that the external world exists, but who cannot test his hypothesis because it takes the status of an *a priori* assumption. Similarly, we saw in chapter 1 that Gregory and Neisser were only able to account for individual cases in which our perceptions may mislead us using the assumption that our perceptions are generally reliable. They were unable to use the evidence presented by the senses to acquire both specific and general knowledge about the external world.

Perhaps the best introduction to the topic of bootstrapping is given by Clark Glymour in his book *Theory and Evidence* (Glymour, 1980). It is essential to note from the outset that bootstrapping is a very complex methodology that can appear in a variety of different guises. Since it is itself merely a suggestion or model of how it may be possible to obtain support for general and specific hypotheses from the same initial evidence, it is hardly



surprising that individual accounts and uses of the methodology will emphasize different aspects of the strategy and a variety of ways in which it is employed.<sup>7</sup>

Moreover, since the strategy is a device for acquiring new knowledge on the basis of sound evidence, it blurs the traditional distinction between discovery and justification.

I shall begin by discussing the methodology as described by Glymour, prior to isolating the key features of the strategy and illustrating how they recur in the form of bootstrapping which better describes the arguments of the philosophers of mind discussed in Section B.

Glymour regards his own variation of the bootstrap strategy as deriving from theories of Reichenbach and Carnap. Both Reichenbach and Carnap sought a way of using evidence derived from our observations to confirm the individual hypotheses of a theory. Glymour writes that the Carnapian legacy within Glymourian theory is

"a stratagem for making the connection between evidence and theory: use some of the hypotheses to deduce from the evidence statements, instances of other hypotheses" (Glymour, 1980, p. 62).<sup>8</sup>

However, Reichenbach and Carnap both held that such deductions were only possible if they were based on premises which incorporated both observational and theoretical predicates<sup>9</sup>. Such premises are, rightly, rejected by Glymour because they are regarded by Reichenbach and Carnap as meaning postulates and have not been tested. Glymour comments that such premises

"have a special status; they are not understood as hypotheses within the language but rather as *meaning postulates* that extend the language. Their truth is entirely stipulative" (Glymour, 1980, p. 59)

It is clear that philosophers who base their theories so firmly upon such premises cannot avoid the charge of drawing conclusions from indefensible assumptions any more successfully than the metaphysical realist. However, the bootstrap methodology need not necessarily rely upon such premises.

### **2.3.1 The Glymourian bootstrap method**

The following passage from Glymour's book encapsulates the essence of the bootstrap strategy.

"Hypotheses are tested and confirmed by producing

instances of them; to produce instances of theoretical hypotheses one must use other theoretical relations to determine values for theoretical quantities; these other theoretical relations are tested in turn in the same way. Ideally, one might hope for bodies of evidence that permit each hypotheses to be tested independently" (ibid, p. 52).

For Glymour, evidence does not confirm a hypothesis absolutely, but only relative to a theory. We proceed, in the first instance, by using a body of evidence  $E_1$  together with an unconfirmed hypothesis  $H_1$  to confirm a hypothesis  $H_2$  relative to a theory  $T$  (where  $T$  is simply a set of hypotheses  $H_1 + H_2 + H_3 + \dots + H_n$ , together with their supporting bodies of evidence,  $E_1, E_2$ , etc). Thus far, there is nothing distinctive about Glymour's procedure, except the emphasis upon the fact that  $E_1$  is only evidence for  $H_2$  relative to  $T$ . Nor must it be possible for any evidence whatsoever to confirm  $H_2$ . What is unique about Glymour's bootstrap method is that it can then be inverted so that  $H_2$  and a fresh set of evidence  $E_2$  can be used to confirm  $H_1$  relative to  $T$ <sup>10</sup>.

In other words, Glymour's strategy enables us to pull ourselves up by the bootstraps by using relatively unconfirmed hypotheses to yield fresh information in the

form of further hypotheses, which may then in turn be combined with new evidence to confirm the original hypotheses. In more general terms, an overall world view may be combined with observational evidence to confirm a more narrow hypothesis relative to a theory; the narrow hypothesis can then be used in conjunction with fresh evidence to confirm the general hypothesis, again relative to the theory. In this way, Glymour might contend, we are able to tear down the veil-of-perception.

Two examples follow of how the Glymourian bootstrap methodology may be used. The first is a rather informal treatment of the use of the strategy in the formulation of a theory about genius in composers. The second is more formal. It consists of a justification of induction based upon the bootstrap strategy.

#### Example 1 - Genius in composers

Let  $T$  be the entire theory of the nature of genius in classical composers.

Let  $H_1$  be the unconfirmed hypothesis that all musical geni come from mainland Europe and  $E_1$  be knowledge of the ages at which those mainland European composers familiar to us first began to show signs of musical promise. So  $E_1$  might consist of:

Mozart - gave public performances aged 6

Bach - showed early promise

Strauss (Richard) - composed from the age of 8

Liszt - played in public aged 9

Chopin - played in public aged 8<sup>11</sup>

From this, we may deduce the hypothesis  $H_2$  that musical genius often begins at an early age. If we then use this hypothesis  $H_2$  together with the details of the birthplaces of those youthful composers familiar to us ( $E_2$ ), we can try to confirm the first hypothesis  $H_1$  that all musical genius comes from mainland Europe. For example,  $E_2$  may consist of the following data:

Mozart was born in Austria

Handel was born in Germany

Bach was born in Germany

Strauss was born in Germany

Liszt was born in Hungary

Chopin was born in Poland

Given that  $E_2$  does contain this information, then it would appear that  $H_1$  is confirmed relative to the theory  $T$ . However, for the bootstrap strategy to avoid circularity, we must be able to say what sort of evidence would count against  $H_1$ . For example, if Purcell were to be included in  $E_2$ , then  $H_1$  would be falsified since Purcell was born in England.  $H_1$  would then need to be weakened to the

hypothesis that "many musical geni come from mainland Europe". Similarly, we must be able to say what evidence  $E_1$  would have to include for it to falsify hypothesis  $H_2$  - in this case, had we included Beethoven, who did not show much musical promise in his youth (despite his father's attempts to make it seem otherwise),  $H_2$  would also need revision.

Note that we could now go on to use  $H_2$  to formulate either another general hypothesis,  $H_3$ , such as all musical geni played the piano, or a more specific hypothesis  $H_4$ , such as, if Purcell was a musical genius, then it is likely that he composed pieces at an early age. ( $H_4$  is confirmed by further evidence  $E_4$  - Purcell did begin to compose when still a young boy.) In this fashion, our initial hypothesis  $H_1$  and evidence  $E_1$  permit inference of both specific and general facts.

#### Example 2 - The use of the bootstrap strategy to justify induction

A more formal example of the use of the bootstrap method in science is given in Hunt's (1990) examination of the question of induction. Hunt argues that the fact that only the existence of regularity can provide evidence for the view that there are necessary connections within our universe does not entitle Hume to infer that there can be

no empirical evidence for necessity. The problem of necessity simply reduces to the problem of induction. If we can empirically justify induction, then necessity exists, even if we cannot know the precise nature of this necessity. This is where the bootstrap strategy comes in.

Although Hume can argue that inductive argument cannot guarantee the existence of a uniform world, it is not open to him to deny *the very possibility* that such a world exists. Call this possibility our theory T. We can assign this possibility a finite positive probability. This probability need not be very high. It therefore follows that we can assign all the hypotheses included in T ( $H_1, H_2, H_3 \dots H_n$ ) an equal probability of  $> 0$ .

If induction is valid (call this hypothesis  $H_1$ ), then it will tend to yield true hypotheses on the basis of the evidence available. So if the probability of T being true is  $> 0$ , then the probability of induction being valid is also  $> 0$ . If evidence  $E_1$  is available which increases the probability of a second hypothesis  $H_2$  being true, given that the probability of induction being valid is  $> 0$ , then this in turn increases the possibility that nature is uniform and that induction is indeed valid.

Once more, we are pulling ourselves up by the bootstraps since we are assuming that nature may be uniform and that

induction may be valid in order to demonstrate these very conclusions. However, Hunt's argument is not circular because

"even though an inductive rule appears both as a rule of inference and as a conclusion, it is nowhere assumed as true but only, initially, as minimally probable" (Hunt, 1990, p. 243).

Moreover, we can say what evidence would count against it.

Hunt comments that

"[t]he emergence of evidence supporting induction is an empirical matter and if contrary evidence appears the probability of [induction being valid] will become low" (ibid, p. 243).



## 2.4 Evaluation of the bootstrap strategy - part I

Hunt's example shows how just how deep-rooted and useful the bootstrap strategy may be if it can withstand the criticisms that have been levelled at the metaphysical realist with respect to the testability of the premises of his theories.

Just how satisfactory is this method of formulating and testing hypotheses? As we have seen in our examples, it is possible to avoid the obvious objection of circularity. Glymour argues (p. 108) that the objection is not pertinent here, because we can restrict the evidence to confirm any hypothesis relative to the theory so that we are able to say at every stage of the process what would falsify the hypothesis. He writes that

"[t]hen it could not be objected that the hypothesis would be confirmed whatever the evidence, and it could not be objected that anything confirms anything else with respect to any theory whatsoever" (Glymour, *ibid*, p.108).

So Glymour seems to do better than the metaphysical realist, at least as far as the strong version of the veil-of-perception issue is concerned. If new evidence calls one of his hypotheses into doubt, then Glymour is able to

reject either all or part of his theory. Compare this with the metaphysical realist who is unable to say what evidence would refute his theory that the external world exists, since this theory is a prerequisite for the rest of his beliefs, and is, as we saw earlier in this chapter (subsection 2.2.1), not testable.

But if hypotheses are only ever justified by evidence relative to a theory, how does Glymour fare against the problem of empirical underdetermination (see earlier, subsection 2.2.2)? What happens if there is more than one possible explanation for the observational evidence we have, so that the evidence supports rival theories in different ways?

As we noted, this scenario may still cause difficulties for the scientific realist. There is little point in granting the existence of an external world if we cannot say which of our theories is a correct description of it. The scientific realist will not want to back the theory that merely saves the phenomena, rather than the one provides the correct explanation. Glymour comments that

"[w]hat makes one theory better than another is a diffuse matter that is not neatly or appropriately measured by any single scale" (Glymour, *ibid*, p. 153).

Nevertheless, as we noted above (subsection 2.2.2), Glymour believes that use of the bootstrap strategy does yield a few general criteria for deciding between theories. Such criteria include (ibid, pp. 153-4) preferring those theories of which the hypotheses are confirmed by the evidence, rather than those containing disconfirmed hypotheses, those that have been tested against a wider variety of evidence, those that explain evidence in a uniform manner, etc. As Glymour himself writes

"[t]he criteria correspond to familiar methodological truisms, and, moreover, they are all aspects of the demand for better tested theories" (Glymour, ibid, p.154).

These criteria are very similar to those mentioned in the discussion of scientific realism earlier in this chapter (subsection 2.2.2)<sup>12</sup>.

However, I believe that the scientific realists have not won the day against the metaphysical realists simply by citing such criteria, since the criteria that they actually invoke in deciding which theory to adopt may in fact be far more metaphysically invidious. I state my case in subsection 2.5.2 of this chapter.

### 2.4.1 Achinstein's criticisms against Glymour and their implications

Peter Achinstein has put forward another criticism of Glymour's theory, which is worthy of consideration. In Achinstein (1983), he argues that Glymour misconceives entirely the desired relationship between theory and evidence. This is because Glymour's concept of evidence cannot account for the role that background information generally plays in our decisions as to whether it is reasonable to believe a theory or not. Achinstein comments that Glymour's bootstrap strategy does not acknowledge one of the basic requirements of evidence, namely that it should increase the probability of the hypothesis that it is intended to support being true (Achinstein, 1983, p.358)<sup>13</sup>.

Nor is Glymour's cause helped any, Achinstein argues, by his (Glymour's) declaration that evidence ( $E_1$ ) is only evidence in support of a hypothesis ( $H$ ) relative to one particular theory ( $T_1$ ), and not necessarily relative to any other similar theory ( $T_2$ ). For the logical consequence of this view is that we can only regard evidence ( $E_1$ ) as providing genuine support for  $H$  if we already know that the theory ( $T_1$ ) to which the truth of  $H$  is relativized is itself true. Yet, given Glymour's concept of evidence, we cannot derive any reasons for believing  $T_1$  to be true from the

evidence evinced in support of the other hypotheses which, together with H, make up  $T_1$ . Achinstein writes that

"[a]lthough I know that (in Glymour's sense)...[E]...is evidence for [the hypotheses that make up  $T_1$ ] with respect to [ $T_1$ ], this fact does not enable me to know that [ $T_1$ ] is true, Glymour will probably say. Nor, indeed, does this fact give me a reason to believe [ $T_1$ ]" (Achinstein, 1983, p. 361)<sup>14</sup>.

Achinstein implies (ibid, p. 361) that it is because of this flaw in the Glymourian concept of evidence that Glymour relies upon the general criteria mentioned earlier (subsection 2.4) as tools to decide which is the best theory from a number of rivals. However, Achinstein rightly comments that even

"if one theory is better than another according to Glymour's criteria for comparing theories, this does not mean that the former is known to be true whereas the latter is not; nor does it mean that there is a good reason to believe the former to be true and the latter false" (ibid, p. 361)

The implications of Achinstein's arguments are very clear. If my conclusions in subsection 2.5.2 (below) are

correct, then philosophers in co-evolutionary disciplines<sup>15</sup> like the philosophy of mind should be wary of employing general evaluative criteria when they are comparing their own theories with those of their rivals, since this strategy may lead to the temptation to assess the theories against yardsticks that favour their own theories. Moreover, Achinstein has shown that this mode of evaluation can tell us nothing about the truth or falsity of the theory that best satisfies such abstract criteria. This means that the scientific realist's position would be dangerously close to collapsing into an instrumentalist stance (cf. chapter 1, subsection 1.2.1).

The instrumentalist holds that it is meaningless to apply the terms "true" and "false" to theories, since theories are merely devices for predicting what will happen in the future. They do not describe anything. Consequently, the problem of empirical underdetermination rears its ugly head once more. If theories only predict, then it is perfectly plausible for two theories to predict events equally successfully. Only one, however, will be the true explanation.

Clearly, it would be disastrous for the scientific realist to lose his grip upon the concepts of truth and falsity. It is perfectly correct to acknowledge that all theories should be testable and liable to revision should

hypothesis be falsified by the evidence, for this strategy cannot be applied without an understanding of what it means for a theory to be true or false. However, it is quite another thing to employ a method for evaluating theories that can have no bearing at all on whether they should be regarded as true or false. It is hard to see how the scientific realist could maintain that his theories have any more explanatory value than the instrumentalist's predictions.

Achinstein's arguments also demonstrate that it is not automatically open to scientific realists who adopt the Glymourian bootstrap strategy to claim that they can indeed be sure that the resulting theories have some correspondence with external reality simply because they are supported by bodies of evidence. Glymour's theory of evidence makes no mention of the requirement that evidence in support of a hypothesis should a) make the truth of that hypothesis more probable by means of an explanatory connection between the two (which demonstrates that the evidence can be explained by the hypothesis) and b) take background information into account. This means that it is perfectly possible for an opponent to accuse the Glymourian scientific realist of selecting evidence which appears to fit in with his hypotheses even though the hypothesis cannot really explain the evidence.

A second accusation may also be justifiable, namely that the scientific realist is using Glymour's strategy to escape from the need to take relevant background information into account, including, perhaps, evidence that might tell against his chosen theory. If this accusation is correct, then the scientific realist would be no less guilty of the metaphysical realist of rendering his theory immune from testability. In seventeenth century terminology, he would simply be hiding evidence against his theory behind the veil of perception, in the hope that critics of his work will not be discerning enough to notice this sleight of hand.

#### 2.4.2 The outlook for philosophers of mind who bootstrap - Part I

I have already noted (subsection 2.3) that the bootstrap strategy may be employed in slightly different forms, depending upon precisely how it is thought by individual philosophers to provide an explanation of our ability to support general and specific hypotheses upon the basis of the same initial evidence.

I shall outline the bootstrap strategies of the philosophers of mind discussed in this thesis in greater detail in Section B. Here, I shall confine myself to a few general remarks.



It is worth noting that the philosophers discussed all incorporate several key features of the Glymourian account into their bootstrapping strategies. Firstly, they all perceive the bootstrap strategy as a way of formulating general and specific hypotheses upon the same initial evidence, thereby allowing a structured search for further information. Secondly, they all claim to be aware of the importance of the possibility of evidence that will refute either one of their hypotheses, if they are to avoid the charge that any evidence whatsoever will support their hypotheses, thus effectively rendering them untestable. Like Glymour, they also try to provide at least two independently obtained sources of evidence in support of their hypotheses. This is intended to make the hypotheses independently testable and to allow for the possibility that one hypothesis may be accepted, and the other rejected if there is not sufficient relatively independent evidence to support it. For example, they may accept the general hypothesis and reject the more specific one, preferring to find a second specific hypothesis that is better supported by independently obtained evidence.

However, there are also divergences from the Glymourian outline of the methodology. Philosophers of mind are concerned with formulating general theories of what the mind/brain is really like. They each proceed by formulating a general hypothesis about the nature of

processes and representations within the mind/brain ( $H_1$ ) and citing evidence ( $E_1$ ) which consists in facts that are best explained by such a hypothesis. They then postulate specific hypotheses ( $H_2$ ,  $H_3$ , etc) about the nature of the mind/brain and how it should be studied on the basis of  $H_1$  and  $E_1$ . If further confirmation of  $H_2$  can be found in the form of an independently obtained body of evidence ( $E_2$ ) then  $H_2$  is supported, and  $H_1$  may also be regarded as more likely to be correct. If, however,  $E_2$  contradicts  $H_2$ , then support for both hypotheses is weakened, and further confirmation for  $H_1$  must be sought elsewhere.<sup>16</sup>

It might be argued that this version of the bootstrap strategy avoids the Glymourian problem noted by Achinstein (see subsection 2.4.1) that evidence cannot provide support for a theory unless we presume the theory to be true, because the evidence does not make a hypothesis appear more or less likely to be true. If philosophers of mind cite some support for  $H_1$  in the form of  $E_1$ , then this difficulty seems to be avoided. However, I shall demonstrate in Section B that the philosophers of mind do not succeed in avoiding these problems any better than Glymour, since the evidence that they cite in support of their hypotheses either systematically excludes sources of possible empirical refutation, or is merely compatible with the hypotheses, rather than genuinely explained by them. This can result, once more, in the possibility of empirical

underdetermination.

In Section B, the theories of modern philosophers of mind will be examined to see if they fall into the traps of the bootstrap strategy outlined thus far. It seems that bootstrapping philosophers of mind may be caught on the horns of a three-way dilemma. Either

a) they will maintain that their evidence provides us with good reason to believe that their theory is true: this is debatable in the light of my earlier comments, or

b) they will adopt the more general evaluative criteria for judging their theories, in which case they will either  
 i) run the risk of reducing their theories to the status of complex predictions (the accusation of instrumentalism), or  
 ii) fall foul of the argument evinced in section 2.5 below, which concludes in subsection 2.5.2 that general evaluative criteria are not suitable instruments for comparing co-evolutionary theories in philosophy of mind.

These considerations indicate that the use of a bootstrap strategy will not automatically ensure that the scientific realist will avoid the difficulties associated with metaphysical realism.

I shall now argue that co-evolutionary theories within the

philosophy of mind suffer from particular problems when employing the bootstrap strategy. These problems may be directly attributable to the novelty of their interdisciplinary approach.

## 2.5 Evaluation of the bootstrap strategy - part II

"Philosophers, for the most part, are constitutionally timid, and dislike the unexpected. Few of them would be genuinely happy as pirates or burglars. Accordingly they invent systems which make the future calculable, at least in its main outlines" (Bertrand Russell, Philosophy's Ulterior Motives, Unpopular Essays, 1950).

It would perhaps surprise Russell to know that many philosophers, including those who attempt to use findings from other disciplines in the construction of their theories, are in fact burglars, since they help themselves unjustifiably to implicit metaphysical, that is, untestable, assumptions.

This will become clear following an examination of an alternative view of the bootstrap methodology put forward by Jon J. Nordby in his article "Bootstrapping While

Barefoot" (henceforth referred to as Nordby (1989)). Nordby believes that his suggestions apply to the use of the bootstrap strategy to solve serial crimes, rather than to its use in science. My view is that his criticisms of the formal aspects of the Glymourian bootstrap approach have a wider application than he thinks, and are particularly pertinent to those philosophers who attempt to develop co-evolutionary theories using both philosophical and scientific considerations.

The remainder of this chapter has two central functions. Firstly Nordby's position will be examined. This will be followed by an outline of the major features of those theories that will fall into the same bootstrap category as Nordby's example. I shall then discuss the consequences for our key issues of evidence and realism .

### **2.5.1 Nordby's account of bootstrapping**

Clark Glymour holds that the bootstrap strategy is at its most useful when new scientific theories are being evolved, rather than in more established areas of scientific activity. He writes that it is generally found

"when quite novel theories are introduced, when systematic theorizing is extended to new subject matters, when theoretically based predictions go

wrong, and when fine-grained questions arise as to the relative importance of various experiments or observations" (Glymour, 1980, p. 172).

As we shall see in a moment, it is the very novelty of the co-evolutionary approach which renders inappropriate the use of the general evaluative criteria suggested by Glymour<sup>17</sup> and endorsed by philosophers of mind (see subsections 2.4 and 2.4.2 above) and which makes Nordby's comments so relevant.

Nordby begins by making a comparison between the models constructed in science on a mathematical basis and those constructed to solve criminal investigations. Examples of mathematical modelling exist, he writes, in many different sciences ranging from geology to physics. Equations are formulated which reflect phenomena in the observable world, but which are not identical with them. Nordby writes that scientific models

"are, for example, mathematical constructions in the form of equations that allow predictions and explanations of earthquakes. Equations are not the same things as moving geological masses" (Nordby, 1989, p. 387, note 3).

Nordby takes a realist's view of the role that scientific

theories play in our relationship with the external world. They predict and explain natural phenomena.

Like scientific models, crime models are simplified representations which aid our understanding of complex matters (Nordby, 1989, p. 374) and, like scientific models, crime models are in some sense idealizations. But whereas the scientific models are evaluated according to their predictive and explanatory power, and

"abstract certain elements at the expense of others and simplify the process of explaining and predicting phenomena" (Nordby, *ibid*, p. 373),

crime type models<sup>18</sup> are

"constructed with legal definitions, which stipulate relevant conditions, and with various phenomenological laws, scientific theories, inductive inferences, and lore. These compose the alternative components identifying features of the model that defines a particular crime" (Nordby, *ibid*, p. 374).

The differences between a scientific model and a crime type model do not end here. A scientific model forms part of a scientific theory. As we saw in Chapter 1, subsection

1.2.1, theories are formulated to explain or predict the nature or occurrence of events using law-like statements linking types of events. Consequently, evidence that tells against a scientific model will falsify some part, or perhaps even all, of the scientific theory in which it is embedded. The explanations provided by scientific theories and the models associated with them are intended to be laws of nature which hold without exception.

This is not true of crime type models. Nordby contends that the crime type model does not explain a crime in the same way as a scientific model explains a phenomenon in nature. In his view, the crime type model is

"only a heuristic device for discovering the actual, concrete token crime story" (ibid, p. 376)<sup>19</sup>.

Whereas the scientific theory will itself be confirmed or refuted by evidence gleaned from individual occurrences, the crime type model does not admit of either confirmation or refutation. Rather, it is the individual crime token models of concrete criminal acts that are confirmed or refuted by the evidence, and that are either true or false *in that particular situation*. The crime type model itself remains valid even if a token model is ultimately discarded. This is because it may still apply to a



different crime. Nordby comments that

"[t]he crime type model does not function as a theory since it is underdetermined by any investigative outcome....No matter how many times a model fails to lead us to the right crime story, it is always possible that there will be a time when it does" (ibid, pp. 376-7).

Nordby then goes on to explain in what way he thinks that bootstrapping in science differs from the use of the bootstrap methodology to solve serial crimes. Firstly, detectives are unused to creating novel hypotheses. They generally just employ the device of arguing to the best explanation from a range of those already available. A decision as to what constitutes the best explanation is facilitated in cases of empirical underdetermination by the employment of scientific techniques from psychology, statistics and computer science, among others (Nordby, ibid, p. 378). Yet very few actual cases ever fit the models generated exactly.

Secondly, Nordby comments that, although it is possible to cite numerous facts about the events surrounding a particular crime<sup>20</sup>, it is not always easy for the detective to sort out which of the facts actually constitute evidence in support of one crime type model

rather than another, when he is trying to build up a crime token model of events. He writes that

"[u]sing current investigative methods, we can show that a fact is relevant or irrelevant only given a specific model: there are no obvious principles governing model type component selection in serial homicide cases" (Nordby, *ibid*, pp. 379-80).

For example, using a crime type model that a husband has murdered his wife, then the fact that she remortgaged their house without his knowledge may be evidence in support of that model, whereas her act may be accorded less weight using a crime type model which assumes that a woman may be murdered by her blackmail victim or by a complete stranger.

This relationship between the selection of evidence and the crime model type that the detective is using should be compared with Glymour's claim that evidence will only ever support a hypothesis relative to a theory.<sup>21</sup> However, Nordby (unlike Glymour, see subsection 2.4.1 above) is aware of the relationship between evidence and probability. He writes that

"[a] fact is relevant for an explanation if and only if it increases or decreases its

probability" (ibid, note 10, p. 387; see also note 9, p. 387).

As we have seen (subsection 2.4.2), this view is also incorporated into the form of the bootstrap strategy used in the philosophy of mind.

So, on Nordby's account, the detective must decide which crime type models may be appropriate starting points for use in constructing a provisional crime token model for the events that he is currently investigating. He must also consider the kinds of evidence that would make each crime type model the most likely explanation. Then he may make use of the bootstrapping strategy to enable him to move from the evidence currently available to him to the formulation of a provisional crime token model. Using that token model, he then makes further use of the bootstrap strategy to predict what other evidence may yet be found. If the evidence is not forthcoming, or evidence is found which contradicts the provisional crime token model, then the model must be either rejected or partially revised to take account of this.

Nordby's bootstrapping method follows a similar pattern to the one described by Glymour (Glymour, 1980) and employed by Hunt (Hunt, 1990) (see subsection 2.3.1 above). The strategy enables us to make inferences from relatively

unconfirmed hypotheses to new evidence, which in turn strengthens our belief in the hypotheses themselves. But whereas Glymour holds that scientific hypotheses yielded by the bootstrap method are confirmed relative to a theory, Nordby thinks that the constituent hypotheses of crime token models gain credibility relative to a crime type model when supporting evidence is discovered. This is not the same thing, since

"crime type models are not theoretical models because they lack the surrounding support of scientific theory, and therefore do not provide the same kind of explanations.....Although such models may have scientific theories as components, any attempt to build theoretical functions into these crime type models simply confuses them with theories or theoretical models" (Nordby, *ibid*, p. 386).

The importance of this distinction is exemplified in the question of empirical underdetermination. Detectives, like philosophers and scientists, have to face the problem of what to do in situations where the available evidence does not seem to point to any one explanation as being more plausible than other possibilities.

As we noted earlier, Glymour and many philosophers of mind

believe that, in cases of empirical underdetermination, we must rank alternative accounts of events according to such criteria as uniformity of explanation, number of hypotheses confirmed etc (see above, subsections 2.2.2, 2.4, 2.4.2). But Nordby holds that this is impossible when bootstrapping in criminology because no overall theory is available to determine which models or sciences should apply to the case under consideration. Nordby writes that

"[a]pplying Glymour's formal criteria for theory choice does not provide firm grounds for crime model choice. Ultimately the choice depends on deciding that certain lore, or natural or social sciences apply, and that is not a decision based on formal grounds" (Nordby, *ibid*, p. 385).

Nordby's use of the word "formal" is misleading in this context, since it implies that Glymour uses a rigid, almost algorithmic method for deciding between theories in cases where the evidence is equivocal between two or more rivals. Glymour's criteria might be better described as "abstract" or "non-particular". The implication in Glymour's work is that it is possible to assess rival theories in terms of how uniform an explanation of phenomena they provide, how many hypotheses of each theory are confirmed, what variety of evidence is obtainable in each case, etc, without ever considering the actual content of the different theories.

Nordby rejects these criteria for the purpose of solving the problem of empirical underdetermination in criminology because crime type models, unlike scientific models, are not embedded in wider theories which tell the detective how he should decide between competing models.

So the essence of Nordby's position is that bootstrapping is a useful strategy for the detective to employ when trying to solve crimes, since it may help him both to explain the evidence currently available to him and to predict what evidence may be discovered at a later date. However, since crime type models, unlike scientific models, are not embedded in wider theories which determine which models are applicable to individual phenomena, abstract criteria for deciding between rival theories are not appropriate tools for the detective. Rather, Nordby declares that

"[t]he issue instead remains which science(s), or other components composing the models, should apply to the problem" (ibid, pp. 382-3).

In the next subsection of this chapter, I shall argue that co-evolutionary theories in philosophy also lack the support of a wider theory. Consequently, the difficulties for the detective isolated by Nordby of determining which scientific and sociological considerations are relevant to the crime currently under investigation will recur for co-

evolutionary theorists who find themselves with no guidance as to which philosophical and scientific considerations should be allowed to carry any weight and which should be rejected. Once more, it seems that abstract evaluative criteria will prove to be inadequate for deciding between two theories whenever the problem of empirical underdetermination arises.

### 2.5.2 Criminal elements in Co-evolutionary Philosophical Theories

It is worth noting from the outset that it is not enough for those philosophers of science who hold co-evolutionary theories to argue that the considerations raised by my discussion of Nordby do not apply to their theories by dint of the very fact that they are theories and not merely models. Did not Nordby himself draw a sharp distinction between scientific theories and the crime type models from which detectives begin their deliberations?

This line of argument simply begs the question at issue, namely whether or not co-evolutionary theories do exhibit many of the features that create problems for the detective who must choose between crime type models. It is true that Nordby does distinguish between the way in which a scientific theory explains and the way in which a crime type model explains, but Nordby's examples are all drawn from theories which fall squarely within distinct branches of sciences such as physics, geology or biology. The mathematical representations used in the modelling of phenomena studied in these sciences are merely tools, and no mathematical conclusions are drawn from them except in so far as they can be transposed into geological theory. The geologist has no information to help the mathematician make new discoveries within his own field of expertise.



Nordby makes no mention of any more modern developments in science, such as the attempts to draw philosophical and psychological conclusions about the nature of the mind/brain from the study of the neurophysiology of the brain and vice versa. It is therefore an open question whether or not these co-evolutionary theories exhibit the same features as geological theories, even though they are undoubtedly intended to explain phenomena in the same way; that is, in terms of lawlike statements which apply universally to all phenomena of the same kind.

Moreover, the differences highlighted by Nordby between scientific theories and crime type models may be at least partially obviated by bad scientific methodology. In Chapter One (see especially subsections 1.2.7, 1.2.8, and 1.3) it was noted that it is possible to construct psychological theories so that some of their constituent hypotheses become untestable. Any failure to understand what kind of evidence is required to confirm or disconfirm a theory may well result in theories being retained when they should be, at the very least, revised. Philosophers of science may be guilty of allowing their own personal views about whether or not a theory is correct to influence their opinions about the nature of the tests and evidence that may be said to support their own theories and rival theories. This aspect of the bootstrap problem will be discussed more fully at the end of this chapter (subsection

2.6).

First, however, it is important to note further similarities between Nordby's crime type models and co-evolutionary theories. Nordby comments that many scientific techniques

"from sciences such as psychology, statistics, archaeology, physical anthropology, physics, and computer science" (Nordby, 1989, p. 378)

have been imported into criminology to assist detectives in finding the best explanation based upon the evidence available.

Nordby writes that the resulting models which are made up using these components from scientific theories are not themselves scientific theories, since they

"do not explain ..... in the same sense that geological theories and tectonic plate models explain earthquakes" (Nordby, *ibid*, p. 378).

However, although philosophers who advocate a co-evolutionary approach are engaged in the search for new theories, they employ very similar methods. In philosophy of mind, for example, they examine not only traditional

philosophical concepts, but also the results and methods of experimental psychology, neuropsychology and neurological research, before attempting to synthesize everything into one theory.

The co-evolutionary approach to the mind/brain question is relatively novel, and hence, if Glymour is to be believed, the bootstrap strategy should be appropriate to their task<sup>22</sup>. But it is *precisely because* this approach is so new, that Nordby's considerations regarding the bootstrap strategy apply. Like crime models, neurophilosophers lack any over-arching theory which determines which of the sciences should be considered in formulating their opinions and which should be ignored or given less weight. Again, this becomes a question for decision, and formal considerations do not apply. Nordby comments that

"Glymour writes formally about bootstrapping and theory choice as if it had already been determined which models, or which sciences, having phenomenological laws embedded within them, apply. In criminal investigations, where there are real choices among competing crime narratives as they develop, this is not true" (Nordby, *ibid*, p. 383).

Nordby goes on to argue that

"[u]sing explanations to bootstrap other explanations within an uncertain model describes bootstrapping while barefoot.. Detectives are barefoot in the sense that crime type models, unlike theoretical scientific models, lack the support of associated theory" (Nordby, *ibid*, p.383).

However, many co-evolutionary theories will also lack this support. Paul Churchland, one of the pioneers of the co-evolutionary approach to the study of the mind/brain comments that

"[g]iven our position as proto-scientists, philosophers are seldom in the position of having a fully-developed theory ready for unambiguous experimental evaluation" (P. M. Churchland, 1986a, pp. 6-7).

Further details of the methodologies of the co-evolutionary theorists of the mind/brain will be given in Section B (see especially Chapters 5 and 6). Suffice it to say for the moment that it is one thing to recognize that there is a need for a guiding theory to govern this new approach, but quite another to avoid the pitfalls that its absence entails.

## 2.6 The outlook for co-evolutionary theories

It is now easy to see how neurophilosophical, or indeed, all co-evolutionary theories may still fall foul of metaphysical dangers even whilst ostensibly maintaining a scientific realist position. If a hypothesis is simply assumed rather than properly tested, or if there is insufficient empirical evidence available to support it, then it may be tainted with metaphysical realism (see above, subsection 2.2.1).

Paul Churchland writes that

"[o]ur discipline [neurophilosophy] appears to be a *a priori* because the 'metaphysical' issues with which we deal generally are a fair distance removed from any obvious empirical resolution. But that is a reflection of the early stage of the inquiry, not of the superempirical nature of the subject matter" (P. M. Churchland, *ibid*, pp. 6-7).

However, this stance is problematic. Even if we accept that co-evolutionary philosophers of the mind/brain are aware of the need for testability within their theories, two problems still remain.

Firstly, are these philosophers genuinely prepared to acknowledge that they will have difficulty in providing evidence in support of their theories and that therefore their theories can only have a provisional status, or do they, like the detectives in Nordby's article (Nordby, 1989, p. 383), sometimes forget that? This question will be discussed at length in Section B.

Moreover, even if the philosopher is prepared to treat his theory as a provisional attempt to make sense of the evidence at hand, any methodology which involves on-the-spot decisions about the admissibility of evidence from a variety of scientific disciplines and about the nature of satisfactory testing may still prove unsatisfactory. This is because such a methodology runs the risk of favouring those sciences and tests that yield the evidence that best supports a preferred theory. Equally, evidence cited in support of rival theories may be rejected out of hand because of a genuine but mistaken belief that this evidence does not come from a credible or important enough source.

The difficulties are exacerbated owing to the disparate nature of the concepts at work in philosophy and in the realms of psychology and neurophysiology. As Hundert comments,

"the relationship of 'fact' to 'theory' is complicated here by the diversity of those three 'approaches to the mind' which we have brought together. As Aristotle warns us, what counts as a fact in a field like neurobiology is something fundamentally different from what counts as a 'fact' in a field like existential psychiatry. The 'data' of shared phenomenological experience cannot be expected to be 'precise' in the same way as the 'data' of the action of growth associated proteins in the central nervous system" (Hundert, 1989, p. 300).

In other words, not only must philosophers contend with the problems of conflicting theories (empirical underdetermination) and conflicting evidence, they must also deal with sources of evidence that do not really fit comfortably together. As a result we may again choose the evidence that supports our own theory best, when we should have retained the evidence that we have just discarded.<sup>23</sup>

Furthermore, the very way in which philosophers set up their theories may lead to the exclusion of tests that are likely to refute preferred, but weaker hypotheses. This point has also been made by Ronald Giere (see Giere, 1983). Giere comments that it is perfectly possible for philosophers to formulate their theories so as to guarantee the desired results under test (with no intention to do so whatsoever). They can do this by systematically building

into the theories and tests so many constraints about what constitutes a successful hypothesis, or about the nature of acceptable evidence, that the probability of the theory passing the test is greatly increased, but this does not make the test a good test.<sup>24</sup> Giere writes that

"it is not in fact unlikely that a model designed to accommodate a given result should in fact do so. And this is true no matter whether the corresponding hypothesis is true or false. It is possible, therefore, to be justifiably mistaken about whether a given experiment constitutes a good test or not" (Giere, 1983, p. 286).

Consequently, philosophers and scientists may well help themselves to metaphysical assumptions to which they are not entitled, even though they believe that they have subjected their theory to rigorous tests. By the time they realize their error, (if at all), it may be too late. Similarly, Nordby thinks that detectives may realise too late that they have been pursuing the wrong crime type model. The ways in which neurophilosophy falls into this trap will be illustrated in Section B (notably in Chapter 6).

This systematic failure to take account of all sources of evidence that do not necessarily support one's preferred



theory is reminiscent of Achinstein's criticism of Glymourian bootstrapping that the notion of evidence upon which it rests takes account neither of the need for evidence to increase the probability of a hypothesis being true, nor of all the relevant background information (see above, subsection 2.4.1). It appears that co-evolutionary philosophers of mind who are forced into deciding between rival theories on the basis of decisions about what evidence is permissible may be liable to the same criticism if it can be shown that they systematically pay insufficient attention to possible empirical sources of evidence, and raise the probability of their theory's satisfying tests only by means of artificially imposed constraints. This supports my argument in subsection 2.4.2 that although the bootstrap strategy employed by philosophers of mind ostensibly takes account of the need for evidence to lend increased support to the hypotheses of a theory, this is not in itself sufficient to remove Achinstein's criticism. The evidence must also be of the right kind. This will be illustrated with respect to the theories of the mind/brain discussed in Section B.

But perhaps Glymourian general evaluative criteria will be enough to decide between rival theories in the philosophy of the mind/brain. Paul Churchland certainly seems to think so.

He writes that

"we must be guided by abstract considerations of simplicity, coherence with what we already presume to know, and intuitive plausibility" (P. M. Churchland, 1986a, pp. 6-7).

However, it may be argued, firstly, that this does not obviate the need for evidence in support of a hypothesis relative to a theory to be good evidence and not evidence attained by poor testing. One could hardly argue that a theory is to be preferred to a rival because it contains more confirmed hypotheses if these hypotheses were not properly confirmed by the evidence cited!

Secondly, the charge of instrumentalism must still be faced by so-called scientific realists who rely upon general evaluative criteria to support their theories when these criteria cannot show that a theory is to be preferred because it is true (see above, pp. 54-5). This charge will also be discussed with reference to co-evolutionary philosophy in Section B.

I have contrasted internal realists with their metaphysical realist opponents. I will now turn in Section B to a closer investigation of the metaphysical tendencies inherent in the theories of philosophers of the mind/brain who purport to be scientific realists, and evaluate the implied criticisms of Chapter 2.

1. See Introduction for a brief explanation of this term. Co-evolutionary theories in the philosophy of mind will be discussed in detail in Section B.

2. Even if such accuracy were possible, this would still not provide us with objective knowledge. cf. Thomas Nagel "What is it like to be a bat?", in Nagel (1979).

3. The term "materialist" is used here to denote the general belief that physical matter exists over and beyond our sensations, rather than in the more narrow sense used by philosophers of mind to indicate that mental processes can be redescribed in terms of brain processes. To avoid unnecessary confusion, I shall refer to this latter group of philosophers as "physicalists".

4. This is a very simplified description of what may, in fact, be a very complicated methodology with many hidden pitfalls. The problems inherent in this approach will be discussed in the course of this thesis.

5. See, for example, Newton-Smith (1981), pp. 40-43.

6. cf: my discussion of Paul Churchland in Chapter 1, subsection 1.1.1. By contrast, Ellis (ibid, p. 72) feels that we should not concern ourselves with such matters. We should only worry about the human position.

7. Note that Glymour himself defends his choice of cases from the history of science to illustrate the bootstrap strategy on the grounds that "[t]hey are diverse in their subject matter, and reveal not only different applications of the strategy ..., but also different features of that strategy" (Glymour, 1980, p.177).

8. For an example of this, see Dorling, 1971.

9. This is a very simplistic explanation of the views held by Reichenbach and Carnap. However, I shall not be concerned with their theories, except in so far as they are precursors of the Glymourian position, so a more detailed exposition of their arguments would be inappropriate here.

10. Peter Achinstein defines the bootstrap condition in terms of the first part of Glymour's strategy only. The bootstrap condition is "that  $e$  is evidence that  $h$  with respect to theory  $T$  if and only if using  $T$  it is possible to derive from  $e$  an instance of  $h$ , and the derivation is such as not to guarantee an instance of  $h$  no matter what  $e$  is chosen" (Achinstein, 1983, p.355). Whilst this aspect of the methodology is important, I do not think it is the whole story, nor does it merit the title of bootstrapping. We pull ourselves up by the bootstraps only if we are able to confirm the hypotheses from which we began in the light of new evidence. If these hypotheses remain

unconfirmed (or, in any event, no more probable than before the new evidence was brought to bear [cf. my discussion of Hunt, 1990, p.000]), then we are not bootstrapping, but merely building castles in the air.

11. All the biographical details in this example are taken from *Chambers Biographical Dictionary*, Revised Edition, ed. J. O. Thorne & T. C. Collocott, W & R Chambers Ltd, Edinburgh, 1984.

12. It is worth pointing out that although Glymour adopts similar criteria for deciding between empirically equivalent theories as do many other scientific realists, he does not share their belief in naive holism. Rather, he maintains that it is possible to regard a body of evidence as supportive of part of a theory, but not necessarily all of it. We may thus reject some of the theory, whilst retaining the rest.

13. Note that Hunt's use of the bootstrap strategy does not suffer from this difficulty since he combines the basic bootstrapping method with probability theory (cf. section 2.3.2 above).

14. I have changed Achinstein's notation in this quotation to fit with mine so that it follows more naturally from my text. The amended notation is given in square brackets. This does not alter the sense of Achinstein's words in any way.

15. See note 1 above.

16. Of course, the process may be more haphazard in reality than this description might suggest. For example, Marr (1982a, p. 331) comments that "part of the fun is that we never really know where the next key is going to come from".

17. This difficulty with the general evaluative criteria for choosing between theories put forward by Glymour does not diminish the importance of the additional problems with these criteria set out in my discussion of Achinstein (1983) in subsections 2.4.1 and 2.4.2 above. Indeed, as will be seen in Chapter 4, subsection 4.5, Achinstein's criticisms of Glymour, taken together with my comments in subsection 2.4.2, suggest that Nordby's remarks are very relevant to the philosophy of the mind/brain.

18. Nordby's distinction between crime type models and crime token models is exactly congruent with the standard philosophical type/token distinction. In the case of the crime of unlawful killing, for example, there may be one crime type model for the crime of a jealous lover, another for patricide, and still another for the opportunistic murder of a complete stranger. The details that make up each model will be generic features which

frequently recur in killings of these types (eg: a missing wallet at the scene of an opportunist murder). The crime token model of the unlawful killing of John Smith, however, will only contain details of the facts specific to that one crime.

19. Once more the bootstrap methodology is being used to construct a general hypothesis and a specific hypothesis from the same initial body of evidence.

20. Nordby uses the example of serial homicide cases throughout his article. In other crimes, such as rape, it is not always as easy to establish the facts about what took place. I am indebted to Dr. Ian Morley for this point.

21. The two situations are not identical, because Nordby believes that crime models do not constitute theoretical models, because criteria of confirmation and disconfirmation do not apply to them. However, my concerns are not Nordby's, and this aspect of his paper does not affect my central argument that co-evolutionary theories in the philosophy of science may well share precisely those features of criminology that make Glymourian considerations an inappropriate characterization of the actual kind of bootstrapping that takes place.

22. The positive attitude of many philosophers of mind towards bootstrapping will be discussed in Section B.

23. For concrete examples, see my discussion of Marr and Churchland in Chapter 6.

24. See Chapter 6 for some concrete examples of this tendency.

Section B - Bootstrapping and Realism in the Philosophy  
of Mind

Introduction

B1 The story so far

As we have already noted in the previous two chapters, it is perfectly reasonable for the realist to acknowledge the existence of an external reality and to make mention of such a reality in his theories. However, the metaphysical realist encounters problems because he fails to see that this very positing of an external reality should itself be regarded as part of a scientific theory. Instead, the metaphysical realist founds his individual theories about phenomena on the basis that this external reality *must be assumed* to exist for any further scientific activity to be possible.<sup>1</sup>

For a theory to be classed as scientific rather than metaphysical it must be possible to refute it by citing empirical observations which conflict with it. As Popper points out, it is impossible to refute pure existential statements in this manner, because

"a strict or pure existential statement applies to the whole universe, and it is irrefutable

simply because there can be no method by which it could be refuted" (Popper, 1969, p. 196).

Clearly, the statement "there exists an external reality beyond appearances which permits the practice of science" falls into this category of pure existential statements, and is therefore, in Popper's opinion at least, irrefutable and hence unscientific.<sup>2</sup>

My discussion of the veil-of-perception issue (Chapter 1, subsections 1.2.3 - 1.2.8) highlighted the problematic features within the theories of metaphysical realists. It should be clear by now that these theories are full of irrefutable (because untestable) statements of the kind condemned by Popper.

By contrast, Popper is prepared to bite the bullet and agree with the Kantian view that

"we cannot possess anything like full knowledge of the real world with its infinite richness and beauty" (ibid, p. 194).

The bootstrap strategy is intended to provide a way out of this difficulty for scientists and philosophers who accept that the problems raised for realism by the veil-of-perception issue are genuine, but who are reluctant to

adopt anything other than a full-blooded realist stance towards the existence of a reality beyond our perceptions. However, as I argued in chapter 2 (subsection 2.2.2), these scientific realists are really defending the view that an external reality exists on the grounds that this general hypothesis provides the best explanation for the success of more specific scientific hypotheses in explaining and predicting phenomena in individual fields of study.

If this view is compared with the claim of the metaphysical realist that there exists an external reality that permits the very inception of scientific activity, we can see that the scientific realist, unlike the metaphysical realist, succeeds in formulating a theory that seems to live up to the Popperian challenge of meeting his criterion for science. If empirical evidence can be found which refutes the specific scientific hypotheses, then the general hypothesis that the world exists is also rendered doubtful. This is because these specific hypotheses play a dual role in the scientific realist's framework. Not only are they intended to help us to explain and to predict phenomena, their success in doing so also constitutes evidential support for the general hypothesis that the world exists beyond our perceptions (and, of course, for any further general hypotheses about what this external reality must be like). This is the essence of the bootstrap strategy discussed in Section A.



However, we also saw in chapter 2, subsection 2.2.2, that it is not enough for philosophers simply to pay lip service to the principles of scientific realism if they are to avoid the problems that have thus far been associated with metaphysical realism: if their methodology is faulty, they may be equally guilty of making unwarranted metaphysical assumptions.

Scientific realists are therefore charged with the obligation of producing theories which provide a true explanation of events rather than theories which simply save the phenomena (see Chapter 2, subsection 2.2.2). We have already noted that this is no easy task. The method of bootstrapping is often used by scientists and philosophers alike to enable them to derive both specific and general conclusions from the same initial evidence. All of the conclusions derived from use of the bootstrap strategy should be regarded as hypotheses which must be rejected or revised if new evidence contradicts them.

It has already been argued in the early parts of this thesis that successful use of the bootstrap methodology depends upon overcoming problems surrounding the criteria available for deciding between competing theories (Chapter 2, section 2.5.2). It is very difficult for the scientific realist to use abstract evaluative criteria (as suggested by Glymour; see chapter 2, subsection 2.4) without either

interpreting them in a subjective or biased fashion (deliberately or unwittingly), or lapsing into an instrumentalist position.

Further, I have intimated that the use of the bootstrap methodology in new realms of co-evolutionary research lacks the safeguard of an overarching theory which may help us to determine which kinds of evidence should be admissible and which rejected. Again, the chances of obtaining unbiased use and interpretation of sources of evidence are significantly diminished.

The remainder of this thesis constitutes an evaluation of the use of bootstrap strategies in both traditional analytic theories in the philosophy of mind (Chapters 3 and 4), and in the more modern co-evolutionary theories (Chapters 5 and 6) in the light of these considerations. I shall be particularly concerned with the question of whether their use of the strategy does enable them to avoid the pitfalls of metaphysical realism.

### **B2 Application of the veil-of-perception issue to theories of the mind/brain**

If theories are to be genuinely explanatory, they must be testable. Problems may arise if an explanatory account is systematically protected from thorough, unbiased testing.

This is precisely the situation delineated by the traditional veil-of-perception issue. Moreover, any attempt to explain what is going on in our minds by reference to what is going on in our brains should also be testable.

How can we explain the activities of our minds in terms of what goes on in our brains? As Kant saw (vide Churchland, 1986a, pp. 46, 248-9), any attempt to understand such activity by means of pure introspection will by no means be assured of success, since our perceptions of ourselves are no less fallible than those we have of the world around us.<sup>3</sup> Scientific theories of the mind/brain exhort us to abandon introspection and to regard our mental experiences as just another subject for scientific study. The theories discussed in this section embrace a range of different views concerning the best ways to learn about the operation of the mind/brain.

Of course, in practice, it may be very difficult to separate what we would normally think of as the workings of the mind from those of the brain. Furthermore, modular theories of mind have cast doubt upon the traditional image of man as a united self which confronts the rest of reality. The discussion of modular theories of the mind in Chapter 5 is included to demonstrate that the veil-of-perception issue does not simply disappear if the traditional gap between the self and the external world is

displaced, and relocated as a gap between perception and cognition as Fodor (1983) suggests. This is because the veil-of-perception issue is really only an illustration of the more fundamental and underlying problem of explanation that all theories of the mind/brain must confront if they are to be acceptable in the light of scientific criteria.

It is one of the major ironies within the current state of the philosophy of mind that philosophers working in this field are scathing about the importance of the original veil-of-perception issue, and yet fail to understand the more general problems of explanation that the issue illustrates so well. For example, McGinn remarks that

"[t]o suppose that the involvement of an inner experience produces perceptual indirection, or enclosure in a world of merely mental objects of acquaintance, is the analogue of use-mention confusion. This is not to say that it is false that we are acquainted with our own experiences - we are, on the contrary, acquainted with them in acts of introspection - but it does not follow that we are not also and primarily acquainted with external objects" (McGinn, 1982, p.45).

I am not concerned with the validity of McGinn's criticisms

of the veil-of-perception issue per se, but wish merely to note that no mention is made in his work of the difficulties of explanation that lie behind the issue. This may perhaps account for his failure to avoid these very difficulties in his treatment of the issue of consciousness. It will be demonstrated in Chapter 3 (especially in subsection 3.2.5) that it is in fact possible to discern in McGinn (1990) the very line of argument that was identified in Chapter 1 as symptomatic of the difficulties associated with the veil-of-perception issue.

David Papineau (1987) has suggested that nothing is to be gained by incorporating the veil-of-perception issue into an examination of the questions surrounding representation within the mind/brain. Papineau holds that the talk in terms of a veil-of-perception fosters an unwarranted sympathy with the Cartesian notion of a dichotomy between what is "given" to the conscious mind and what is outside it. I think that Papineau is mistaken for two reasons.

Firstly, I have already shown in Section A<sup>4</sup> that the situation set up by the veil-of-perception issue can be described in terms of explanation and testability that make no reference to notions of the "given", and that are currently widely used by the scientific community.

However, this may not be enough to dissuade Papineau, since he too (1987) believes that the essential question raised in the issue of the veil-of-perception can be rephrased. I need to demonstrate that the veil-of-perception issue presents the problem in a manner which is positively advantageous for my argument and that its terminology is not responsible for seducing those philosophers who do recognise its importance for the study of the mind/brain issue, such as P.S. Churchland (1986a, pp.246-8), into believing that there are unassailable givens in the philosophy of mind.

Churchland comments that

"common sense is theory laden, that there is no Given, that there are neither epistemological foundations nor unrevisable theories nor even First Philosophies. I unabashedly avow my persuasion to these claims" (P.S. Churchland, 1980a, p.153).

Moreover, Churchland discusses the use of the bootstrap strategy in terms which strongly suggest that she sees it as a solution to the problems of explanation and evidence raised by the veil-of-perception issue. She writes that

"[a] remarkable thing about the human brain is

that it can use .... primitive theories to bootstrap its way to ever more comprehensive and powerful theories - to find the reality beyond the appearances" (P.S. Churchland, 1986a, p.265).

It is precisely because there are no indubitable givens in philosophy or in science that the bootstrap strategy is so necessary for explanations of mind/brain activity. Indeed, bootstrapping is the key feature of Churchland's co-evolutionary methodology. She comments that

"the co-evolutionary process [is like] two rock climbers making their way up a wide chimney by bracing their feet against the wall, each braced against the back of the other" (P.S. Churchland, 1986a, p. 374).

Unlike McGinn and Fodor, who are quite happy to rely upon the traditional linguistic representations of the world as a basis for any theory of representation within the mind/brain, Churchland takes a far more open-ended stance.

Whilst she does not object to the use of the sentential paradigm as an initial starting-point, Churchland does not believe that we are entitled to simply postulate the existence of such representations *inside* the head without any independent proof. Indeed, she believes that these

starting-points may well be revised once we know more about our internal representational systems. Hence, far from being seduced into the opinion that these starting-points constitute Cartesian "givens", she takes the view that they may well be leading us down the garden path. It will become clear in my discussion of Churchland in Chapter 6 that she *genuinely* believes that she endorses only testable hypotheses. However, there are systematic flaws in her methodology that render some of her hypotheses untestable. Nonetheless, Papineau has still not won the day, since it is the very use of the veil-of-perception issue that demonstrates these flaws in Churchland's argument, hence it is far from useless within the philosophy of the mind/brain.

### **B3 - An Introduction to the issue of representation in the mind/brain**

All the theories examined in this section of the thesis are concerned how information is represented within the mind/brain. For this reason, a few introductory comments upon the importance of representation for the philosophy of mind (and, in particular, for those theorists discussed in this section) are in order.

Both cognitive psychologists like Jerry Fodor and co-evolutionary theorists like Marr and Churchland accept the



idea that the mind/brain operates by means of computations upon symbols or representations. This is the traditional idea of the mind/brain as an information processor.

There is also extensive agreement about the purpose of representations within the mind/brain. Teleological accounts of representational systems abound among philosophers of mind. Creatures, including human beings, need an accurate representation of the world in which they live in order to survive. McGinn writes (1982, p. 82) that

"[e]volved creatures are intent upon preserving their lives, and bodily action is (for many of them) essential to their survival; but actions need to be guided by information about the world if they are to serve the end of survival".

Similarly, Churchland comments that

"[o]bviously, the organism that flees in the absence of predators and feeds willy-nilly is doomed to be prey for those more lucky organisms fitted out with cells coordinating *representations of the world with movement in the world*" (P.S. Churchland, 1986a, p.1, her italics).

In a later article, she writes that

"[i]mprovements in sensorimotor control confer an evolutionary advantage: a fancier style of representing is advantageous so long as it is geared to the organism's way of life and enhances the organism's chances of survival" (P.S. Churchland, 1987, p.548, her italics).

Marr comments along the same lines that

"[t]he usefulness of a representation depends upon how well suited it is to the purpose for which it is used. A pigeon uses vision to help it navigate, fly, and seek out food. Many types of jumping spider use vision to tell the difference between a potential meal and a potential mate..... because vision is used by different animals for such a wide variety of purposes, it is inconceivable that all seeing animals use the same representations; each can confidently be expected to use one or more representations that are nicely tailored to the owner's purposes" (1982a, p.32).

Jerry Fodor, too, thinks that an

"organism has available means for representing not only its behavioural options but also: the probable consequence of acting on those options, a preference ordering defined over those consequences and, of course, the original situation in which it finds itself" (Fodor, 1976, p.31).

Nevertheless, these quotations provide a suggestion that this alliance is uneasy and at best superficial. There is little or no agreement as to the way in which the teleological representations are realized within the mind/brain. For example, Colin McGinn is happy to talk in terms of folk psychological concepts such as beliefs and desires. Jerry Fodor, on the other hand, postulates the existence of a language of thought or Mentalese, which literally consists of sentences in the head. The symbols within these sentences are combined by means of computations over their syntax. In Fodor's opinion, the language of thought constitutes the paradigm of representation. For example, he comments that

"only sentences in the language of thought represent in, as it were, the first instance, and they represent in virtue of the natural teleology of the cognitive mechanisms. Propositional attitudes represent qua relations to sentences in

the language of thought. All other representation depends upon the propositional attitudes of symbol users" (Fodor, 1984b, p. 247).

McGinn does not share Fodor's computational view of internal representation. He writes that if

" we compare the manipulation of sentences in thought to the operations performed by computers on the sentences of the languages with which they are programmed.....[this]...is suspect because the 'languages' used by computers are not languages in the ordinary and required sense: the computer does not *understand* the sentences it operates on, and printing out symbols on a tape is not a kind of assertion" (McGinn, 1982, p.67).

Fodor does show some signs of acknowledging this objection, since he comments (Fodor, 1980) that the meaning of the sentences which constitute the language of thought is not given by the computations that are performed upon the constituent symbols. However, this does not render his view unproblematic. Fodor's arguments are examined in detail in Chapter 4.

I have already alluded to Marr's view that different species employ different representational systems that are orientated towards the specific information about the world that is essential for their survival. Other information is irrelevant and needs to be filtered out by the representational system. Consequently, although we need to understand the neurophysiological facts to fully appreciate how vision works, this should not be our main concern. Rather, our study of vision should be dictated chiefly by an attempt to isolate the abstract, high-level computational<sup>5</sup> constraints that govern the information selected for representation by each species in the light of the singular problems that it must overcome in order to survive. Marr writes that

"[w]hat higher nervous systems must do is determined by the information-processing problems that they must solve" (Marr, 1982a, p. 349).

This is why

"one has to exercise extreme caution in making inferences from neurophysiological findings about the algorithms and representations being used, particularly until one has a clear idea about what information needs to be represented and what

processes need to be implemented" (ibid, p. 26).

Marr's views will be examined in greater detail in Chapter 6 (subsections 6.1 - 6.1.4).

Patricia Smith Churchland disagrees fundamentally with both Fodor and Marr, particularly over the correct method of studying the mind/brain.

Churchland's own view is that there is little to be gained from the Fodorean method of studying the mind/brain which merely assumes the validity of folk psychological concepts such as knowledge, belief, and desire. Rather, we should be prepared to renounce even these, our most familiar and entrenched ways of seeing the world and our own activity within it, if other, more neurophysiological descriptions are better supported by the available evidence. For example, Churchland comments that although Fodor recognizes that

"it is indeed an empirical question whether the sentential presupposition ... is actually true, .. he is nonetheless quite innocent of any real suspicion that it might be false, or that the stresses and strains already visible in the sentential approach could be signs that we have to re-assess the prospects of that dear and

comfortable old shoe, the common sense psychological theory" (P.S. Churchland, 1980a, p. 154).

This view is also supported by Paul M. Churchland, who wrote as early as 1975 that

"[o]ne could of course nurture a hope that the structure, elements, and operations of human language systematically reflect or mirror all of the theoretically relevant structures, elements and operations of the brain, but there is no *empirical* evidence to sustain such a hope, and one would expect on the contrary that linguistic structures/operations reflect brain structures/operations only very grossly, selectively, and superficially" (P.M. Churchland, 1975b, p. 156).

Indeed, Patricia Smith Churchland is sharply critical of Fodor's theory of the language of thought, precisely because she rightly believes that he does not sufficiently concern himself with the dearth of satisfactory and testable evidence in support of his view. For example, she claims that

"[s]urely there is something bizarre about the

idea that a theory of meaning that has nothing whatever to do with human psychology or neurophysiology can explain the meaningfulness of language and how representational structures relate to the world?" (ibid, p. 545)

and denies that

"knowledge in general is sentential; rather, representations are typically structures of a quite different sort" (P.S. Churchland, 1987, p.545).

In Chapter 4, I shall demonstrate that her criticisms can be reformulated as an accusation against Fodor of metaphysical realism.

Churchland also rejects the amount of emphasis placed on abstract, high-level constraints upon the nature of representation within the mind/brain in David Marr's theory of vision. In her view, the co-evolutionary approach should be more thoroughly orientated towards the importance of "bottom-up" neuroscientific study. She writes that

"[w]e want a theory of how the mind-brain



represents whatever it represents, and of the nature of the computational processes underlying behavior. The collective effort to devise such a theory will be constrained by empirical facts at all levels, including neurophysiological, ethological, and psychological facts" (P.S. Churchland, 1986a, p. 5).

In sharp contrast to McGinn and the early Fodor, Churchland rejects the traditional methodology of conceptual analysis within philosophy and advocates a strongly empirical approach. She notes that

"it is the aim of neuroscience to discover how the mind-brain represents, how information about the world is filtered, stored, and transmitted, and in so discovering, to provide a theory characterizing 'information' and 'information processing'" (P.S. Churchland, 1980b, p.194).

It should be clear by now that it is possible to describe the activity of the mind/brain at different levels. Sloman makes a similar point. He writes that

"in some cases, the intelligent thing to do is to find a special-purpose representation, tailored to the problem domain. Of course, it would be

even more intelligent to have a deep understanding of the nature of the representation and the reasons why it is appropriate. It may prove best for the second-order reasoning to use a quite different type of representation, eg: logic" (Sloman, 1985, p.175)

Like Sloman, Churchland is quite prepared to acknowledge that there are different levels of description that may be applied to the mind/brain. Her insistence that we must not allow temporary constraints upon the language available to us for this task to bias our judgements on ontological facts about the mind/brain is wholly consonant with her remarks about the dubious validity of the sentential paradigm as a representational model for the mind/brain (as described above)<sup>6</sup>. Hence her comment that

"it would be foolish to suppose folk psychology must be true because at this stage of science to criticize it implies using it. All this shows is that folk psychology is the only theory available now" (P.S. Churchland, 1986a, p. 397, her italics).

Churchland believes that judicious use of the bootstrap strategy opens up the possibility of other more appropriate means of describing the mind/brain and its functions. This

belief will be evaluated in Chapter 6 (subsections 6.2 - 6.2.5).

All the theories delineated above will be evaluated in the light of the issues of realism and explanation discussed in Section A.

#### **B4 - Structure of Section B**

I shall begin in Chapter 3 by discussing the work of Colin McGinn on the nature of consciousness, as McGinn can be regarded as representative of the tradition of conceptual analysis within the philosophy of mind.<sup>7</sup> This will be followed in Chapter 4 by an evaluation of the Representational Theory of Mind (or sentential paradigm) endorsed by Jerry Fodor. I shall demonstrate that both of these philosophers employ the bootstrap method, but without regard to the need for their hypotheses to be supported by sufficient and systematically testable evidence. As a result, they may both be regarded as metaphysical realists within the philosophy of mind, despite any protestations to the contrary.

In the final chapters of this section (Chapters 5 and 6), the discussion will be extended to cover those theories of the mind/brain that claim to take more account of evidence from other disciplines, such as psychology or

neurophysiology. These are the co-evolutionary theories of the mind/brain, and include the modularity theories of Jerry Fodor and Howard Gardner, the theory of vision put forward by David Marr and the eliminative materialism of Churchland. I shall show that the incorporation of further sources of evidence in support of hypotheses does not necessarily yield the desired result of a theory of the mind/brain that can truly be said to avoid the problems of metaphysical realism so clearly exemplified in my earlier discussion of the veil-of-perception issue (Chapter 1, subsections 1.2.4, 1.2.5 and Chapter 2, subsection 2.2.1). Indeed, I shall argue that, in these theories at least, these difficulties are exacerbated by the inclusion of findings from a variety of disciplines, because of the resemblance of the bootstrap argument employed by these theorists to the methodology employed by detectives described in Nordby (1989) (see subsections 2.5.1, 2.5.2).

1. See especially subsections 1.2.4, 1.2.5 and 2.2.1 for relevant discussion.
2. This is not to say that Popper regards such metaphysical statements as meaningless or false. On the contrary, they may form part of rational theories which have evolved as attempts to provide solutions to problem-situations. They may still be critically discussed in this context, even though they can never be refuted. To declare that metaphysical statements are meaningless is, in Popper's eyes, to confuse the question of meaning with the question of demarcation between science and metaphysics.
3. For an outline of the Kantian influence upon Churchland's theory of the mind/brain, see Chapter 6, subsection 6.2.1.

4. See especially Chapter 1, subsections 1.2.4, 1.2.5 and Chapter 2, subsection 2.2.1.

5. Marr is using the term "computational" in a much narrower and specific sense than Fodor or McGinn. For Marr, true computational constraints are high-level, abstract, specifications of a problem that needs to be solved by the representational system of an organism. The solution is then provided by a lower-level algorithm which determines the details of how information is represented and manipulated within an organism and which is instantiated by neurons (in a biological organism - it is worth noting that Marr's three-level theory of vision could also be applied to computers). Clearly, the Fodorean sense of "computational" is more nearly reflected in Marr's theory at the algorithmic level.

6. cf: Paul M. Churchland's insistence that there is no such thing as a theory-neutral foundation for knowledge (see Chapter 1, subsection 1.1.1). In essence, it will be recalled, he believes that we must be aware of the bias which is inherent in our use of language and in the theories which govern our observations and avoid drawing rash ontological conclusions.

7. McGinn himself remarks that "the philosopher seeks to discover a priori necessary truths about the phenomena of mind - truths that can be ascertained without empirical study of the mind and its operations, and truths that hold good for any conceivable exemplification of the mental phenomenon in question" (McGinn, 1982, pp. 3-4).

## Chapter 3 - Metaphysical Elements in McGinn's Account of Consciousness, Modularity and the Mind

### 3.1 Introduction

As I noted in Section A, it is perfectly reasonable for the realist to acknowledge the existence of an external reality and to incorporate the notion of such a reality in his theories. However, the metaphysical realist encounters problems because he fails to see that this very positing of an external reality should itself be regarded as part of a scientific theory. This means that it may be subjected to systematic and rigorous empirical testing and revised if necessary in the light of new evidence.<sup>1</sup> Instead, the metaphysical realist bases his individual theories about phenomena on the assumption that this external reality *must be assumed* to exist if any further scientific activity is possible. The problems delineated in my discussion of the veil-of-perception issue (Chapter 1, Sections 1.2.3 - 1.2.8) characterise the theories of metaphysical realists.

However, we also saw in chapter 2, subsection 2.2.2 that it is not enough for philosophers simply to pay lip service to the principles of scientific realism if they are to avoid these metaphysical difficulties: if their methodology is faulty, they may be equally guilty of making unwarranted metaphysical assumptions.<sup>2</sup>

Scientific realists are therefore charged with the obligation of producing theories which provide a true explanation of events rather than theories which simply save the phenomena. As I remarked in Chapter 2, subsection 2.2.2, this is no easy task. The method of bootstrapping is often used by scientists and philosophers alike to enable them to derive both specific and general conclusions from the same initial evidence. All of these conclusions should be regarded as hypotheses which may be subjected to tests, and rejected or revised if new evidence contradicts them.

However, I have argued that successful use of the bootstrap methodology depends upon overcoming problems surrounding the criteria available for deciding between competing theories (Chapter 2, subsection 2.5.2). It is very difficult for the scientific realist to use abstract evaluative criteria (as suggested by Glymour; see chapter 2, subsections 2.4, 2.4.1) without either interpreting them in a subjective or biased fashion (deliberately or unwittingly), or lapsing into an instrumentalist position.

Further, I argued that the use of the bootstrap methodology in new realms of scientific and philosophical research lacks the safeguard of an overarching theory which may help to determine in a systematic and impartial manner which kinds of evidence should be admissible and which rejected.

Again, the chances of obtaining unbiased use and interpretation of sources of evidence are significantly diminished.

### 3.2 Metaphysics and McGinn

This chapter examines the way in which traditional conceptual analysis can fall foul of the metaphysical problems associated with the veil-of-perception issue, even though it is claimed that the concerns of scientific realism are of paramount importance. The example I have chosen is Colin McGinn's (1991) treatment of the problem of consciousness. Although it is possible to identify a bootstrap argument within his discussion, his understanding of the need to subject his hypotheses to rigorous and systematic tests is minimal. Thus his views may be regarded to all intents and purposes as those typical of a modern metaphysical realist.<sup>3</sup>

Although I have chosen to illustrate the metaphysical difficulties that McGinn creates for himself in just one text, The Hidden Structure of Consciousness in McGinn, 1991, (pp. 89-125), his work is littered with unjustifiable metaphysical assumptions.<sup>4</sup>

It will be pertinent to begin with a brief outline of the



structure of McGinn's argument.

### 3.2.1 An outline of McGinn's theory of consciousness

The first part of McGinn's argument for the hidden structure of consciousness is grounded in largely philosophical considerations about the role of consciousness in our mental life, and about the nature of the link between consciousness and physical processes within the mind/brain.

McGinn draws an analogy between language and consciousness to argue that the true logical structure of our conscious thoughts is not the one revealed to us by introspection. He comments that

"[e]arlier logicians had mistakenly assumed that the real inner nature of propositions is made manifest in their ordinary linguistic vehicles, and it was this assumption that retarded the progress of logic. What has to be recognized is that language (meaning) possesses a level of structure that transcends, or underlies, the way it presents itself to us" (McGinn, 1991, p.93).

Similarly, McGinn continues, the deep structure of our conscious thoughts is not that revealed to us by

introspection. In fact, McGinn argues that it is this very division between the surface properties and the deep structure of our conscious thoughts that is responsible for the corresponding structural dichotomy within language.

McGinn's view that introspection may not tell the whole story derives at least some initial plausibility from the well-known experimental evidence that introspection is unreliable when it comes to giving reasons for our actions.<sup>5</sup> However, this evidence simply indicates that we are sometimes influenced by thoughts of which we are unaware. It sheds no light whatsoever upon the question of whether the thoughts of which we are aware have some hidden structure beyond that revealed to us by introspection. This remains an empirical question.

McGinn's suggestion is that it is impossible to explain consciousness without reference to hidden structure, despite the reluctance of most philosophers to even consider this as an option. Moreover, he argues, we have every reason for adopting an attitude of scientific realism towards the hidden structure of consciousness, just as we have done towards other hidden structures such as DNA. Such a structure would furnish us with a better understanding of the true logical character of our thoughts. It would be a

"stratum of mental reality, an inner

configuration of conscious states themselves, predating the development of the notation that captures it" (McGinn, *ibid*, p.99).

McGinn goes on to suggest that the hidden structure of consciousness would provide an (in his opinion, necessary) additional link between the surface properties of consciousness and the physical facts which cause them.

There are many difficulties inherent in McGinn's view, but comment upon these will be deferred until subsections 3.2.3 - 3.2.6 of the chapter. For the present, I am solely concerned with the structure of his argument.

McGinn writes that

"[t]he considerations advanced in the previous section in support of the idea of a hidden structure to consciousness are of a broadly metaphysical character; they concern the philosophical problems that might be solved by accepting that idea. In this section I shall argue that there also exist empirical data that are best explained by adopting the idea; we now have empirical evidence for the thesis that consciousness extends further than it introspectively seems" (*ibid*, p. 109)

McGinn then goes on to discuss the kinds of empirical evidence that he thinks supports his theory of the existence of a hidden structure of consciousness, including the phenomenon of blindsight. These arguments will be considered in subsection 3.2.5 of this chapter.

### 3.2.2 The bootstrapping device in McGinn's argument

McGinn starts his chapter with a statement of his purpose, namely, to establish that consciousness has a hidden structure in addition to the surface structure revealed to us by introspection. He cites three reasons for accepting this argument right at the outset (ibid, p. 91): i) it is needed to provide an explanation of the logical properties of conscious thoughts, ii) the need for an explanation of the relationship between conscious states and the physical body and iii) it explains phenomena such as blindsight. However, the points at which these reasons are introduced into the main thread of the argument, the work that they each do and the manner in which they support each other indicate that McGinn is using a bootstrapping argument. It is possible to reconstruct McGinn's line of argument using the same notation as I have employed in previous chapters.

Within the theory of consciousness,  $T$ , let  $H_1$  be the general hypothesis that consciousness plays an important role in our mental lives, and that it is somehow rooted in a

physical process within our brains<sup>6</sup>, yet the properties of consciousness of which we are aware do not seem to account for the central role it plays, nor for this connection between the mental and the physical.  $E_1$  is the evidence that McGinn thinks we have from a commonsense empirical consideration of our experience coupled with conceptual analysis that consciousness does indeed play an important role in our mental lives, but that surface properties of consciousness (revealed to us through introspection) are neither able to provide an adequate account of how this is possible, nor to yield a satisfactory explanation of the nature of the link between the mental and the physical aspects of consciousness.<sup>7</sup>

As a result of the considerations arising from a combination of  $H_1$  and  $E_1$ , McGinn formulates a more specific hypothesis  $H_2$ .  $H_2$  asserts that there is a deeper structure to consciousness hidden beneath the surface aspects that alone are accessible to introspection.

McGinn continues to combine conceptual and empirical arguments and evidence in seeking support for the existence of the hidden structure of consciousness put forward in  $H_2$ . This evidence,  $E_2$ , consists in a discussion of phenomena such as blindsight, which McGinn regards as supportive of hypothesis  $H_2$ .

The bootstrapping device is operative in McGinn's argument since he is using his second specific hypothesis of the hidden structure of consciousness and the phenomena of blindsight to provide further confirmation of his first, more general hypothesis that consciousness is able to play a major role in our mental lives and yet is linked in an explicable manner to the physical. This role is better understood and explained if we postulate a hidden structure within consciousness. Similarly,  $E_1$  (our introspective awareness of consciousness) vindicates the role attributed by McGinn to consciousness in  $H_1$ , and the two together suggest the possibility of the hidden structure of consciousness postulated in  $H_2$ , relative to McGinn's overall theory of consciousness,  $T$ .

We have already seen that McGinn advocates taking a scientific realist's attitude to the issue of consciousness. Further, he is making use of the bootstrap strategy. We have already seen in Chapter 2 (subsections 2.3, 2.4.2) that this is a device frequently employed by those who uphold the principles of scientific realism (including philosophers of mind) to ensure that their hypotheses are informative and explanatory, and yet are properly supported by sufficient evidence. However, we also noted in the course of Chapter 2 (subsection 2.3) that, if the bootstrap strategy is to have any real success, it is preferable for each hypothesis within the

theory to be supported by relatively independent bodies of evidence. Moreover, it is essential that each hypothesis is genuinely and systematically testable, that is, we must be able to say what kind of evidence would count against it.

These considerations make it clear that McGinn cannot claim that his theory is a good one by simply stating that he is an adherent of scientific realism. The following four subsections examine the evidence that he cites in support of his hypotheses in greater detail to see if they meet the strict criteria for a successful bootstrapping strategy outlined in the preceding paragraph.

### 3.2.3 Putting the con into consciousness

We will begin with a consideration of what McGinn himself has termed the metaphysical side of his argument (ibid, p. 109).

McGinn implies that consciousness is involved in a very large proportion of our mental activity, although he does mention the existence of

"unconscious beliefs and desires of psychoanalysis [and] ... the subpersonal states of computational psychology" (ibid, p. 98).<sup>8</sup>

Indeed, when discussing the phenomenon of blindsight, McGinn comments that

"[o]n some possible interpretations of the data, ["sight"] would not be the right word, since the assumption of those interpretations is that nothing mental is going on in these cases; the conscious experience is wholly abolished, leaving only mechanisms that at best accompany experiences" (ibid, p. 114, McGinn's italics, underlined emphasis mine).

Leaving aside McGinn's specific interpretation of the phenomenon of blindsight for the moment, this quotation is still very important for it demonstrates that McGinn readily slips into a Cartesian identification of the mental with the conscious. Clearly, it is an empirical question just how much of our mental activity falls within the domain of conscious activity. Freud, for one, argued that this identification is far from correct. Moreover, the doctrine of functionalism raises the whole question of whether or not the qualia that are traditionally associated with our mental experiences are in truth an essential aspect of our mental states. McGinn may be correct to argue that it is poor methodology to refuse to ascribe a hidden structure to consciousness, on the grounds that



"[i]t simply begs the question against the thesis of hidden structure to insist that what is hidden must ipso facto belong to something other than a state of consciousness" (ibid, p.98).

However, the onus is on McGinn to provide us with some satisfactory evidence that consciousness is indeed responsible for so much of our mental activity. If he is not able to do so, then it is questionable whether he is really entitled to the very assumption that motivated him to postulate the hidden structure in the first place. Perhaps it is he, and not his opponents, who is begging the question. Consequently, doubt may be cast upon that part of  $H_1$  which states that consciousness plays such an important role in our mental lives. This in turn renders his argument for  $H_2$  much weaker, since he has less reason to declare that

"the demands of theory make the attribution of hidden structure to consciousness unavoidable. Only thus can we explain what needs to be explained" (ibid, p.91).

#### 3.2.4 Consciousness and the brain

Even if we accept McGinn's view that consciousness is so vital a component in our mental lives for the moment, there are still other difficulties inherent in McGinn's body of evidence ( $E_1$ ) that the important role of consciousness and the link between the mind and the body cannot be accounted for by the surface properties of consciousness. He writes that

"[c]onscious states, we know, depend causally and constitutively on physical states, but the way these states appear to us renders the dependence problematic. The physical sciences tell us about the nature of body and brain, and this nature seems far removed from the nature of consciousness; yet it must *somehow* be in virtue of physical facts that conscious states come to exist in the first place and have their causes and effects." (ibid, p.100).

McGinn continues

"The surface properties are not enough on their own to link conscious states intelligibly to the physical world, so we need to postulate some deep properties to supply the necessary linkage."

(ibid, p.100)

This argument is not a good one. McGinn is simply asserting that because he cannot see intuitively how the surface properties relate to physical properties of the brain, then the two cannot be identified with one another, and must be linked via a third set of properties, namely those which pertain to the hidden structure of consciousness postulated by  $H_2$ .<sup>9</sup> There are obvious parallels with the fallacious argument that, since I know the identity of my father, but do not know the identity of the masked man, then the masked man cannot be my father. In other words, McGinn dismisses summarily the possibility that the surface properties of consciousness may themselves be redescribable in terms of physical processes, without any need for a hidden structure of consciousness. Yet this possibility remains a live one irrespective of whether McGinn's own knowledge of these properties extends that far, provided that we are able to say what kind of evidence would count against this hypothesis, thus ensuring that it is systematically testable. It would seem therefore that there is a degree of empirical underdetermination between McGinn's inference to  $H_2$  relative to  $T$  and the possibility of a rival hypothesis within a second theory of consciousness,  $T_2$ , claiming that we may be able to redescribe the surface properties of consciousness in physical terms. The evidence cited by McGinn in favour of

his own view is inadequate. A mere analysis of what he perceives to be possible is not a systematic or rigorous test of a hypothesis, but rather a systematic, a *priori* exclusion of evidence from other sources that may refute his preferred view.

Of course, it is perfectly possible that McGinn's hypothesis that consciousness has a hidden structure is correct, even if his arguments that such a structure is necessary for our understanding of consciousness are, in themselves, poor. Further, if McGinn's hypothesis were correct, this would indeed cast doubt upon an identification of the surface properties of consciousness with physical processes. However, McGinn's body of evidence ( $E_2$ ) which he claims provides independent support for his specific hypothesis ( $H_2$ ) is also quite simply inadequate, in spite of McGinn's assertion that the operative properties of the hidden structure of consciousness

"really would explain how it is that chunks of matter can develop an inner life. There would be nothing hard-to-swallow or take-it-on-faith about the theory that detailed the powers of these properties" (McGinn, *ibid*, p. 104)

These properties would be

"neither at the phenomenal surface nor right down there with the physical hardware; they would be genuinely deep and yet they would not simply coincide with physical properties of the brain. Somehow they would make perfect sense of the psychophysical nexus, releasing us from the impasse that seems endemic to the topic." (ibid, p. 104)

Yet, just as we are all set to hear further details of these hidden aspects of consciousness, McGinn pours cold water upon our expectations by informing us (ibid, p. 104) that our conceptual apparatus is simply not up to working out these details. This is no mere contingent situation; McGinn is not making the point that our scientific study of the mind/brain is not yet sufficiently developed to enable us to discover the precise nature of the hidden structure of consciousness<sup>10</sup>. On the contrary: our ignorance of these facts is a necessary consequence of the way that our conceptual apparatus is constructed. McGinn refers the reader to Chapter One of the same volume for further elaboration upon this topic and it will be instructive to take a brief look at his arguments as they appear there.

He begins by defining the concept of cognitive closure:

"[a] type of mind M is cognitively closed with

respect to a property P (or theory T) if and only if the concept-forming procedures at M's disposal cannot extend to a grasp of P (or an understanding of T) .... But such closure does not reflect adversely on the reality of the properties that lie outside the representational capacities in question; a property is no less real for not being reachable from a certain kind of perceiving and conceiving mind .... Thus cognitive closure with respect to P does not imply irrealism about P." (ibid, p. 3)

We noted in Chapter 2, subsection 2.2.2 that the scientific realist is able to postulate the existence of a reality which lies beyond his perceptions on the grounds that such a hypothesis yields the best explanation of those very perceptions. However, McGinn's stance deviates from a genuine attempt at scientific realism on two counts. Firstly, he is prepared to postulate the existence of properties which we are not only unable to perceive, but which also fall outside the scope of our conceptual understanding. Secondly, he fails to see that the scientific realist must, at the very least be able to say what kind of evidence would support or refute the existence of any non-observable entities. Moreover, this evidence must be obtainable, at least in principle. McGinn, on the other hand, denies that introspection, conceptual analysis

or neuroscience can ever yield any evidence in support of his hypothesis. Consequently, any arguments that a hidden structure of consciousness may provide the best explanation for the situation described in his general hypothesis ( $H^1$ ) are cut off at a stroke. For example, McGinn writes that

"the property of consciousness is cognitively closed with respect to the introduction of concepts by means of inference to the best explanation of perceptual data about the brain."

(ibid, p. 13)

But if McGinn systematically disallows the possibility of any empirically testable inferences to the truth of his hypothesis ( $H_2$ ) that consciousness has a hidden structure, then his claim to have succeeded in showing that there is nothing metaphysically strange or problematic about the link between consciousness and the brain is unwarranted. By his own admission, we can never know anything about the properties of such a hidden structure. McGinn's assertion that there is nothing non-naturalistic about this structure is simply begging the very point at issue. Indeed, McGinn is hoisting himself on his own petard, since the very arguments he employs to show that we must be pessimistic about our chances of ever discovering the true explanation of the link between consciousness and the physical brain are the very arguments which must also preclude his

optimism concerning the character of this explanation, and the reason why his theory lacks testability.

### 3.2.5 Blindsight and blind judgment

McGinn attempts to mitigate these difficulties by providing us with some indirect evidence for the hidden structure of consciousness. Perhaps the phenomenon of blindsight will provide us with some support for his specific hypothesis that consciousness has a hidden structure thereby providing further confirmation of the original, more general, hypothesis,  $H_1$ , that consciousness plays an important role in our mental lives and is somehow rooted in physical processes in the mind/brain.

McGinn begins this section of his argument (McGinn, 1991, p. 110) by asserting that the need for a hypothesis of a hidden structure within consciousness would be obviated if it could be demonstrated that whenever the phenomenal features associated with any particular ability are absent, then that ability breaks down too. Here, McGinn does attempt to describe the kind of evidence that would refute his hypothesis. However, he can afford to be generous in this respect, since the findings of neuropsychological research present evidence that the reverse is sometimes the case. There are instances in which a person has no conscious awareness of being able to see and is yet able to



locate and identify shapes presented in their former visual field with a level of accuracy that is better than chance. This is the phenomenon of blindsight.

McGinn comments, correctly, (ibid, p. 111) that this phenomenon illustrates that the subjective experience of seeing is sometimes dissociated from the actual ability to see. Moreover, he continues

"I take it that what is most controversial in my interpretation of blindsight is the thesis that it demonstrates a hidden causal structure to conscious visual states *themselves*. It would presumably not be denied that *some* causal structure exists in common to ordinary sight and blindsight; the moot question is whether this structure is intrinsic to experience itself or merely exists alongside it" (ibid, p. 112).

McGinn is quite right to anticipate objections to this part of his interpretation of blindsight. Of course, it is not impossible that the causal structures responsible for the preservation of some of the blindsight subject's ability to see may be situated within that part of consciousness that McGinn claims is not open to introspection. However, the evidence by no means supports this interpretation unequivocally. Indeed, Weiskrantz, who was responsible for

much of the pioneering work in this field, is chary about drawing too many inferences concerning the nature of these structures. Whilst he admits that it is important for us to consider the dissociation of function from experience, he goes on to comment that this

"does not mean that philosophical questions regarding it are not tangled and complex, and we make no absurd claim here to find a ready answer to mind-body problems." (Weiskrantz, 1986, p. 165)

Further, it is interesting to note that one of Weiskrantz's few remarks concerning the implications of the gap between the verbal reports of blindsight subjects and their preserved visual abilities takes us in completely the opposite direction from the one advocated by McGinn. Weiskrantz writes that

"if it is universally accepted by ordinary men and women that we all have phenomenal experience, then what would be the evolutionary value of such a claim: not only for our belief in its existence, but even beyond to the reason for its very existence? The phenomenon of blindsight, and other "non-conscious" processes, does at least challenge one to speculate about this and

in doing so, to put the matter in a biological and hence scientific context." (ibid, p. 165, my italics)

This quotation demonstrates that it is equally possible to interpret the phenomenon of blindsight as evidence against McGinn's theory of consciousness. Weiskrantz suggests that blindsight is a *non-conscious* process, rather than one created by the preservation of the deeper structures within consciousness when its surface properties have been destroyed. Moreover, Weiskrantz implies that the discovery of such phenomena calls into question the importance that we have hitherto assigned to conscious experiences.

Consequently, McGinn's discussion of blindsight, far from strengthening his argument, may shed further doubt upon his initial hypothesis that consciousness plays such a major role in our mental lives. McGinn's failure to discuss the merits of such alternative interpretations of blindsight demonstrates a strong bias in favour of the story that will support his own hypothesis.<sup>11</sup> Once more, McGinn has systematically neglected to consider sources of empirical evidence that may refute his theory.

Further, McGinn's arguments in favour of his interpretation of blindsight are themselves not good. Once more, his insistence that we are constitutionally incapable of

discovering the precise nature of the deep structure of consciousness precludes the possibility of our ever being able to test his claim that such a structure is responsible for the preservation of visual capacities in blindsight subjects.<sup>12</sup> This means that the bootstrap strategy he employs is dealt a severe blow, for he needs to assume hypothesis  $H_2$  (the existence of a hidden structure within consciousness) in his interpretation of blindsight (evidence  $E_2$ ). McGinn is once more begging the very point that he is trying to prove.<sup>13</sup>

Secondly, McGinn relies upon linguistic arguments to support his view that the structures responsible for the preservation of the visual capacities of blindsight subjects are intrinsic components of consciousness. This strategy is perverse to say the least, since one of his main initial arguments for the postulation of a hidden structure within consciousness was based upon an analogy with the ways in which our linguistic expressions of our thoughts fail to provide an accurate reflection of the true logical structure of those thoughts. For example, McGinn writes that

"the apparent form of sentences actively tempts us into certain kinds of logical and metaphysical error; the surface of language naturally generates various intellectual illusions, which

can only be avoided by discounting or downplaying the appearances and acknowledging a level of hidden logical structure." (McGinn, 1991, p. 93)

Yet he comments elsewhere that

"[w]e explain my discriminative behaviour precisely by saying that I have visual experiences of certain kinds; they are the causal ground of my discriminative behaviour. We do not say that I have these experiences *and* there is some other causal source at work inside me" (ibid, p. 112)

By the lights of his earlier arguments, McGinn should acknowledge the possibility that our ordinary way of talking about visual experiences may be radically misguided, so that the very postulation of mysterious causal powers within a hidden structure of consciousness is itself a source of "metaphysical error" and "intellectual illusion". Hence, that part of evidence ( $E_2$ ) that is concerned with blindsight does not provide McGinn with any satisfactory support for  $H_2$ . His bootstrap argument is therefore proved to be unsatisfactory in yet another respect.

### 3.2.6 McGinn and the veil-of-perception issue

We saw in subsection 3.2.4 that McGinn maintains that our cognitive closure with respect to the hidden depths of consciousness does not prevent him from asserting the existence of these depths, nor from knowing that their properties bear no resemblance to previously suggested solutions to the problem of how consciousness can exist within a physical substance. His line of argument in support of H<sub>2</sub> involves the acceptance of all four of the following claims:

- i) there is a hidden deep structure of consciousness beneath the surface properties
- ii) this structure is causally responsible for much of our mental life
- iii) we only ever have direct awareness of the surface properties of consciousness; these introspectable surface properties mediate between the deep structure of consciousness and our day-to-day experiences and thought processes. We are constitutionally incapable of direct knowledge of the hidden structure of consciousness.
- iv) yet this hidden structure exists and that it

provides a naturalistic and non-occult explanation of the link between consciousness and the physical brain, since nothing else that we know of can account for this link.

McGinn's argument exhibits the very kind of difficulties that have traditionally been associated with the veil-of-perception issue (see Chapter 1, subsection 1.2.4). The four claims above form a semantically inconsistent tetrad similar to that isolated in the original veil-of-perception issue. They therefore fail to provide a satisfactory argument for the existence of a hidden structure within consciousness. Since we are constitutionally incapable of direct knowledge of such a structure (claim iii), we have no independent corroborating evidence for claims ii) and i) respectively, namely that this structure is causally responsible for so much of our mental life, or indeed, that it exists at all<sup>14</sup>. McGinn's argument therefore exhibits the difficulties that I identified with the strong or ontological version of the veil-of-perception issue (Chapter 1, subsection 1.2.4).

Further, even if we grant McGinn's assumption that consciousness does have a deeper hidden structure, we still have no reason to believe that this structure will provide a non-occult, naturalistic explanation of the link between consciousness and the brain (claim iv). Therefore,

McGinn's argument exhibits the weaknesses that I have associated with the weak or epistemological version of the veil-of-perception issue, as defined in Chapter 1, subsection 1.2.4.

Scientific realists must allow for the possibility that their hypotheses may be refuted as a result of empirical evidence against them that is discovered in the course of rigorous and systematic searches for evidence. In claiming *a priori* that we shall never know the structure of the hidden part of consciousness, McGinn fails to be a true scientific realist. Rather, he is a metaphysical realist who pays lip service only to the concepts of the modern philosopher influenced by science.

It is clear from the above that McGinn fails to understand the burden of proof that the scientific realist is required to provide in support of hypotheses which make reference to non-observable entities. His bootstrap argument has wholly failed to screen out hypotheses that are really based upon metaphysical assumptions rather than upon solid, empirically testable foundations.

McGinn regards his theory as an advance upon those that have postulated occult entities such as immaterial substances and disembodied minds (McGinn, 1991, pp. 105 - 107). Such Cartesian entities are, in his view, dreamt up



as a consequence of paying too much attention to the illusions created by introspection. However, given that McGinn admits that we have no way of knowing the nature of the deep hidden structure of consciousness, his own theory seems no less occult. Locke's arguments concerning the validity of our use of the general term "substance" would be just as appropriate here. Locke wrote that

"here, as in all other cases where we use words without having clear and distinct ideas, we talk like children; who, being questioned what such a thing is, which they know not, readily give this satisfactory answer that it is *something*; which in truth signifies no more, when so used, either by children or men, but that they know not what, and that the thing they pretend to know, and talk of, is what they have no idea of at all" (Locke, 1981, p.186).

It is hard to see precisely why McGinn is so adamant that his hypothesis that consciousness has a hidden structure will be so useful when we are necessarily incapable of discovering its nature.

### 3.2.7 Conclusion - McGinn the metaphysical realist

We are left with the uncomfortable impression that McGinn's self-avowed scientific realism is an illusion, and that his arguments contain more than a few unwarranted metaphysical assumptions. It is acceptable to postulate hypotheses that are as yet unproven, and equally acceptable to attempt to describe the benefits that will accrue for philosophers and scientists if the hypotheses turn out to be correct. However, it is not acceptable to extol the virtues of a hypothesis which contains within itself the seeds of its own inherent untestability. Nor is it acceptable to ignore in a systematic fashion interpretations of empirical evidence that tell against one's hypotheses. Such strategies bear the unmistakable stamp of metaphysical realism.

McGinn is aware to some extent of the metaphysical status of the assumptions that are incorporated into his theory. He writes that

"the metaphysical merits of the present conception should be allowed to overrule any discomfort it may induce in the epistemologically sensitive" (ibid, p.122).

However, he fails to realise the import of what he is

saying for his intended stance of scientific realism. This tension between the metaphysical realist and the scientific realist runs throughout McGinn's argument. Unfortunately, McGinn's neglect of the importance of systematic and rigorous testing of hypotheses within a bootstrapping argument result in victory for the metaphysical realist.

1. See also my comments on Popper in the Introduction to Section B, subsection B1.
2. This point will be made illustrated repeatedly in the course of this section of the thesis.
3. I shall justify this rather contentious statement in the course of this chapter.
4. See, for example, his comments in Could A Machine Be Conscious? (ibid, pp. 202-213) that a chemical substance in the brain

"would not be the kind of thing to explain consciousness" (ibid, p. 208).

These comments are not backed up by any kind of convincing philosophical arguments. Moreover, the lines of argument employed in Could A Machine Be Conscious? are very similar to those used in the chapter under scrutiny here. This indicates that the problems that I am about to highlight are endemic in McGinn's work.

5. Hofstadter and Dennett (1982) contains a useful bibliography of relevant literature.
6. This hypothesis is implicit in McGinn's whole discussion of why it is so important for us to be scientific realists about consciousness, rather than stated explicitly.
7. See, for example, McGinn's comments that "Nothing overt is up to the job [of accounting for the place of consciousness in the physical world].... There has to be more to consciousness

than there seems to be or else it could not depend upon the physical world in the way we know it does. As it were, nature could not create consciousness out of living matter without first constructing a substructure for consciousness to rest on; levitation is not an option" (ibid, pp. 100-101).

8. Compare my earlier comments in subsection 3.2 on the existence of thoughts that are not revealed by introspection.

9. However this does not stop McGinn from postulating some kind of direct link between the surface properties of consciousness and physical properties of the brain when it suits him. For example, he indulges in speculation as to what might happen to a person if "by surgery, we took away the surface structure of his thoughts and left him with only their hidden logical structure" (ibid, p. 113).

10. A point frequently made by many neurophilosophers eg: Paul Churchland. Paul Churchland comments (P. M. Churchland, 1984, p.160) that our introspective powers will become increasingly discriminating as we acquire more concepts with which we can describe our experience. Hence, eliminativism does not deny that consciousness exists as McGinn (1991, p. 122) implies; rather, it claims that the vocabulary that we use to describe our conscious experiences may be in need of revision.

11. This omission on McGinn's part is particularly striking in view of the central role played by Weiskrantz in the study of the blindsight phenomenon. Indeed, McGinn himself refers to Weiskrantz's work (McGinn, 1991, p. 110, n. 26).

12. This aspect of McGinn's arguments also prevents him from giving a satisfactory account of why functionalism is an inferior explanation of the phenomenon of blindsight (cf: McGinn, 1991, p.115).

13. Moreover, McGinn cannot claim that this hypothesis is simply the best explanation of blindsight that we have, since he has not argued sufficiently well against the alternative interpretation suggested by Weiskrantz. The most favourable outcome possible for McGinn under the circumstances is, once more, empirical underdetermination.

14. Recall that McGinn himself asserts that introspection gives us no awareness of this deeper aspect of consciousness, so the claim within statement (iii) that the surface properties of consciousness mediate between it and our experiences remains uncorroborated except for the assumption that there is no other possible explanation for much of our behaviour. However, as I have already mentioned, this assumption is the very point at issue and cannot therefore be regarded as evidence in support of the hypothesis without begging the question.

## Chapter 4 - Taking Fodor to task: Churchland, Mentalese and metaphysical realism<sup>1</sup>

### 4.1 Introduction

The purpose of this chapter is two-fold. Firstly, it is intended to yield further illustration of the point made in the preceding discussion of Colin McGinn's theory of consciousness (Chapter 3), namely, that the difficulties traditionally associated with metaphysical realism are not necessarily avoided through the mere use of a bootstrapping line of argument to derive specific and general theories about the mind/brain from the same evidence. Here, these issues are discussed in the context of Fodor's representational theory of mind.

Like Colin McGinn, Fodor relies chiefly upon conceptual analysis to establish his conclusions that sentential representations exist within the mind/brain and are causally responsible for our overt linguistic abilities. However, some attempts are made to consolidate his views by reference to psychological findings, notably in Fodor (1976). It will be demonstrated that these attempts by Fodor to provide evidence in support of his bootstrap argument are unsatisfactory, and do not exonerate him from the charge of metaphysical realism, notwithstanding his

claims to the contrary (see, for examples, my subsection 4.3 below).

The second, subsidiary aim of this chapter is to establish that the criticisms made of the Fodorean representational theory of mind by Patricia Smith Churchland amount to the claim that the theory suffers from the very difficulties relating to explanation that may be illustrated by the veil-of-perception issue, and that Fodor is therefore a metaphysical realist. Such an awareness of the questions associated with realism as they apply to other theorists of the mind/brain *should*, we hope, yield more positive efforts on Churchland's part to avoid being hoisted upon the same pétard when developing her own position. However, I shall show in Chapter 6, subsections 6.2.3 - 6.2.5 that her own regard for the principles of scientific realism lacks any real depth. I am using the Fodorean position as a stalking horse, so that Churchland's own ambivalent attitude can be more clearly exposed.

I shall begin by describing some of the main features of Fodor's Representational Theory of Mind, then go on to highlight the use of the bootstrapping argument in the creation of the theory. In the second half of the chapter, Churchland's criticisms of Fodor and the Fodorean line of defence will be examined, and then evaluated in the light of the discussion of realism, explanation and the

veil-of-perception issue in Section A. The chapter will conclude with a short summary of the contribution that this chapter makes to the overall argument of the thesis as a whole.

#### 4.1.1 What is mentalese?

In developing a representational theory of mind, Fodor's main line of argument is for the existence of a language of thought or mentalese in Fodor (1976). However, the basic theory is amplified and defended in later works (Fodor 1980, 1981, 1984b, 1985a, 1986a, 1986b, 1987), although its main features do not undergo any radical change.

Fodor adopts a stance consonant with realism towards mental representations or symbol systems, claiming that they literally exist within the mind/brain. Mental processes consist in computations which are defined over these representations. Computations are symbol manipulations which are only possible because the symbols possess a structured syntax by virtue of which they may be combined to generate further mental processes, or to instigate overt action or speech.

It is this symmetry between syntax and semantics that, in Fodor's view, makes a language of thought hypothesis possible. He writes that

"[w]hat makes the story a Language of Thought story, and not just an Intentional Realist story, is the idea that these mental states that have content also have syntactic structure - constituent structure in particular - that's appropriate to the content that they have" (Fodor, 1987, p. 137).

This symmetry also accounts for the success of explanations of behaviour in terms of the intentions of an agent (in other words, folk psychological explanations). Cognitive states consist of propositional attitudes identified via their relationship to statements in the language of thought.

Fodor stresses that the constituent symbols of the language of thought must be explicitly represented, even if the computational processes that make use of them are not. In Fodor's opinion, we can account for our ability to generate an infinite variety of inner mental states in exactly the same way as we explain our ability to utter an infinite number of different sentences; that is to say, in terms of the number of potential combinations of meaningful symbols that may be created through the use of diverse computational rules which operate upon their syntax. In the case of overt languages, of course, these computational rules are known collectively as a grammar.



One consequence of Fodorean theory is that it results in methodological solipsism (Fodor, 1980, p. 65). Although Fodor is adamant that mental representations *do* have semantic content and that it is *in principle* possible for us to construct a naturalistic account of how each mental representation acquires its meaning, he is equally convinced that the naturalistic account will *in practice* remain forever unknown to us. Since mental processes are computational, they only have access to the formal, syntactic properties of the symbols or representations that constitute the language of thought. No account is taken of the semantic content of these representations. Study of these processes could therefore never yield any information about the meaning of these states.

These features of the computational processes within the mind/brain force us not only to query the accuracy with which our mental states represent the world outside the mind/brain, but also to consider the more radical question of whether they can represent anything at all. These doubts correspond exactly to the weak and strong versions of the traditional veil-of-perception issue identified in Chapter 1 (subsection 1.2.4). If we have no testable explanation of how our representations come to have the semantic content that Fodor assures us they do, how can we ever trust them? Our concerns are compounded by Fodor's unfortunate comment that his theory

"is, in fact, a Good Old Theory - one to which both Locke and Descartes ... would certainly have subscribed" (Fodor, 1981, p. 26).

Fodor goes on to draw an explicit comparison (ibid, p. 26) between the mental representations or symbols involved in his inner language of thought and the "ideas" postulated in seventeenth century philosophy. This is disturbing, since such "ideas" were an essential component of the theories of Locke - one of the philosophers most frequently alleged to have fallen into the trap of the veil-of-perception (see also my comments on Locke in Chapter 1, subsection 1.2.5).<sup>2</sup>

As late as Fodor (1986), we find casual comments from Fodor about these difficulties. He writes

"there is a worry about how symbols(/beliefs) can be false. These are hard problems; but at least they're the *right* problems" (Fodor, 1986, p. 21).

However, Fodor does not appear altogether happy with this state of affairs, and in Fodor (1987) he attempts to put forward a naturalistic account of mental representations which retains the language of thought as an integral feature. His success in avoiding the difficulties associated with the impossibility of naturalism will be assessed in subsection 4.4.3, below.

## 4.2 Fodor's use of the bootstrap strategy

### 4.2.1 Fodor's support for the bootstrap strategy as a device for the development of psychological theories

Fodor is well aware that the bootstrap strategy has been used for a long time by psychologists with the express intent of deriving from one set of data both general and specific conclusions about the way that the mind works. He comments in the *Preface* of Fodor (1976) that

"[w]hat speculative psychologists did was this: They thought about such data as were available about mental processes, and they thought about such first-order psychological theories as had been proposed to account for the data. They then tried to elucidate the general conception of the mind that was implicit in the data and the theories" (Fodor, 1976, p. vii).

Fodor then goes on to associate himself with this methodology by declaring that

"[t]his book, in any event, is unabashedly an essay in speculative psychology" (ibid, p. viii).

Nor is Fodor unaware of the implications of the bootstrap strategy: namely, that empirical work must go hand in hand with theorizing. He argues that

"One wants to say: 'If our psychology is, in general, right then the nature of the mind must be, roughly, this...' and then fill in the blank. Given the speculative elucidation, the experimentalist can work the other way around: 'if the nature of the mind is roughly..., then our psychology ought henceforth to look like this:...', where this blank is filled by new first-order theories. We ascend, in science, by tugging one another's bootstraps" (ibid, p. ix).

Indeed, Fodor's overt response to the possibility that fundamental mistakes in our theories may be found as a result of the use of the bootstrap strategy is a good one. He argues (ibid, p. ix) that this possibility should be recognised as a real one, but should not prevent us from carrying out the work that will demonstrate these errors. However, when faced with evidence which tells against his own theory, Fodor's actual reaction is very different. He is reluctant to admit that either his specific or general theories of the mind/brain may be wrong<sup>3</sup>.

#### 4.2.2 The bootstrap line of argument within Fodor's Representational Theory of Mind

There is a clear bootstrapping argument within Fodor's theory. His general hypothesis  $H_1$  within his representational theory of mind, RTM, is that cognitive processes are computational<sup>4</sup>. This hypothesis is thought by Fodor to derive support from a variety of sources which together constitute a body of evidence that I shall term  $E_1$ .

Firstly, Fodor claims that all hitherto plausible specific theories of cognitive processes within the mind/brain have been based upon the notion of computation. He writes that

"[t]he only psychological models of cognitive processes that seem even remotely plausible represent such processes as computational....Remotely plausible theories are better than no theories at all" (ibid, p. 27).

Theories that do not make use of the computational hypothesis have, in Fodor's opinion, all been unsatisfactory. He regards behaviourism, for example, as incapable of providing us with a satisfactory explanation of human action, because it does not allow for the possibility that an agent has chosen to act to bring about one particular event rather than another. The behaviourist

ascribes causal powers only to actual events, and not to ones that are merely possible. Fodor writes that

"the behaviorist requires us to view considered behaviors as responses to actual inputs, when what we want to do is view them as responses to possible outcomes.

It is, conversely, one of the great advantages of computational theories of action that they allow us to acknowledge what everybody knows: that deciding what to do often involves considering what might turn out to be the case" (ibid, p. 33).

In Fodor (1986b), Fodor is still insistent that our ability to predict and explain human behaviour would not exist were it not for computational systems within the mind/brain. This is because such explanations and predictions are based upon propositional attitude psychology. Moreover,

"one can say in a phrase what it is that computational psychology has been proving so successful at: viz. the *vindication of generalizations about propositional attitudes*" (Fodor, 1986b, p. 3).<sup>5</sup>

The very existence of such a successful method of explaining and predicting human behaviour, and the failure of psychologists and philosophers to provide an acceptable

non-computational theory of cognitive processes together constitute two of Fodor's main sources of evidence ( $E_1$ ) for the general hypothesis ( $H_1$ ) concerning the computational nature of mind/brain processes.

In the light of his general position that any satisfactory theory of the mind/brain must make reference to computational processes, Fodor goes on to develop his specific hypothesis ( $H_2$ ) about the way that our mind/brain actually realizes these processes.  $H_2$  consists in Fodor's language of thought hypothesis and is arrived at by the following argument.

If we accept Fodor's general hypothesis  $H_1$ , and the evidence he cites in its favour,  $E_1$ , then cognitive processes are computational. However, it is impossible for any computation to take place if there are no representations or symbols within the mind/brain over which such computational processes can be defined. It is the syntax of these representations that facilitates their computational manipulation. Fodor comments that

"certain kinds of very central patterns of psychological explanation presuppose the availability, to the behaving organism, of some sort of representational system.....For, .... deciding is a computational process; the act the agent performs is

a consequence of computations defined over representations of possible actions. No representations, no computations" (Fodor, 1976, p. 31).

This representational system, in Fodor's opinion, constitutes a language of thought. Like a natural language, our internal representational system must be able to deal with hypothetical situations and to create a potentially infinite variety of scenarios from a large, but finite number of constituent symbols. What better way to describe the representational system within the mind/brain than as a language of thought? Fodor sums up his argument graphically as follows:

"OK, so here's the argument: Linguistic capacities are systematic, and that's because sentences have constituent structure. But cognitive capacities are systematic too, and that must be because *thoughts* have constituent structure. But if thoughts have constituent structure, then LOT [the Language of Thought thesis] is true" (Fodor, 1987, pp. 150-51).

Fodor believes that our undisputed ability to learn concepts furnishes us with yet another source of evidence in support of the view that the mind/brain operates in a computational fashion. In his opinion, we owe our ability



to learn overt languages, including our native language, to the existence of the language of thought or Mentalese. This is because the learning of a concept in an overt language such as English consists in the formulation and testing of hypotheses. If we do not yet know a concept in English, then it has to be introduced into our vocabulary via a Mentalese equivalent. Fodor comments that

"there is only one kind of theory that has ever been proposed for concept learning ... and this theory is incoherent unless there is a language of thought" (Fodor, 1976, p. 36).

The hypothesis and confirmation theory of concept learning constitutes part of the body of evidence ( $E_1$ ). As the previous quotation illustrates, Fodor combines this evidence with the general computational hypothesis ( $H_1$ ) to form an argument in support of the language of thought hypothesis ( $H_2$ ). He writes that

"the analysis of concept learning is like the analysis of considered choice; we cannot begin to make sense of the phenomena unless we are willing to view them as computational and we cannot begin to make sense of the view that they are computational unless we are willing to assume a representational system of considerable

power in which the computations are carried out" (ibid, p. 36).

Here, Fodor makes an explicit link between the general hypothesis ( $H_1$ ) and the specific language of thought hypothesis ( $H_2$ ) via the evidence provided by his view of concept learning. Concept learning is only explicable by reference to computation, but the language of thought hypothesis is needed to account for the representational complexity demanded by the computational processes involved. This is, of course, wholly in keeping with the traditional use of the bootstrap strategy to formulate a general hypothesis and a more specific hypothesis on the basis of just one set of data. (See also my earlier comments in subsection 4.2.1, and the discussion in Chapter 2.) If independent evidence could be adduced in support of the language of thought hypothesis, further confirmation would also be lent to the more general hypothesis concerning the fundamentally computational character of the mind/brain and the bootstrap argument would be complete.

Fodor's views about how we learn concepts commit him to two further claims about the nature of the language of thought. Firstly, it is innate. Secondly, since animals, and children who have not yet learned to speak are capable of acquiring concepts, a language of thought must be

attributed to them too, since such computational processes require representations, and

"the representational systems of preverbal and infrahuman organisms surely cannot be natural languages. So either we abandon such preverbal and infrahuman psychology as we have so far pieced together, or we admit that some thinking, at least, isn't done in English" (ibid, p. 56).

Fodor cites a mixture of reasons why we should be convinced of the truth of hypothesis  $H_2$ . This pot-pourri of reflections and supposed experimental support for the language of thought hypothesis constitutes  $E_2$ , the second body of evidence at work within Fodor's bootstrapping line of argument. The main thrust of this stage in Fodor's reasoning is an argument by analogy which is intended to demonstrate that the language of thought hypothesis must be correct because our inner representational system has so many similarities with natural languages. For example, Fodor writes that

"[t]he general idea is that facts about natural languages will constrain our theories of communication, and theories of communication will in turn constrain our theories about internal representations ... messages are most plausibly

construed as formulae in the language of thought" (Fodor, 1976, p. 109).

Fodor's basic idea is that, for a speaker and a listener to understand each other, there must be some correspondence between the mental states of one and the mental states of the other. When two people communicate, a process of coding and decoding messages is taking place. Fodor believes that this process is only possible if each participant in the communication process is able to construct representations of these messages which stand in systematic relation to the structure of the sentences which make up the messages. If no such representations exist, then there is nothing for our cognitive processes to manipulate in the course of interpreting the information carried by the messages.

Moreover, this representational medium must also be able to convey information from the senses in such a form that it can both be communicated to others and be used to confirm remarks uttered either by oneself or by others (ibid, pp. 111-112). Fodor suggests that

"[a]n obvious way to achieve this would be to translate all perceptual inputs into a common code and then define the confirmation relation for formulae in that code" (ibid, p. 112).

This common code is the language of thought. Fodor explains that

"if the kind of theory of communication I have been sketching is right, messages must be formulated in the language of thought; ie, they must be formulae in whatever representational system provides the domains for such cognitive operations as apply (inter alia) to linguistically carried information" (ibid, p. 115).

Fodor is anxious to argue that his specific language of thought hypothesis ( $H_2$ ) will, when combined with  $E_2$  (the body of evidence and arguments which supports  $H_2$ ), yield further confirmation of his general hypothesis ( $H_1$ ) that cognitive processes are computational. This would complete the bootstrapping cycle. In Fodor (1987), this move is made very clearly in two slightly different ways.

Firstly, Fodor simply states what a representational system like a language of thought actually makes possible. He argues that

"[i]f you think of a mental process - extensionally, as it were - as a sequence of mental states each specified with reference to its intentional content, then mental representations provide a mechanism for the construction of these sequences; they allow you to

get, in a mechanical way, from one such state to the next *by performing operations on the representations*" (Fodor, 1987, p. 145).

Secondly, Fodor looks at the problem from the other side and considers the consequences of rejecting the language of thought hypothesis. His conclusion is radical.

"As things stand now, the cost of not having a Language of Thought is not having a theory of thinking" (ibid, p. 147).

These passages provide a lucid illustration of the Fodorean stance. The language of thought hypothesis is not merely regarded as a consequence of the hypothesis that the mind/brain operates in a computational manner - it is also a prerequisite of computation<sup>6</sup>. Acceptance of the general hypothesis  $H_1$  (on the basis of evidence  $E_1$ ) entails acceptance of the language of thought hypothesis  $H_2$ , and this more specific hypothesis, combined with supporting evidence  $E_2$ , makes it possible to claim that computational processes occur within the mind/brain.

#### 4.3 Fodor's attitude to realism

It is one of the supreme ironies of Fodor's work that he repeatedly declares his respect for the view that all

hypotheses should be supported by satisfactory evidence, and yet fails to respond to criticisms of his theory in the same spirit. His apparently open mind about the ultimate plausibility of his language of thought hypothesis is revealed in this comment concerning a comparison of the language of thought with natural languages. Fodor claims that it is merely

"a suggestion that I regard as entirely speculative but very interesting to speculate about...." (Fodor, 1976, p. 156)

He continues

"It is pertinent to finish by emphasizing that these views may very well all be wrong ... The thesis I care most about is that claims .. about the character of internal representations are empirical in the sense that empirical data would tend toward their confirmation or disconfirmation" (ibid, p. 156).

Fodor is concerned to emphasize that he is not simply concerned to make sweeping statements to the effect that the language of thought hypothesis is correct and supportable by empirical evidence. Indeed, he sees his work as part of a larger, active research programme. He admits that

"it is not enough to argue that the notion of an internal language is conceptually coherent, that it is demanded by such cognitive models as sensible people now endorse, and that, in principle, claims about the structure of that language connect with empirical issues in psychology and linguistics. What now needs to be shown is that some progress can in fact be made in the assessment of such claims" (ibid, p. 122).

Unfortunately, Fodor does not always show himself to be so willing in practice to consider questions of evidence and to provide detailed arguments in support of his theory. This becomes apparent in Fodor's (1976) discussion of the linguistic evidence which he takes to favour the language of thought hypothesis. Fodor asserts that he wishes only

"to try to convince the reader that the internal language hypothesis is not, in the pejorative sense of the term, 'metaphysical': that there are factual considerations which constrain theories about the internal code. I shall therefore be content if it is accepted that the *kinds* of arguments I will rehearse are pertinent to the confirmation of such theories (ibid, p. 99).

Fodor's reluctance to claim any decisive authority for the specific evidence that he goes on to cite would be



commendable if it were combined with a genuine willingness to give due consideration to arguments and evidence brought forward against his theory of the nature of the mind/brain. However, he displays a tendency in all his work on the language of thought hypothesis to reject any arguments or data that conflict with his own views (see subsection 4.4.3 below for some examples of this Fodorean tendency). Fodor fails to acknowledge criticisms of his theory with the result that he creates the impression of systematically rendering it immune to tests or refutation. His insistence that he adopts the approach of a scientific realist to the issue of representation within the mind/brain is therefore misplaced. This point will be emphasised at length in subsection 4.4.4.

#### **4.4 Churchland's criticisms of the language of thought hypothesis**

##### **4.4.1 Introduction**

This part of my argument has two key objectives. Firstly, I shall demonstrate that the arguments of Patricia Smith Churchland against the language of thought hypothesis may be developed into the accusation that Fodor's representational theory of mind falls into the traps revealed in the earlier discussion (Chapter 1, subsection 1.2.4) of the traditional veil-of-perception issue.

Consequently, Fodor is a metaphysical realist and not the scientific realist he aspires to be.

Secondly, my use of Patricia Smith Churchland's arguments to criticize Fodor is intended to show that she is not unaware of the issues of testability that relate to the question of scientific realism.<sup>7</sup> The contrast between her ability to isolate Fodor's failure to adhere to the rigid standards of testability required and her own neglect of these standards will become apparent in the course of the remainder of this chapter and Chapter 6 (subsections 6.2 - 6.2.5).

Churchland does not object to Fodor's use of the bootstrap strategy within his theory of the mind to work from a general computational hypothesis towards a language of thought hypothesis which, if correct, would yield further confirmation of the computational hypothesis. There is absolutely nothing wrong, in her view, with working from a position that may ultimately turn out to be wrong, provided that independent evidence can then be produced in support of any conclusions drawn<sup>8</sup>.

But this is precisely where Fodor fails. It will be argued that Churchland is correct to castigate Fodor on the grounds that he is not entitled to draw the conclusions that he does solely on the basis of the hypotheses and

sources of evidence cited in the course of his bootstrap argument. This means that Fodor's use of the bootstrap strategy does not prevent his argument from suffering from the difficulties associated with the traditional veil-of-perception issue and with the failures of metaphysical realism. This line of argument will be developed in detail in subsection 4.4.4.

For the sake of clarity, my initial discussion of Churchland's criticisms of the Fodorean bootstrap strategy will be divided into two halves. The first half (subsection 4.4.2) will deal with her criticisms of Fodor's use of the general computational hypothesis ( $H_1$ ) and the supporting body of evidence ( $E_1$ ) to infer the more specific hypothesis ( $H_2$ ) of the language of thought per se.

The second half (subsection 4.4.3) of the discussion examines Churchland's analysis of the remaining portion of Fodor's theory, including the implications of  $H_2$  itself, the strength of the evidence contained in  $E_2$ , and Fodor's resulting claim that a combination of the two provide further confirmation of the truth of hypothesis  $H_1$  with respect to his overall representational theory of the mind, RTM.

In subsection 4.4.4 it will be demonstrated that Churchland's criticisms of Fodor's theory can be

reformulated to demonstrate Fodor's failure to avoid the difficulties that were associated with the traditional veil-of-perception issue in Chapter 1, subsection 1.2.4 and subsequently identified as characteristic of metaphysical realism in Chapter 2, subsection 2.2.1.

#### 4.4.2 Churchland's criticisms of Fodor's inference to $H_2$ from $H_1$ and $E_1$

It was noted above (subsection 4.2.1) that a sizeable proportion of Fodor's evidence ( $E_1$ ) in favour of his general computational hypothesis ( $H_1$ ) consists in the claim that such a hypothesis is needed in order to account for the success of common sense propositional attitude psychology. It is for this reason that Fodor comments that

"at the heart of the picture, the fundamental *explicandum*, is the organism and its propositional attitudes: what it believes, what it learns, what it wants and fears, what it perceives to be the case. Cognitive psychologists accept, that is, what the behaviorists were most determined to reject: the facticity of ascriptions of propositional attitudes to organisms and the consequent necessity of explaining how organisms come to have the attitudes to propositions that they do" (Fodor, 1976, p. 198).

Since Fodor believes that having a propositional attitude is quite simply being in a specific relation to an internal representation, then a systematic way of manipulating such representations is required to account for the systematicity of human behaviour. Hence, Fodor claims that

"only symbols have syntax, and our best available theory of mental processes - indeed, the *only* available theory of mental processes that isn't known to be false - needs the picture of the mind as a syntax-driven machine" (Fodor, 1985a, p.94, his *italics*).

This argument is thought by Fodor to establish the point that only a computational theory can perform the task of accounting for our systematic behaviour and thought patterns. However, although Churchland is prepared to accept both the general point that our cognitive processes are computational, and that there is a superficial regularity in our behaviour, she denies that these two facts taken together constitute adequate grounds for the inference that Fodor then makes to the existence of a specific, sentential language of thought (H<sub>2</sub>). Churchland objects to the use of the sentential paradigm as a model for all internal representation, and argues that Fodor has been seduced into ignoring the possibility that other, non-

linguistic forms of representation may exist within the mind/brain.

In Churchland's opinion, Fodor has overlooked valuable neuroscientific evidence and adopted an excessively top-down *a priori* stance upon what is really a wholly empirical issue. In an article written with her husband, Paul M. Churchland, Patricia Smith Churchland notes that

"[t]he bulk of cognition may take place in other sub-systems, and follow principles inapplicable in the linguistic domain. What those other representational systems are, and how they are knit together to form human cognition, these are empirical questions, begging empirical answers" (Churchland & Churchland, 1983, p.16).

Churchland's criticism of Fodor raises the possibility that Fodor may be confusing two different issues. The first of these is the undeniable fact that our descriptions of the internal representations of the mind/brain must, for the present, be couched in sentences, for we have no other means available to us. Whether these representations themselves take a sentential form is a second, wholly separate question. Churchland herself distinguishes clearly between these two very different points. On the first point, she remarks that, for example,

"within the confines of that very theoretical framework [of folk psychology] we are bound to describe the eliminativist<sup>9</sup> as believing there are no beliefs; however, this is not because folk psychology *is bound to be true*, but only because we are confined within the framework the eliminativist wishes to criticize and no alternative framework is available.....It would be foolish to suppose folk psychology must be true because at this stage of science to criticize it implies using it. All this shows is that folk psychology is the only theory available now" (P.S. Churchland, 1986a, p. 397).

However, this stance is by no means incompatible with the view expressed below about the nature of representations within the mind/brain. Churchland claims that

"[t]here is no doubt that some of the information-bearing states of the central nervous system are not a species of sentential attitude; that is, they are *not describable in terms of the person's being in a certain functional state whose structure and elements are isomorphic to the structure and elements of sentences*" (P.S. Churchland, 1980a, p. 147, *my italics*).

Churchland's point is lent additional support in Clark (1989). Clark, unlike Churchland, is not an eliminative materialist, and so does not share her view that folk psychological explanations of human behaviour should be rejected just as soon as we have found an alternative theory of mind to work with. However, Clark does share Churchland's conviction that Fodor's conclusions about the form of internal representations within the mind/brain are unwarranted<sup>10</sup>.

Clark's reasoning begins from slightly different premises from those employed by Churchland. He argues along the following lines. Fodor is wrong to claim that the systematicity of our thoughts is a contingent, empirical fact (Clark, 1988, p. 146). Rather, it is the very ability to use and understand concepts in a systematic fashion that allows us to ascribe thoughts to a person. Systematicity is therefore a conceptual fact which helps us to explain the undoubtedly empirical fact that we behave in a comprehensible manner (most of the time).<sup>11</sup>

Clark states his case thus:

"[w]hat stands in need of empirical explanation is not the systematicity of thoughts but the systematicity of the behaviour, which grounds thought ascription. Such systematicity indeed suggests recurrent and



recombinable elements. But there is no reason to suppose these to have to have a conceptual-level semantics" (Clark, 1989, p. 149).<sup>12</sup>

Fodor's failure to distinguish the conceptual from the empirical arguments within his body of evidence ( $E_1$ ) has resulted in his making an unwarranted inference about the nature of the representations within the mind/brain. Although thought is systematic in the way that Fodor describes, this fact is not explained by hypothesising the existence of a language of thought modelled upon a sentential paradigm. Such a hypothesis could only explain empirical facts, not conceptual ones of this kind. Indeed, this part of Fodor's argument exhibits one of the common faults in bootstrapping identified by Achinstein (1983) and discussed in Chapter 2 (above, subsections 2.4.1, 2.4.2), namely that of using evidence which is compatible with the hypotheses contained within a theory, but which is not actually explained by these hypotheses. Consequently, the first half of the Fodorean bootstrap argument is already upon shaky ground.

One of the key problems with the Fodorean hypothesis of the language of thought is that Fodor attempts all too frequently to support it via the same evidential paths that he uses to support his general computational hypothesis<sup>13</sup>. For example, Fodor argues that our ability to acquire

concepts is dependent upon the computational processes within the mind/brain. As we noted in subsection 4.2.2 (above) Fodor writes that

"the analysis of concept learning is like the analysis of considered choice; we cannot begin to make sense of the phenomena unless we are willing to view them as computational" (Fodor, 1976, p. 36).

Once more, Fodor attempts to justify his language of thought hypothesis by relying upon the line of argument that computation can only take place if a structured representational system exists within the mind/brain. Moreover, he has again employed the implicit assumption that this representational system must follow the sentential paradigm. The onus will clearly be upon Fodor to provide evidence in support of this model of the mind/brain in the second part of the bootstrap strategy<sup>14</sup>. See below (subsection 4.4.3) for an evaluation of Fodor's success in this task.

The flaws in Fodor's methodology might perhaps be mitigated if he could provide even one satisfactory source of evidence in support of his hypotheses. However, the Fodorean account of concept-learning is itself extremely problematic. It therefore rightly attracts much of

Churchland's critical attention, particularly in P.S. Churchland (1978) and P.S. Churchland (1980a).

It was recorded in subsection 4.2.2 that Fodor regards the learning of concepts in one's native language as a matter of formulating hypotheses and testing them. In other words, he believes that we learn our native language in exactly the same way as we later learn a foreign language. For example, when we learn German, we do so by forming biconditional hypotheses which match English predicates with predicates in German. For instance, we might formulate the hypothesis:

"Vater(x)" is true (in German) iff father(x) is true (in English)<sup>15</sup>.

That is to say, our comprehension of the German word "Vater" is dependent upon the realization that its usage corresponds exactly to that of the English word "father". Similarly, Fodor argues, we learn English by matching predicates in the inner language of thought (or Mentalese) with English predicates. For example:

"Father(x)" is true (in English) iff F(x)  
where "F" is a predicate of the language of thought.

This account of language learning therefore postulates an complete inner language of thought which predates our ability to express ourselves in our native language. As Churchland remarks

"the sum and substance of language learning thus turns out to be translation" (P.S. Churchland, 1978, p. 150, her italics).

Churchland is quick to give Fodor credit for his insistence that his hypothesis of the language of thought is not intended as an instrumentalist view of the mind/brain. That is to say, Fodor does not claim simply that concept-learning is better understood if we think of the activities of the mind/brain as if they operated upon sentential representations. Rather, as Churchland acknowledges,

"[t]he idea is that insofar as we take the organism to be going through a series of psychological states, sententially characterized, we mean that the structure and elements of his physical states stand in the relevant relations to the structure and elements of sentences" (P.S. Churchland, 1980a, p. 161).

However, Churchland makes it abundantly clear that Fodor must then live with the consequences of championing a realist hypothesis about sentential representations within the mind/brain. Many of these consequences will be discussed in subsection 4.4.3 (see also note 8 above). However two points must be made at once.

Firstly, I noted Fodor's belief (subsection 4.2.2) that we are often justified in attributing concepts to non-verbal animals and to children who have not yet learned to speak. Moreover, Fodor recognises that there are strong evolutionary arguments in favour of psychological theories that treat animals and humans in much the same way. For example, he comments that

"there are homogeneities between the mental capacities of infraverbal organisms and those of fluent human beings which, so far as anybody knows, are inexplicable except on the assumption that infraverbal psychology is relevantly homogeneous with our psychology" (Fodor, 1976, p. 57).

Fodor therefore believes that his language of thought hypothesis applies with the same force to animals and to preverbal children as it does to adult linguistic human beings. He holds that they too have sentential representations in their heads.

Here, Fodor is guilty once more of making unwarranted inferences from what may be valid premises. He ignores the possibility that the situation may be empirically underdetermined (see Chapter 2, subsection 2.2.2) with respect to his own hypothesis and rival, non-sentential views of representation within the mind/brain. For

example, Patricia Smith Churchland often shares Fodor's premises, but would not infer the same conclusions from them. She is prepared to recognize that processes within the mind/brain may be computational, and like Fodor, she argues (P.S. Churchland, 1980a, p. 160) that it is likely that the representational systems of man bear a strong resemblance to those of creatures lower down the evolutionary scale. However, Churchland disagrees with Fodor's conclusion that animals have concepts because they too have a sentential representational system. On the contrary, Churchland takes these very same evolutionary considerations to suggest that language is not a basic representational format in either animals or human beings. Fodor completely overlooks this possibility, thereby creating the impression that his inference is, at worst, wrong or, at best, not well supported by the available evidence.<sup>16</sup>

I now turn to the second problem inherent in the positing of sentential representations within the mind/brain. In Chapter 1, it was remarked that, for a theory or a hypothesis to provide a satisfactory explanation of an event or a situation, it must furnish us with new information about what is happening. Explanations cannot be informative if the explanans is merely couched in the same vocabulary as the explanandum. This renders doubtful their claim to be considered as explanations. Yet, it may

be argued that the Fodorean account of concept-learning itself fails to be explanatory because of this very failing.

Dennett (1981, p. 560) terms this kind of breakdown in explanation the *virtus dormitiva*, following Molière's example (cited in 1.2.2). Dennett comments that

"there can be no doubt that convicting a theory of relying on a *virtus dormitiva* is fatal to that theory, but getting the conviction is not always a simple matter" (Dennett, 1981, p. 57).

However, Fodor makes it remarkably easy to gain such a conviction. Churchland (1978, p. 153) provides us with the initial stages of the case for the prosecution in a shrewd comparison of the Fodorean account of concept-learning with primitive theories of the growth of the foetus in the womb. Such theories failed to provide an adequate explanation of foetal development because they simply assumed that it proceeded along the same lines as postnatal growth, despite the absence of any supporting evidence for this view. Indeed, Churchland argues that

"*prima facie*, whatever was going on *in utero* had to be radically different" (P.S. Churchland, 1978, p.151, *her italics*).

She goes on to ask

"whether the model of translation as poorly fits the learning of the mother tongue" (ibid, p.151).

Although Churchland makes an explicit and valid comparison between ancient theories of foetal development and Fodor's account of concept-learning, this is not in itself enough to establish that the Fodorean explanation falls foul of the *virtus dormitiva* objection. All that Churchland has established so far is that it is not good practice for philosophers or scientists to assume in the absence of independent evidence that unobservable processes are modelled upon observable processes. If we are to succeed in applying the *virtus dormitiva* objection as formulated by Dennett to Fodor's account of concept-learning, we require the additional premise that our ability to learn the concepts of our native language is *both explained by and caused by* our possession of a second language (the language of thought) within the mind/brain. However, this premise can easily be found within Fodor's own writings. For example, Fodor comments that

"one cannot learn a first language unless one already has a language" (Fodor, 1976, p. 64).



This shows that Fodor is indeed attempting to account for our ability to learn our native language by postulating the existence of an inner mental language containing equivalents of all the English concepts that we will ever need. His hypothesis amounts to a causal explanation of our ability to perform linguistic feats which is expressed entirely in terms of other inner linguistic feats. Consequently, Fodor's language of thought hypothesis has exactly the same explanatory status as a *virtus dormitiva*. Mentalese, like the hidden structure of consciousness so vaunted by McGinn (see Chapter 3), is simply one of those

"expressly occult qualities or faculties .. like little demons or goblins capable of producing unceremoniously that which is demanded, just as if watches marked the hours by a certain horodeictic faculty without having need of wheels, or as if mills crushed grains by a fractive faculty without needing anything resembling millstones" (Dennett, 1981, p. 56).<sup>17</sup>

It has been emphasized repeatedly in the course of this chapter that Fodor regards his account of concept-learning as a central feature of his body of evidence (E<sub>1</sub>) in support of a computational system within the mind/brain which requires sentential representations in order to operate. If Churchland's objections to this evidence can be

developed along the lines that I have suggested, then further doubt must be cast upon the hypothesis of a language of thought<sup>18</sup>. A related criticism of Fodor is made by Colin McGinn who comments that

"[t]he suggestion is that outer speech has meaning by being connected with propositional attitudes, so we cannot hope to explain what it is to have a propositional attitude by claiming that attitudes consist in inner speech... And if the inner saying theory can no longer be regarded as explanatory, the motivation lapses for insisting that we recognise the existence of a language of thought" (McGinn, 1982, p. 70).

Thus far we have established that there are grounds for doubting that Fodor is entitled to claim that his general computational hypothesis ( $H_1$ ), combined with the evidence he cites in its favour ( $E_1$ ), yield support for his specific hypothesis of the language of thought ( $H_2$ ). This claim has been criticised upon two grounds. First, that even if the evidence and the initial hypothesis are sound, they do not in themselves warrant Fodor's inference that there is a sentential language of thought within the mind/brain. Second, there are good reasons for believing that at least some of the evidence thought by Fodor to support his hypotheses is fundamentally flawed, in particular his

account of concept-learning<sup>19</sup>. In subsection 4.4.4 these difficulties in Fodor's theory will be evaluated in the light of the issues of realism, explanation and the veil-of-perception (see Section A), which must act as constraints upon the nature of any satisfactory theory of the mind/brain.

Moreover, I hope to have shown that Patricia Smith Churchland has been aware of all the major failings of the Fodorean argument isolated thus far. It will be seen in subsection 4.4.3 that she has identified still further flaws in the second half of Fodor's bootstrap argument. As noted earlier (subsection 4.4.1), Churchland's awareness of the difficulties inherent in the relationship between evidence and hypothesis within the Fodorean bootstrap strategy, and indeed, within the philosophy of psychology in general, contrasts sharply with her own failure to deal with these difficulties when evincing her own theory of the mind/brain.

I now turn to a critical examination of the second part of the Fodorean bootstrapping argument.

#### 4.4.3 Churchland's criticisms of Fodor's inference to $H_1$ from $E_2$ and $H_2$

It was mentioned in the previous subsection of this chapter that Fodor's account of concept-learning commits him to the view that the language of thought must be innate. If we learn the concepts of our native language through a process of translation from their Mentalese equivalents, then the question arises as to how we learned the concepts in Mentalese. Fodor cannot claim that these were learned in the same way as the concepts of our native languages without creating an infinite regress. Instead, he argues that the language of thought is innate.

The evidence that Fodor cites in the second part of his bootstrap argument ( $E_2$ ) (see subsection 4.2.2) is predominantly concerned to establish that, in many respects, the language of thought resembles overt languages. Fodor is clearly hoping to avoid objections to his hypothesis based upon the accusation that he is unable to characterize this inner language, and that therefore it cannot possibly be hypothesized.

Once more, Fodor's hypothesis of the language of thought may be criticised along lines suggested by Churchland's (1978) comparison of Mentalese with primitive theories of foetal development. Fodor is taking an unnecessarily

narrow view of the possible nature of representations within the mind/brain, which is similar to

"the narrowness ... in slavishly modelling growth in utero on growth ex utero" (P.S. Churchland, 1978, p.151).

Churchland's opinion, shared by many commentators on Fodor's theory, is that the body of evidence  $E_2$  is inadequate to the tasks that it is intended to perform, namely, supporting hypothesis  $H_2$  and combining with it to provide further confirmation of the truth of the general computational hypothesis  $H_1$  relative to Fodor's entire representational theory of mind.

For example, Churchland (P.S. Churchland 1978, 1980a, 1980b) emphasizes the sheer number of concepts contained within our native languages and the way in which language has evolved over the centuries in synchrony with the increases in our knowledge of the world. We now describe the world in an infinitely more sophisticated fashion than our ancestors did. In so doing, we have developed new concepts and discarded those that no longer reflect the way that we think about the world. The term "phlogiston", for example, no longer has any application<sup>20</sup>. In a nutshell, Churchland's position is that the capacity of our native languages to keep pace with our theories about the world

can only be accounted for by their flexibility. By contrast, the language of thought cannot, *ex hypothesi*, exhibit the same flexibility since it is innate and therefore fixed. Such a fundamental difference between our overt linguistic systems and the inner one postulated by Fodor can only cast grave doubts upon the validity of Fodor's argument by analogy which forms such a large part of his body of evidence ( $E_2$ ) in support of his language of thought hypothesis. Churchland writes that

"what is so implausible about Mentalese is that it is surpassingly rich, and that it has this richness independently of any learning processes...The richness together with the innateness constitute a rigidity, a resistance to modification, an imperviousness to learning from one's mistakes that is wholly uncharacteristic of the language of communication" (P.S. Churchland, 1978, pp. 151-152).

We can extend Churchland's argument to see how Fodor is caught between the horns of a dilemma. The first horn takes the following form: if the language of thought is to be rendered capable of formulating hypotheses containing new concepts so that English equivalents can be learned in the way Fodor describes, then Fodor is forced to give up the view that Mentalese is innate<sup>21</sup>. This would enable him to counter the criticism expressed by Churchland that

"we may be innately disposed to represent space Euclideanly, or to represent the sky as a vault, or to represent dolphins and whales as fish, and so on. But even if we are, I think it is fair to say that we have shown some alacrity and aptitude for reform" (P.S. Churchland, 1978, p. 153).

However, such a manoeuvre would remove the final obstacle to the criticism that Fodor's account of concept-learning suffers from an infinite regress. Fodor is only able to counter this objection by insisting that all concept-using creatures are born with a complete language of thought (see above, subsection 4.2.2). If this criticism were allowed to go through, then the Fodorean account of concept-learning would be even less satisfactory than it appears in the light of the arguments already evinced in subsection 4.4.2.

We now come to the second horn of the dilemma in which Fodor finds himself. Fodor can opt to accept that there is a yawning gap between the apparent rigidity of mentalese and the flexibility of our overt linguistic expression. However, this gap simply underlines the unlikelihood that our inner thought processes can be modelled along the lines of public language statements. Fodor's attempt to provide evidence for his language of thought hypothesis by using an

argument from analogy would therefore be dealt a severe blow by this admission.

Moreover, given that our native languages seem capable of representing our beliefs and judgements about the world in a satisfactory manner, the language of thought, by implication, could not achieve this. Yet, as Churchland rightly points out,

"for any Weltanschauung, for any structure allegedly exploitable to represent how the world is, it is pertinent to inquire whether it might not be in certain ways inaccurate, or unfaithful to the facts - whether how it *represents* the world might not be at odds with how the world *actually is*" (P.S. Churchland, 1978, p. 152, her italics).

If the language of thought is not capable of representing the world faithfully and of thereby providing us with a suitable supply of concepts with which we can describe it, then it is not doing the job for which it was deemed necessary in the first place. This greatly weakens the case in support of the hypothesis.

Churchland is not the only commentator to have noticed this fundamental flaw in the Fodorean enterprise of characterising the processes and representations within the



mind/brain. For example, Samet argues that Fodor's theory opens up

"the possibility that although the world that triggers our concepts might vary indefinitely, we remain trapped within the range of our innate endowment. We have only a limited spectrum of conceptual responses; we are not indefinitely flexible; we do not necessarily have the resources to accurately represent the variety that's out there in the world" (Samet, 1986, p. 592)<sup>22</sup>.

It was noted above (subsection 4.3) that Fodor displays a singularly cavalier attitude towards criticisms of his theory. This is perhaps even more irritating and saddening than the flaws in his argument, as his spurious defences of his hypotheses create the impression that he is systematically trying to place them beyond the reach of criticism and possible refutation. It need hardly be said that this is scarcely the stance that one would expect of a committed scientific realist.

A graphic illustration of Fodor's style of response to criticism of his language of thought hypothesis is contained in Fodor (1980). Here, Fodor is actually responding to points raised by Cohen in the peer commentaries section of the same paper, but it is easy to

imagine that he might counter Churchland's line of attack with the same words. He writes that

"[t]he system of mental representations will have to be precisely as rich as it has to be to explain the data, and not one whit less so. I'm very much afraid that 'put up or shut up' is the only form of serious argument in this area" (Fodor, 1980, p. 100).

Fodor's response to Cohen's reasoning is most worrying. This statement in support of his hypothesis hardly amounts to a well-constructed argument in favour of his theory! Faced with the accusation that it is implausible to postulate a wholly innate, yet very rich and flexible language of thought, it is simply not enough to claim that this is *the only possible explanation* of the data available. Indeed, Fodor is once more dangerously close to the objection that his argument is circular. If the richness of the language of thought is the only possible way to explain the data, and yet the only evidence for the richness of the language of thought is this very same data, then the circle is complete.

Fodor requires independently obtained evidence to explain how the language of thought can be so rich but, as we have seen, he has been completely unable to reconcile the

abundance of concepts contained within Mentalese with its status as an innate representational system.

It is clear that Fodor's comparison of the language of thought with overt languages has utterly failed to yield the unequivocal support for his sentential hypothesis that he anticipated. The relationship between the body of evidence  $E_2$  and hypothesis  $H_2$  within the bootstrap argument therefore breaks down.

Moreover, the plausibility of the intended comparison is decreased still further by Fodor's insistence (see, for example, Fodor 1980) that the syntactic relations between the representations of the language of thought allow computational processes to operate within the mind/brain, but cannot account for how the representations have meaning. This is the formality condition of computational accounts of the mind/brain, which precludes such accounts from ever yielding an explanation of how the inner language can accurately represent the world. Yet, as we have just seen (see also subsection 4.2.2), it is paramount that the language of thought should be able to do so, if it is to account for our ability to learn concepts and apply them correctly within the world.

Fodor argues (Fodor, 1980) that only a naturalistic psychology can provide a theory of how the language of

thought can represent the external world. However, such a theory would require a wholly physical vocabulary to characterise nomological relations between an organism and its environment. As a consequence of this, Fodor concludes (Fodor, 1980, pp. 70-71) that such a naturalistic psychology will not be possible until all nonpsychological, physical sciences have been completed and can yield such a vocabulary. That is to say that, *in practice*, naturalistic psychology is impossible. Nonetheless, Fodor claims that these conclusions should not lead us into fears of solipsism. He comments that

"[m]ethodological solipsism isn't, of course, solipsism *tout court*. ... Heaven only knows what relation between me and Robin Roberts makes it possible for me to think of him (refer to him, etc.), and I've been doubting the practical possibility of a science whose generalizations that relation instantiates. But I *don't* doubt that there *is* such a relation or that I do sometimes think of him. Still more: I have reasons not to doubt it; precisely the sorts of reasons I'd supply if I were asked to justify my knowledge claims about his pitching record ... nothing in the preceding tends to impugn these truths" (Fodor, 1980, p. 71).

Fodor's views about the formality condition are shared by other philosophers and psychologists of widely differing persuasions, including Paul and Patricia Churchland<sup>23</sup>. Nonetheless, his claim that a naturalistic psychology is (in practice, at least) impossible is disastrous for his language of thought hypothesis which was postulated in the first instance to explain how we are able to learn concepts which accurately reflect the world. If no account of how this is actually done is possible, then the case for the language of thought is once more greatly weakened<sup>24</sup>. How can we ever be sure that the language of thought even contains any references to an external world whatsoever? The traditional veil-of-perception issue recurs before our very eyes! The point is made graphically clear in this lengthy quotation from Harman's response to Fodor (printed in the peer commentaries to Fodor 1980). Harman writes that

"[w]hat leads us to attribute mental representations to a creature in the first place is that the creature is aided in achieving its presumed goals by being able to detect, come to know about, certain features of its situation. We study its capacity for mental representation first by investigating what features it can detect and what it cannot detect; then we study what sort of mistakes it can make that would lead it to act inappropriately. At that point we may be able

to form hypotheses about the content of some of its mental representations, perhaps even hypotheses about the form or syntax of these representations. *But there is no way we could ever come to these hypotheses without considering how the creature's mental states are connected with things in the outside world*" (Harman, in Fodor, 1980, p. 81, my italics).

It appears that Fodor has recognised that his language of thought hypothesis is in difficulties if no naturalistic psychology is possible, for his (1987) contains an attempt to construct just such a psychology. He comments that

"[w]e really do need to know at least roughly what [a naturalist psychology] might be like, on pain of having the metaphysical worry that - excepting psychophysical concepts<sup>[25]</sup> - we have no idea at all what a naturalized semantics would be like for the nonlogical vocabulary of Mentalese" (Fodor, 1987, p. 118).

Fodor believes that some concepts, which he terms "psychophysical", will automatically occur to us in the presence of the right kind of causal conditions. "Red" is one such concept. Fodor writes that

"there are circumstances such that red instantiations control 'red' tokenings whenever those circumstances obtain; and it's plausible that 'red' expresses the property red in virtue of the fact that red instantiations cause 'red' tokenings in those circumstances; and the circumstances are nonsemantically, nonteleologically, and nonintentionally specifiable" (ibid, p. 112).

Fodor claims that a naturalistic semantics can also be given for other more complex concepts, including new scientific ones such as "proton" (which, it will be recalled, have already caused major difficulties for the language of thought hypothesis). Crudely speaking, his theory amounts to the view that a proton will activate the concept "proton" in the mind/brain by creating effects that are describable in terms of psychophysical concepts. For example, the presence of protons in the environment may turn a photographic plate red (Fodor, 1987, p. 120). Fodor readily admits that the concept of "proton" will only be activated in the mind/brain of someone who knows that protons have this effect upon photographic plates, but argues that this does not affect the possibility of a naturalistic semantics. Fodor asserts that .

"[f]or purposes of semantic naturalization, it's the existence of a reliable mind/world correlation that

*counts, not the mechanism by which that correlation is effected*" (ibid, p. 122, his italics).

Several points need to be made at this juncture. The first is that Fodor's naturalistic account of semantics seems to have the same explanatory status as a *deus ex machina*. Fodor needs some kind of causal relation between mental representations and the world to get his theory out of a tight spot so, lo and behold, he postulates that one exists, even though we cannot actually know its exact nature. Clearly, the criticisms I levelled at McGinn's hypothesis of the hidden structure of consciousness in Chapter 3 (see especially subsection 3.2.4) and at the language of thought as an explanation of concept-learning in subsection 4.4.2 are equally applicable to this part of Fodor's argument. Fodor's failure to find a solution to the problem of how we are to account for the meaning of our mental concepts is far more extreme than any of his caveats in Fodor (1987) imply (see for example, p.118). Fodor's actual progress from his (1980) stance is minimal.

The explicit aim of Fodor's (1987) treatment of this issue is to disprove the sceptic's claim that it is impossible even in principle to construct a naturalized semantics. He is not trying to provide a cast-iron argument that such a semantics actually exists. Fodor's limited objectives may partially explain his failure to shed very much light upon



the problem for practical purposes<sup>26</sup>. This brings me to my second point. Fodor allows that a proton will only trigger the Mentalese concept "proton" within the mind/brains of those who have theoretical knowledge about protons and their effects upon the environment. However, Fodor has still not provided us with a satisfactory account of how we could ever acquire an understanding of modern concepts like "proton" given that the language of thought is innate. (See my earlier criticisms in this subsection). It is therefore unclear how any of us could ever know that the appropriate circumstances for a tokening of the Mentalese concept "proton" were being instantiated.

Finally, Fodor does not seem to have realized that the relationship between the existence of a satisfactory naturalized semantics and the fate of Mentalese is not symmetrical. I have argued that the absence of a satisfactory naturalistic psychology is bad from Fodor's point of view since it deprives Fodor of yet another source of evidence that some system exists within the mind/brain which does the job of representing the world that the language of thought was hypothesised to do. However, even if a naturalistic psychology of how our mental representations are causally related to the world does ultimately prove to be attainable, this would not necessarily vindicate a sentential representational system.

For example, P.M. Churchland and P.S. Churchland's (1983a) argues that

"if we want to know how cognitive creatures hook up to the world they inhabit, neuroscience holds out the best hope for an enlightening account" (p. 17).

Before I turn to a consideration of how the arguments against Fodor put forward in this subsection and the previous one demonstrate that his representational theory of mind does indeed suffer from the problems of explanation and metaphysical realism delineated by the traditional form of the veil-of-perception issue, one final remark needs to be made with respect to the second half of Fodor's bootstrap strategy.

Fodor hopes to derive further support for his computational view of the mind from the success of his language of thought hypothesis. More specifically, it is the systematic relations between the representations within the language of thought which he thinks provide the support he requires to complete the bootstrap argument. Fodor declares that

"arguments that suggest that mental states have constituent structure ipso facto favor Turing/Von Neumann architectures, which can compute in a language

whose formulas have transportable parts" (Fodor, 1987, p. 139).

Clearly, the problems outlined above with regard to the viability of Mentalese make it difficult for Fodor to continue to assert that he can obtain support for his computational hypothesis within his theory of mind from the more specific hypothesis relating to the existence of the language of thought. This is not to say that mental processes are not computational, but that the demise of the language of thought hypothesis renders it perfectly possible that other, rival computational hypotheses may be equally viable. Perhaps images are manipulated, rather than languages with a constituent structure, or perhaps as Clark (1989) suggests, computation may take place along connectionist lines - a possibility acknowledged, but not taken seriously enough by Fodor (Clark, 1989, p. 150). In any event, Fodor fails to provide us with sufficient independently obtained evidence for his claim that the language of thought hypothesis ( $H_2$ ) should increase our confidence in the truth of the more general computational hypothesis ( $H_1$ ), and the bootstrap argument breaks down once more.

#### 4.4.4 Fodor unmasked - the metaphysics behind Mentalese

It will be recalled that one of the key problems isolated in Fodor's development of the representational theory of mind was the precise way in which the language of thought could really be said to represent the world and to yield adequate concepts for us to describe this world subsection 4.4.3). The root of this difficulty lies in the innateness of the language of thought, although Fodor's claim that a naturalistic psychology is not currently possible may be regarded as a contributory factor. Nonetheless, it is worth repeating Churchland's remark that

"for any Weltanschauung, for any structure allegedly exploitable to represent how the world is, it is pertinent to inquire whether it might not be in certain ways inaccurate, or unfaithful to the facts - whether how it *represents* the world might not be at odds with how the world *actually* is" (P.S. Churchland, 1978, p. 152, her italics).

This quotation could easily be adapted to question the accuracy of Fodor's view of the representational system of the mind/brain, by simply replacing the word "world" with the words "mind/brain". We have already noted in the Introduction to Section B (subsection B2) that those who choose to investigate the workings of the mind/brain should

be no less constrained by the need to provide satisfactory evidence in support of their theories. We may therefore ask whether Fodor's representational theory of mind is an accurate account of what is really going on inside our mind/brains.

In Section A, it was argued that the problems of explanation associated with metaphysical realism are exemplified by those theories that exhibit the same kind of semantic inconsistency as that identified within the traditional veil-of-perception issue. It is now time to see in what way the arguments put forward against Fodor in the previous two subsections (many of which were influenced by Churchland) can be construed as an attack upon a metaphysical realist whose theories suffer from a veil-of-perception type of defect. This subsection will indicate not only that Fodor's bootstrap strategy has failed, but also that it has failed because Fodor adopts a metaphysical realist's attitude towards the relationship between the bodies of evidence and the hypotheses within his theory, rather than that of a scientific realist. In the next subsection, I shall attempt to suggest reasons why this might be so.

It is easy to reconstruct Patricia Smith Churchland's arguments against Fodorean Mentalese as a renewed attack upon the metaphysical realism raised by the

veil-of-perception issue. To see this, we have only to compare the following outline of Fodor's views with the central features of the traditional representational theories of perception I highlighted earlier (subsection 1.2.4). Fodor takes the following statements to be true:

I) there is an inner language of thought over and above any overt linguistic skills in human mind/brains and those of many dumb animals

II) this inner language of thought stands in relation to our overt mother tongue as cause does to effect, since we can only learn a language like English by matching its concepts with those of the inner language

III) we are only ever directly aware of our acquired linguistic capacities; the existence of mentalese is recognised only because these acquired linguistic capacities reflect it, and because the success of our commonsense computational and psychological explanations and our abilities to learn concepts cannot be otherwise accounted for

IV) yet we nevertheless have an accurate idea of what this inner language of thought must be like, since our overt linguistic capacities are both caused by and exact reflections of our mentalese faculties.

All of the claims made in these four statements occur at some point within the Fodorean bootstrap argument. However, it will be instructive to clarify the precise relationship between statements I-IV above and the individual hypotheses and bodies of evidence that together constitute Fodor's Representational Theory of Mind.

Statements I and II represent a summary of the specific language of thought hypothesis  $H_2$ , together with Fodor's main reason for postulating it, namely that it accounts for how men and animals have concepts. (Compare Fodor's comment, also quoted above, subsection 4.2.2, that "there is only one kind of theory that has ever been proposed for concept learning ... and this theory is incoherent unless there is a language of thought" (Fodor, 1976, p. 36).

Statement III acknowledges that it is not possible for us to have direct knowledge of the inner language of thought, but reiterates Fodor's two main sources of evidence for the claim made in statements I and II that such a language of thought exists. These are the success of our commonsense psychological explanations yielded by computational processes (a combination of hypothesis  $H_1$  and evidence  $E_1$  within the bootstrap strategy, see also subsection 4.2.2, above) and the argument from analogy with overt languages (the body of evidence  $E_2$ , see subsection 4.2.3). Finally, statement IV embodies a claim fundamental to Fodor's

bootstrap strategy, namely, that that these sources of evidence provide us with adequate indirect knowledge of the nature of the language of thought.

These four statements, like their counterparts in my reconstruction of the traditional formulation of the veil-of-perception issue (Chapter 1, subsection 1.2.4) form a semantically inconsistent tetrad. The semantic inconsistencies between the statements can be seen easily if the arguments against the Fodorean representational theory of mind put forward in subsections 4.4.2 and 4.4.3 are considered.

There are two levels of semantic inconsistency within the Fodorean tetrad, just as there were within the traditional veil-of-perception issue. Both of these levels are revealed by the arguments evinced in the previous two subsections (many of which, it will be recalled, have been developed from lines of attack suggested by Churchland).

The first inconsistency is revealed by a critical examination of the evidence cited by Fodor in statement III in support of his claim that the language of thought exists (statements I and II). Fodor readily admits that we can only have direct awareness of our overt languages, and the other evidence cited in statement III in support of his



language of thought hypothesis was shown to be inadequate in subsections 4.4.2 and 4.4.3 above.

In subsection 4.4.2, it was argued that our success in explaining the systematicity of human behaviour does not depend upon the existence of a sentential representational system within the mind/brain, that the Fodorean account of concept-learning is itself not well supported by sufficient evidence, and that evolutionary considerations do not necessarily suggest that all creatures with an apparent ability to reason or to exhibit behaviour appropriate to a given situation have a sentential language of thought.

In subsection 4.4.3, it was concluded that Fodor's other body of evidence in support of the language of thought hypothesis, an argument from analogy with overt languages, fails because of the incongruities implicit in the Fodorean view that an innate language of thought can contain all the concepts of our overt, flexible languages and thereby represent the environment with any degree of accuracy.

This means that the criteria evinced for the existence of the language of thought in statement III are unsatisfactory, so that this statement, when fully evaluated, is incompatible with statements I and II. This semantic inconsistency constitutes the strong or

*ontological*<sup>27</sup> version of the veil-of-perception issue as it occurs within Fodor's representational theory of mind.

Secondly, even if we permit Fodor to postulate the existence of some kind of language of thought within the mind/brain, the considerations put forward in subsections 4.4.2 and 4.4.3 indicate that Fodor is still not entitled to make the further assertion that it resembles our overt languages. The breakdown of the argument from analogy discussed in subsection 4.4.3 alone renders such an assertion dubious. Moreover, subsection 4.4.2 contains more than one illustration of Fodor's habit of drawing unwarranted conclusions from premises that may, in fact, themselves be valid. Indeed, it was remarked that Fodor sometimes makes inferences which are consonant with the evidence he is discussing, but which do not actually *explain* this evidence. Thus Fodor's failure to provide satisfactory evidence which tells in favour of his own view of the representational system within the mind/brain and against rival theories renders him open to the weaker, *epistemological* version of the veil-of-perception issue (see Chapter 1, subsection 1.2.4). Churchland sees this point clearly. She comments that

"[t]he onus of proof rests on the sententialist to provide evidence for taking various neural structures to process information in accordance with his favoured

model of information processing" (P.S. Churchland, 1980a, p.165).

We can see that the hypothesis of the language of thought put forward in statements I and II and claim IV, that the nature of this representational system can be readily inferred from the sources of evidence mentioned in statement III cannot be upheld in the light of the criticisms of Fodor's theory in the previous two subsections. Additional evidence would be needed for the language of thought hypothesis to withstand these criticisms, but this evidence is not forthcoming.

The conclusion that Fodor's representational theory of mind suffers from the difficulties associated with metaphysical realism discussed in Chapter 2, subsection 2.1 (despite Fodor's protestations to the contrary noted in subsections 4.1, 4.3) follows easily from the reconstruction of his argument on the veil-of-perception model.

The arguments evinced in this chapter demonstrate clearly that Fodor is attempting to assert the existence of a language of thought within the mind/brain to explain our systematic thought and behaviour, and our ability to learn concepts in the absence of any independently obtained evidence in support of this assertion.

This pattern of explanation is identical to that employed by metaphysical realists (see also Chapter 2, subsection 2.2.1). The metaphysical realist takes it to be a prerequisite of scientific activity that the world exists more or less as we perceive it. Fodor takes the existence of a language of thought to be a prerequisite of our undisputed linguistic ability. Metaphysical realists presuppose a correspondence theory of truth, so that the truth of our perceptions is determined by their accuracy in representing the world. Similarly, Fodor thinks that our mental representations provide an accurate reflection of the external world.<sup>28</sup>

Indeed Fodor's only real argument for the existence of the language of thought is based upon its ability to explain our linguistic capacities and systematic thought, but this line of argument clearly begs the very point at issue. Moreover, Fodor systematically disregards sources of possible empirical evidence against his views and persistently ignores all criticism of his arguments, thereby rendering his theory increasingly untestable and immune to refutation (see above, subsections 4.3 and 4.4.3 for details of Fodor's evasive tactics).

The moment that Fodor adopts an approach that is genuinely congruent with scientific realism and allows evidence which contradicts his hypothesis to be brought forward, his

theory is likely to be refuted or, at the very least, unsatisfactory to the same degree as other theories which provide an equally satisfactory explanation of the observable facts.

#### 4.5 Fodor, the bootstrap strategy, and the problem of empirical underdetermination

##### 4.5.1 Introduction

It was argued in Chapter 1 (subsection 1.2.5) that the problems of explanation epitomized by the traditional veil-of-perception issue arise from the lack of a suitable methodology from making justifiable and testable inferences from experience to unobservable systems. Fodor needs just such a methodology to permit him to make inferences about the nature of representational systems within the mind/brain from the overt sources of evidence readily available to him. Like Glymour (see Chapter 1, subsection 1.2.5, also Chapter 2, subsections 2.3, 2.3.1), Fodor turns to bootstrapping as a way of providing well-supported explanations of observable facts in terms of entities or processes that cannot be directly experienced (see also subsection 4.2.1).

However, the use of the bootstrap strategy is patently not in itself a sufficient guarantee of a well-confirmed

theory. If it were, then Fodor would not be in the difficulties that he is.

#### 4.5.2 Empirical underdetermination in Fodor's representational theory of mind

In subsection 4.4.2, it was noted that Fodor has a marked tendency to dismiss the possibility that his specific conclusions may be drawn from premises that actually lend equal support to rival hypotheses. Indeed, there is a high degree of empirical underdetermination in this area of study (see also Chapter 2, subsections 2.2.2, 2.4).

One example of empirical underdetermination is to be found in the possible consequences of evolutionary factors for the question of representation within the mind/brain (see above, subsection 4.4.2). Churchland and Fodor agree that the available evidence suggests that the representational systems within the human mind/brain may, in many respects, resemble those found in other species (particularly, other primates such as chimpanzees). Yet, as we saw in subsection 4.4.2, Churchland regards this as a reason for rejecting the sentential paradigm as a basic form of representation in either animals or in human beings, whereas Fodor confronts the issue by ascribing a language of thought to any animal he deems capable of entertaining concepts. Similarly, I have argued (following Clark, 1989)

that the undisputed systematicity of human behaviour does not necessarily require a sentential system of representation within the mind/brain. Clark suggests that other forms of representation such as those contained in a connectionist network may be able to account for our behaviour in an equally convincing manner.

Glymour (1980) has suggested that the bootstrap strategy is capable of circumventing such problems of empirical underdetermination, since it is unlikely that two theories produced by this line of argument will be equally acceptable when are examined in the light of a welter of abstract criteria. (See Chapter 2, subsection 2.4). When faced with two theories which apparently yield equally plausible explanations of a body of data, we should follow a policy of endorsing the theory that best conforms with general methodological criteria such as: prefer theories containing confirmed hypotheses rather than ones containing disconfirmed hypotheses, prefer theories that provide a uniform explanation of phenomena and that have been tested against a wide variety of evidence. However, Glymour's criteria have been shown by both Achinstein (1983) and Nordby (1989) to be vulnerable to criticism and abuse. These issues were discussed in Chapter 2 (subsections 2.4.1, 2.5.1). The next two subsections will demonstrate their relevance to the Fodorean use of the bootstrap methodology.

### 4.5.3 Achinstein's criticisms of the bootstrap strategy as they apply to Fodor's representational theory of mind

I remarked in Chapter 2 (subsection 2.4.1) that Achinstein attributes much of the initial plausibility of the Glymourian bootstrap methodology to the fact that the evidence used *does* fit with the hypotheses of the theory, although it can only support them relative to that theory. In fact, Achinstein argues, this methodology presupposes the truth of the theory and therefore begs the very point at issue. Once more, a supposedly scientific methodology falls into the difficulties associated with metaphysical realism.<sup>29</sup>

This flaw is clearly also visible in the Fodorean bootstrap strategy and was demonstrated in subsection 4.4.4 of mind to be the root of the veil-of-perception problems with which his theory is afflicted. Fodor's evidence in support of his language of thought hypothesis is only acceptable given the framework of his overall theory. Indeed, it was argued in subsections 4.4.2 and 4.4.3 that evidence which conflicts with the very foundations of this theory has been systematically ignored, despite Fodor's assertions (see above, subsection 4.3) that his hypotheses are empirical and may be either confirmed or disconfirmed by a wide variety of evidence.



This illustrates the point made in Chapter 2 (subsection 2.4.3) that it is often possible to find examples of philosophers and psychologists making use of criteria which implicitly favour their own theories whilst claiming to use the abstract methodological criteria for selecting the best theory suggested by Glymour.

#### 4.5.4 Bootstrapping in Fodor: the Nordby paradigm

These problems should make us question the appropriateness of general methodological criteria for the evaluation of bootstrapping arguments in the philosophy of mind. Furthermore, we must ask whether other criteria are actually being employed instead.

It was explained in Chapter 2, subsection 2.5.1 that Nordby (1989) extends the application of the bootstrap methodology to cover the deliberations of detectives, who may employ it to account for existing evidence and to predict what other evidence would be forthcoming if a crime is viewed as a token of a particular crime type. However, Nordby rejects Glymour's abstract criteria for choosing between crime type models in cases of empirical underdetermination as unhelpful in the absence of wider theories which prescribe strategies for deciding between competing models or for weighting evidence that originates from a diverse number of sources (eg: forensic, eye-witness accounts, expert

witnesses, etc). It is worth repeating Nordby's comments that

"[a]pplying Glymour's formal criteria for theory choice does not provide firm grounds for crime model choice. Ultimately the choice depends on deciding that certain lore, or natural or social sciences apply, and that is not a decision based on formal grounds" (Nordby, 1989, p. 385)

and that

"[t]he issue instead remains which science(s), or other components composing the models, should apply to the problem" (ibid, pp. 382-3).

Subsection 2.5.2 of chapter 2 contends that the distinction made by Nordby between bootstrapping in the philosophy of mind and bootstrapping in detective work may not be as pronounced as he suggests. In the study of the mind/brain, just as in detective work, relevant information come increasingly from a wide variety of different fields, each with their own standards of evidence. Co-evolutionary philosophers must therefore develop a strategy for evaluating this information and for determining the relative importance of each field and its criteria for reliable evidence. No such strategy has been established

as yet, so like Nordby's detective, the philosopher of the mind/brain bootstraps while barefoot.

Fodor rejects a fully co-evolutionary approach to the study of central information processing in his (1976), although he does accept that there is some link between his sentential representations and their neurological implementation.<sup>30</sup> At the time when *The Language of Thought* was written, philosophers took relatively little interest in neurophysiological studies, but Fodor, disturbingly, has made no attempt to revise his theory or to consider the neurophysiological evidence brought forward against his theory in more recent times, even in Fodor (1987)<sup>31</sup>. He continues to rely on the conceptual arguments that he evinces against the possibility of reduction in Fodor (1976).

In the absence of any overarching theory of what constitutes relevant and sound evidence, Fodor finds himself beset by exactly the same difficulties as his more committed co-evolutionary counterparts. We are frequently given the impression that many of the gaps in his evidence for the hypotheses that constitute the representational theory of mind arise because Fodor must make decisions about what kind of evidence is needed to support his theory in the absence of any tangible guidelines. Consequently, he champions evidence that appears to fit with his

hypotheses (at least superficially), dismisses criticisms of this evidence summarily, and neglects less favourable information. This approach was illustrated in subsections 4.4.2 and 4.4.3. Fodor's representational theory of mind is therefore open to the line of criticism developed in Chapter 2, subsection 2.6. There, I argued that the lack of any overarching theory of how disparate sources of information relevant to the study of the mind/brain should be synthesised renders the philosopher liable to judge the available evidence in the light of his own preferred explanations. Not only may his interpretation of the relative importance of evidence from different sources be coloured by this form of bias; he may also simply not see some facts as relevant evidence at all.

#### **4.6 Conclusions**

Firstly, it has been demonstrated that Fodor's use of the bootstrapping argument in the construction of his representational theory of mind has not resulted in a well-confirmed, scientific theory. Indeed, I have shown in subsection 4.4.4 that the problems isolated in subsections 4.4.2 and 4.4.3 render Fodor open to the same difficulties regarding explanation and the nature of evidence as those that have been associated with the traditional veil-of-perception issue. As a result, Fodor's claim to have produced a testable and explanatory theory of mind is

untenable. He is shown to be a closet metaphysical realist who does not remain true to the tenets of scientific realism that he professes to uphold.

Moreover, I have shown in subsection 4.5.2 that Achinstein's criticisms of the Glymourian bootstrap strategy are pertinent to the version employed by philosophers of mind. I also suggested in subsection 4.5.3 that Fodor's difficulties in evaluating evidence may be exacerbated by the absence of any guiding theory relating to the combination of data from a variety of sources. To this extent, Fodor, like Nordby's detective, bootstraps while barefoot.

Secondly, it has been emphasised that the criticisms of the Fodorean enterprise urged in subsections 4.4.2 and 4.4.3 follow readily from the remarks made by Patricia Smith Churchland in her discussions of Fodor's theory. This demonstrates that she is aware of how the abstract issues of scientific realism, explanation and the need for testable sources of evidence should be incorporated into specific theories of the mind/brain. In Chapter 6, it will be argued that Churchland's understanding of these questions as they pertain to the theories of others is in stark contrast to her treatment of them within her own attempts to construct a coherent account of the mysteries of the mind/brain.

I now turn to a discussion of the issue of modularity within the mind/brain. This will not only evaluate Fodor's attempt to grapple once more with the issues of realism, explanation and evidence, but will also demonstrate that the image of the veil-of-perception does not disappear if we cease to regard man as a united entity which confronts the rest of the world. Rather, the issue simply relocates at the junction between perception and cognition.

1. Much of this chapter is based upon the contents of a paper entitled "Realism and Representation in the Philosophy of Mind" which I read at King's College, London, in November 1990. I am indebted to Professor David Papineau and his department for some helpful comments.

2. Yet Fodor explicitly dissociates his own theory of the mind from those theories which have traditionally been associated with the difficulties of the veil-of-perception issue. He even uses Platonic imagery in his denial of such views (cf. Chapter 1, subsection 1.2.3): witness his comment that "It is easy to picture the mind as somehow caged in a shadow show of representations unable, in the nature of the thing, to get in contact with the world outside. And it's easy to go from there to an indefinite yearning for epistemic immediacy; a yearning which is none the less impassioned for all that it is largely incoherent...It is therefore pertinent to insist that this picture isn't the one that I have been developing, nor is it implied by anything that I have had to say" (Fodor, 1976, p. 204). However, this disclaimer by Fodor does not prevent his theory from encountering the difficulties of the veil-of-perception issue, if he draws conclusions to which he is not entitled solely on the basis of the flawed evidence cited in his premises.

3. See also below, subsections 4.3 and 4.3.3.

4. Although Fodor tends to refer simply to computation, he really means classical computation (ie: non-connectionist computation). For example, Clark comments that Fodor's "claim is that .. operations, resources, etc. are fundamentally classical; they consist of structure-sensitive processes defined over internal, classical, conceptual-level representations" (Clark, 1989, p. 150).

5. Fodor thus differs from other philosophers of psychology such as Stich, who are prepared to espouse a computational view of the mind/brain, but who believe that such a view dispenses with the need to talk in terms of propositional attitudes (Fodor, 1986b, p. 3).

6. The potential for circularity in this argument will be explored in sub-section 4.4.2.

7. Of course, Fodor's theory has been criticized at length by other philosophers and scientists. However, since Churchland's attitude towards the issues of realism and the veil-of-perception forms one of the key themes of this thesis, I have chosen to concentrate upon her arguments against Fodor. Criticisms of Fodor by others are included only in support of specific points made by Churchland in order to demonstrate that her arguments do not result solely from her own very individual views upon the nature of the mind/brain.

8. Indeed, it will be seen in Chapter 6 that Patricia Smith Churchland employs the bootstrap strategy herself. We shall regard this as an implicit endorsement of the methodology for the study of the mind/brain!

9. See Chapter 6, subsection 6.2.1 for further details of the eliminativist position.

10. It is ironic that Clark agrees with Churchland that Fodor's sentential hypothesis about the nature of representation within the mind/brain is incorrect, for he also holds that the failure of this very hypothesis will provide further grist to the mill of Churchland's eliminative materialism. This reinforces my view that Churchland's criticisms of Fodor do not arise solely from her own preconceptions about the mind, but are rather genuine criticisms of poor arguments and faulty evidence.

11. This blurring of the distinction between conceptual and empirical facts makes Fodor's argument seem more scientific than it really is. Cohen makes a similar comment in the peer commentaries of Fodor (1980). He argues that "[w]hat Fodor calls 'the representational theory' and 'the computational theory' are not scientific theories at all, but bits of conceptual analysis..... A scientific theory would reveal causes, effects or explanatory principles, and would have experimentally testable consequences" (Cohen in Fodor, 1980, p. 75).

12. Further ramifications of this point will be mentioned in the discussion of the success of the second half of the Fodorean bootstrap argument (sub-section 4.4.3).
13. This is the source of the potential circularity noted above (subsection 4.2.2). Although the bootstrap strategy is meant to permit the use of the same initial body of evidence to frame both a specific and a general hypothesis, Fodor must include additional sources of strong independent evidence in  $E_2$ , which is supposed to support the language of thought hypothesis. However, it will be argued in sub-section 4.4.3 that this body of evidence is equally unsatisfactory.
14. It will be seen in the next part of this discussion that it is a necessary consequence of the Fodorean account of concept-learning that the language of thought is innate. However, an examination of Churchland's views concerning the plausibility of an innate sentential system will be deferred until subsection 4.4.3.
15. This hypothesis has been constructed using the same notation as Fodor uses in Fodor (1976).
16. Once again, the line of argument put forward by Churchland is supported by the work of Clark. Clark argues (1989, pp. 69-72) that organisms do not necessarily employ systems that provide the neatest solutions to the tasks that they have to perform. Rather, the systems utilized frequently take the form of a "kludge", which Clark defines as "a solution dictated by available materials and short term expediency" (Clark, 1989, p. 71). Thus, if we were designing a creature from scratch, it might be apposite to give it sentential representations so that it could have concepts, but it is very unlikely that this will happen in nature. For further relevant discussion, see Chapter 6, subsection 6.1.3.
17. It has already been noted Chapter 1, subsection 1.2.2) that explanations have been criticised for displaying a *virtus dormitiva* since the time of Molière. This quotation, which encapsulates the problem perfectly, is taken from the Preface to Leibniz's *New Essays on the Understanding*, (1704).
18. The philosopher who tries to construct his argument using a bootstrap strategy but fails to support it with adequate sources of evidence will discover that the strategy has a sting in its tail. Good evidence may be combined with an initial hypothesis to yield further hypotheses which, when coupled with sound evidence from an independent source, will increase support for the first hypothesis. However, poor evidence will cast doubt upon any conclusions drawn from it in such a way that the entire theory may ultimately fall. This is a necessary consequence of the need to construct bootstrap arguments so that they avoid the charge of vicious circularity. Vicious circularity occurs if any evidence whatsoever would confirm the hypotheses contained within the theory. See also my discussion of Hunt (1990) in Chapter 2 for further explanation of how circularity can be avoided.



19. Indeed, other commentators have cast doubt upon  $H_1$ , Fodor's general computational hypothesis. For example, Wasow argues that Fodor's discussion "would have benefitted from more careful attention to the distinction between the use of computational models and the need for computational models. The fact that a certain phenomenon has been described in terms of computational models gives us little reason to accept the consequences of attributing psychological reality to such descriptions" (Wasow, 1978, p. 163). For further development of this point, see also Clark, 1989, pp. 152-154.

20. It has, however, acquired a whole new lease of life from being used so frequently as an example of a superceded concept. See for instance P.M. Churchland, 1989, p. 14.

21. It may seem initially that Fodor's (1987) attempt to provide a naturalistic account of the meaning of concepts within the language of thought resolves the problem of how we acquire new concepts without relinquishing the innate status of Mentalese, although this is not the main purpose of this work. However, a careful consideration of the arguments readily yields the conclusion that this move is equally unsuccessful. A more detailed evaluation of the arguments is provided later in this subsection.

22. For other criticisms along these lines, see the peer commentaries by Cohen and Hayes following Fodor (1980).

23. I do not propose to discuss the question of the formality condition at length, since it lies outside the scope of this chapter. It is raised merely to demonstrate that, if Fodor is correct to assert that naturalistic psychology is in practice impossible, then this closes off yet another potential source of evidence in support of his language of thought hypothesis, or at least, in support of the view that some inner representational system exists within the mind/brain.

24. Not all philosophers and psychologists share Fodor's pessimism about the possibility of a naturalistic psychology. It is surprising that Fodor does not consider the use of the bootstrap strategy to construct such a theory. This is, in effect, the route suggested by Davis in the peer commentaries of Fodor (1980). Davis asks (Fodor, 1980, p. 77) "why can't naturalistic psychology get started with what we have? If it turns out that [for example] salt really isn't NaCl, then when we find that out, we will revise what has to be revised".

25. See the next paragraph for an explanation of this term.

26. This remark in no way precludes the criticism that it is deceptively easy to assert that something is in principle possible in the very trivial sense that it is not logically impossible, but that it is much harder to put any flesh on this argument in the absence of any practical details.

27. See Chapter 1, subsection 1.2.4 for my use of this term.

28. Fodor tries to defend his claim that he adopts an attitude consonant with scientific realism on the grounds that whilst he holds a correspondence theory of truth, this is not incompatible with a coherence theory of evidence. He comments that he sees "nothing compelling in the inference from 'truth is a matter of the correspondence of a belief with the way the world is' to 'ascertaining truth is a matter of "directly comparing" a belief with the way the world is.' Perhaps we ascertain the truth of our beliefs by comparing them with one another, appealing to inferences to the best explanation whenever we need to do so" (Fodor, 1980, p. 68). However, this statement does not render the charge of metaphysical realism against Fodor invalid, since he plainly violates all commonsense standards of coherent evidence. It is scarcely coherent to claim that, for example, the language of thought accurately reflects the state of the world and yields the flexibility of our overt languages whilst at the same time maintaining that it is innate (see also subsection 4.4.3)!

29. In subsection 2.4.1 I suggested that Achinstein's view that the existence of evidence which supports a hypothesis relative to a theory says nothing about the truth of that theory may result in instrumentalism. However, Fodor is not an instrumentalist with respect to sentential representations, since he believes that they literally exist in the mind/brains of concept-users. Yet the evidence he provides in support of this realist hypothesis is, as I have stressed repeatedly in the course of this chapter, unsatisfactory.

30. For example, Fodor comments that "I think that it is very likely that all of the organismic causes of behavior are physiological" (Fodor, 1976, p. 9).

31. This reticence may be partly explained by the view he expresses in Fodor (1983) concerning the difficulties of investigating the central processes within the mind/brain. (See also Chapter 5, below). However, the bold claims that he makes for the existence of a language of thought contrast sharply with this caution. Nonetheless, I think that I am justified in ascribing to Fodor at least a modicum of acceptance of the relevance of neurophysiological data for the study of the mind/brain, since he considers these issues in his seminal work on the modularity of mind (Fodor, 1983).

## Chapter 5 - Modularity and Metaphysics<sup>1</sup>

### 5.1 - Introduction

So far, I have examined two theories of the mind/brain in the light of the issues of realism and evidence raised in the course of Section A. Both McGinn's theory of the hidden structure of consciousness and Fodor's Representational Theory of Mind are based upon conceptual analysis, even though they contain some evidence from psychology and from neurophysiology.<sup>2</sup> It was clear that, although both McGinn and Fodor rely implicitly upon a bootstrapping argument to support a general and a specific hypothesis about the mind/brain on the basis of just one body of initial evidence, neither of them avoided the traps created by the use of untestable hypotheses and unsatisfactory evidence. Both theories therefore suffered from the metaphysical difficulties associated with the traditional veil-of-perception issue.

I now turn to a discussion of the more modern co-evolutionary theories of the mind/brain to see whether their more overtly empirical hypotheses avoid these difficulties. I begin in this chapter with the modular theories of mind.

There are two main reasons for reviewing the modularist debate. Firstly, such theories call into question the very

image which has traditionally been at the root of the veil-of-perception issue, namely the image of man as a united being which confronts the rest of the world. However, it will become clear in the course of this chapter that this by no means obviates the difficulties raised by the veil-of-perception issue. They are simply relocated in the question of the relationship between the structures within the mind/brain, modular and otherwise.

Secondly, it is possible to isolate a bootstrapping strategy within the co-evolutionary lines of argument employed by at least two of the modularists, Jerry Fodor and Howard Gardner. Their arguments will be discussed in detail in the course of this chapter. Moreover, most of the arguments put forward by the modularists rely heavily upon the possibility of support in the form of scientifically testable evidence. However, this does not mean that these theorists are any more successful in negotiating the pitfalls of scientific realism than either McGinn or the early Fodor. The citation of scientifically obtained evidence does not necessarily entail the correct and objective interpretation of this evidence.

These considerations uphold my hypothesis (outlined in Chapter 2, subsection 2.5.2) that it is not always possible for philosophers of mind employing the bootstrap strategy to determine by means of Glymourian abstract criteria which

of two theories is actually the correct one when the theories are underdetermined by the evidence. This difficulty is explained by the novel and complex multidisciplinary character of the modular approach to the mind/brain issue. As yet philosophers have no over-arching theory to assist them in the selection and evaluation of evidence. Philosophers in innovative areas are faced with the evaluation of evidence from a variety of disciplines and the weight accorded to each source of evidence will depend upon the precise model of explanation that they have in mind. An incorrect decision may lead to the prosecution of the wrong theory, unsatisfactory testing of the evidence and the systematic acceptance of a false metaphysics. All these problems may be very difficult to correct or revise at a later stage.

The remainder of this chapter is given over to an elaboration and justification of these arguments.

## **5.2 What are modular theories of the mind?**

Modular theories of the mind developed from a belief that traditional psychology is making a fundamental mistake in regarding the mind as nothing other than a set of "horizontal" central executive processes which operate across all content domains, no matter how diverse these might be. Modularists all share the view that there are at

least some autonomous content-specific production systems or modules, but the details of their individual theories vary considerably.

Fodor (1983, 1985) describes a module as an informationally encapsulated inference-making computational system within the brain. This means that it does not have access to all the information contained within the rest of the brain, but only to the information within its own "database" together with information about the external stimulation that it is receiving at any one time.

Modules have their own particular function of providing our central cognitive processes with information about the world by making inferences from the way in which our sensory apparatus is affected by outside stimuli. (Fodor sometimes refers to them as input analyzers.) Consequently, they form part of our perceptual system, whilst our cognitive abilities are general and nonmodular. Our linguistic abilities are also based upon a modular structure, since they, like perception, have the function of providing us with information about the world beyond our senses.

Since modules are computationally autonomous (or "bullheaded" as Fodor likes to call them), they can only react in one way. Moreover, since they are informationally

encapsulated in the way just described, they are unable to take account of the background beliefs and knowledge possessed by the brain as a whole, because they have no access to them. Fodor argues that this explains why we continue to be subject to optical illusions even when we know that appearance is not reflecting reality. Our visual modules can only respond to the stimuli that come to them from the eyes and receive no information from our memory, or our knowledge of the rules of geometry, for example.

Visual modules transform the stimuli that we receive from our senses so that our cognitive capacities receive the information in a format that they can understand. There is therefore a division between the perceptual and the cognitive capacities within the brain which causes us to rethink the traditional image of the human brain as one entity which confronts the rest of the world. Just as microscopes enhance the presentation of information about the world so that we can understand it better, the visual input analyzers within the brain are performing a similar function which differs only in temporal location within the visual process. We may come to regard our perceptual capacities as an extension of the visual aids we use that are situated outside the brain, so that our concept of self is restricted in application to the cognitive capacities. However, the traditional veil-of-perception issue may simply be relocated in the question of the reliability of

the information presented to us by our visual input analyzers. We have already seen that Fodor attributes the existence of optical illusions to the mandatory operation of the modules. The possibility of other incorrect information reaching the cognitive processes via the modules cannot be overlooked. It may not always be so easy for the cognitive processes to spot the errors.

Informational encapsulation and computational autonomy have a further consequence. Since modules are unable to consider the relevance of background facts to their operation, they react very quickly (as do reflexes, which are also informationally encapsulated). Fodor regards this speed as vital for our survival and that the advantages of informational encapsulation more than outweigh the disadvantages caused by the occasional misidentification of objects that might perhaps be prevented if modules had access to more background information. As Fodor writes

"we want the perceptual identification of panthers to be very fast and to err, if at all, on the side of false positives" (1983, p. 70).

Modular systems frequently exhibit other properties. These include domain or content specificity. In other words, different modules deal with different perceptual stimuli, such as speech, other noise, colour, and shape. This is



what is meant by a vertical theory of perception. A horizontal theory would operate from the assumption that all perceptions could be processed by the same system, irrespective of their content.

Finally, Fodor holds that modules are frequently associated with specific neuroanatomical architecture or even specific locations within the brain.

Although modules provide us with an interpretation of the information yielded by incoming sensory stimuli about the world and therefore seem to be intelligent, Fodor believes that they are really dumb mechanisms since their operation is mandatory and restricted by their informational encapsulation. By contrast, Fodor thinks that our real intelligence inheres in our general cognitive processes, which he thinks are unencapsulated and slow, but which are not modular and function across all domains, irrespective of their content.

If Fodor's theory seems less precise in its delineation of the characteristics of such central processes, this is no accident. Fodor believes that we can know very little about our cognitive processes precisely because they are not modular. He writes that

"[t]he condition for successful science ... is that nature should have joints to carve it at: relatively simple subsystems which can be artificially isolated and which behave, in isolation, in something like the way that they behave in situ. Modules satisfy this condition; ... If, as I have supposed, the central cognitive processes are nonmodular, that is very bad news for cognitive science" (Fodor, 1983, p. 128).

So, Fodorean modules are very small information processors within the mind/brain and are delineated by reference to their function of interpreting information from the senses about the external world. By contrast, Howard Gardner holds that the mind is divided up into only seven modules or "intelligences", which are classified by their content. These seven intelligences are the linguistic, the logical-mathematical, the spatial, the musical, the bodily-kinaesthetic, the interpersonal and the intrapersonal.

Gardner regards his theory as an alternative to the traditional Piagetian view of intelligence as a general capacity which operates equally successfully in any context and which develops at a constant rate in each child and in all domains. In Gardner's opinion, the modular theory of intelligence provides a better explanation of why an individual may perform well in some tasks and very badly in

others.

Moreover, Gardner tries to adopt a more open attitude to the influence of different cultures upon the manifestation of intelligent behaviour. Traditional intelligence tests have, in his view, placed undue emphasis upon the logical and mathematical abilities prized in Western culture. If this bias is removed, he argues, then we may learn to acknowledge that tasks important in other societies such as navigation by the stars also require intelligence.

There are some similarities between Fodor's modules and Gardner's intelligences. Both are essentially computational mechanisms within the brain which are accessed by certain kinds of informational input. Gardner believes that at birth we have only the cores of what will ultimately become our fully developed intelligences. These cores are similar to Fodor's own modules, but also include primitive motor skills. Thus Gardner writes that

"[o]ne might go so far as to define a human intelligence as a neural mechanism or computational system which is genetically programmed to be activated or "triggered" by certain kinds of internally or externally presented information. Examples would include sensitivity to pitch relations as one core of

musical intelligence, or the ability to imitate movement by others as one core of bodily intelligence" (Gardner, 1985, pp. 63-4).

However, whereas Fodorean modules will always remain autonomous, Gardner believes that the cores of his intelligences are not informationally encapsulated, but rather will begin to interact as they develop against a particular cultural background. He comments that

"[i]n the normal course of events, the intelligences actually interact with, and build upon, one another from the beginning of life" (Gardner, 1985, p. 280).

Nonetheless, the intelligences may still develop at different rates and, in cases of brain damage, it is possible that some intelligences will be impaired or even destroyed whilst others continue to function as before. Indeed, it is only in such cases, or in autistic people, that we see the isolated core mechanisms of the intelligences exposed. Once more doubt is cast upon the accuracy of the view of a united self which inspired the traditional formulation of the veil-of-perception issue. On Gardner's theory, there is a live possibility that a person might react in vastly different ways to diverse stimuli from the external world depending upon which

intelligence is activated. Moreover, autistic people, prodigies and those who have suffered brain damage are living examples of the fact that cross-fertilization of skills and learning abilities between the different spheres of intelligence may be impossible. Despite his theory that intelligences do interact in the normal course of events, Gardner does not think that such interaction will result in the formulation of any general, non-modular cognitive processes.

### 5.3 Modular theories and the use of the bootstrap strategy

It has already been noted in Chapter 2 (section 2.5.1) that Glymour takes his bootstrap strategy to be most useful in the development of new theories. The attempt to understand the mind/brain in terms of a partially or wholly modular structure, and to support this with findings from neuropsychological studies, whilst not entirely new<sup>3</sup>, is certainly an area in which very little interest had been shown. Shallice, for example, comments that

"[f]or 100 years, it has been well known that the study of the cognitive problems of patients suffering from neurological diseases can produce strikingly counterintuitive observations.... However, in general, neuropsychology has had little impact on the

study of normal function" (Shallice, 1988, p.3).

This seems to be the ideal situation in which to employ the bootstrap strategy, and it is indeed possible to find evidence of the strategy within the writing of both Fodor and Gardner. Both use the strategy to bootstrap from a general hypothesis about the mind/brain to a more specific modular hypothesis which, with the right kind of evidential support, is intended to lend further confirmation to the original hypothesis.

Fodor begins the main thrust of his argument with the postulation of a

"trichotomous functional taxonomy of psychological processes; a taxonomy which distinguishes transducers, input systems, and central processors, with the flow of input information becoming accessible to these mechanisms in about that order" (Fodor, 1983, pp. 41-2).

It is a more elaborate form of this claim that may be said to constitute hypothesis  $H_1$  within Fodorean modularity theory (T).  $H_1$  is the general hypothesis that linguistic and perceptual systems have a functional similarity which is not shared by the central cognitive systems. Fodor's

reason or evidence  $E_1$  for this hypothesis is that

"both serve to get information about the world into a format appropriate for access by such central processes as mediate the fixation of belief" (ibid, p. 46)

This leads to Fodor's specific modular hypothesis  $H_2$ , which states that the functional class constituted by perceptual and linguistic systems also constitutes a natural kind, known as modules. The essential, defining feature of these modules is that they are informationally encapsulated in the sense explained above (subsection 5.2).

However, Fodor takes it that these modules share other features of psychological interest. The evidence that he cites in support of this claim constitutes evidence  $E_2$ . This evidence is in two parts. Firstly, Fodor attempts to show that we have some knowledge of the features of modules and secondly, he argues that we have no knowledge of our cognitive processes. Both of these points are, he holds, evidence which not only supports  $H_2$ , but which also strengthens our belief in  $H_1$ .

It seems that Fodor is well aware that he is making full use of a bootstrapping methodology in the face of uncertain evidence. He comments that

"I'll be trying to say what you might expect the data to look like if the modularity story is true of input systems; and I'll claim that, insofar as any facts are known, they seem to be generally compatible with such expectations" (ibid, p. 46).

Moreover, Fodor is also aware of the need for  $E_1$  and  $E_2$  to be independent from each other so that the bootstrap argument is not circular. He writes that

"if we undertake to build a psychology that acknowledges this functional class as a natural kind, we discover that the processes we have grouped together do indeed have many interesting properties in common - properties the possession of which is not entailed by their functional homogeneity (ibid, p. 46).

The individual sections of the Fodorean bootstrapping line of argument will be evaluated in the next few subsections. However, Fodor is not the only one to employ a bootstrapping strategy. It is possible to detect a similar line of argument in Gardner (1985).

It was noted in the previous subsection (5.2) that Gardner regards his modular theory of intelligences as a significant advance upon the traditional theories which



treat intelligence as one general capacity. Let us call Gardner's initial claim or hypothesis  $C_1$ <sup>4</sup>.  $C_1$  makes the general claim that traditional theories of intelligence as one unified capacity across all manner of contents and activities are unable to explain many aspects of human behaviour. Gardner writes that

"current methods of assessing the intellect are not sufficiently well honed to allow assessment of an individual's potentials or achievements in navigating by the stars, mastering a foreign tongue, or composing with a computer.....Only if we expand and reformulate our view of what counts as human intellect will we be able to devise more appropriate ways of assessing it and more effective ways of educating it" (Gardner, 1985, p.4).

Gardner derives much of his evidence for this hypothesis from the existence of many feats from different cultures which would often be overlooked by the Western world with its emphasis upon pencil and paper tests. Moreover, there are people within our own culture whose abilities force us to question the idea that intelligence is one capacity that functions equally well in all domains. There are prodigies who are gifted in, for example, music from a very early age, but who are of otherwise average ability. Equally,

autistic children sometimes have one remarkable talent which may far exceed the abilities of most people in that domain, whereas their performance of other tasks is well below average. It is cases like these which cast doubt upon the notion of a unified intelligence. They form Gardner's body of evidence  $D_1$ .

Gardner's conclusion is that these facts would be better explained by the specific hypothesis of multiple intelligences or modules within the mind/brain. So, hypothesis  $C_2$  is constituted by the

"conviction that there exist at least some intelligences, that these are relatively independent of one another, and that they can be fashioned and combined in a multiplicity of adaptive ways by individuals and cultures" (ibid, p.8).

$D_2$ , the body of evidence cited in support of  $C_2$  is, alas, not wholly independent of  $D_1$ , since further reference is made by Gardner to the existence of idiots savants and to prodigies. However, Gardner also alludes to neuropsychological evidence relating to subjects who have suffered brain damage and consequently lost some of their intellectual skills, whilst others have remained intact. This both supports  $C_2$  since it provides evidence that

intelligences are localized within the brain, and it also strengthens belief in  $C_1$ , namely, that such selective impairment cannot easily be explained by traditional theories of intelligence.

Gardner, like Fodor, does seem aware that he is using a bootstrap strategy. He writes that

"[m]y review of earlier studies of intelligence and cognition has suggested the existence of a number of different intellectual strengths, or competences..... The review of recent work in neurobiology has again suggested the presence of areas in the brain that correspond, at least roughly, to certain forms of cognition; and these same studies imply a neural organization that proves hospitable to the notion of different modes of information processing" (ibid, p. 59).

It is worth noting at this stage that Gardner regards his theory of modular intelligence as an instrumentalist one. It is useful to refer to the intelligences as if they really existed when we are trying to formulate an innovative theory of the mind/brain, yet

"[t]hese intelligences are fictions .... for discussing processes and abilities that (like all of

life) are continuous with one another; Nature brooks no sharp discontinuities of the sort proposed here. Our intelligences are being separately defined and described strictly in order to illuminate scientific issues and to tackle pressing practical problems .... they exist not as physically verifiable entities but only as potentially useful scientific constructs" (ibid, pp. 69 - 70).

However, Gardner's instrumentalism does not free him from the obligations of examining all the evidence available to him to ensure that he is not guilty of dismissing rival theories (instrumentalist or realist) without giving them due consideration, or of rendering his own theory immune from any refutation.<sup>5</sup> Gardner is prepared, at least nominally, to acknowledge this point. In the course of a discussion of the kinds of evidence that might refute his theory he comments

"[a]fter all, if M. I. [multiple intelligence] theory can explain (or explain away) all potentially disconfirming evidence, it is not a valid theory in the scientific sense of that term" (ibid, p. 298).

However, it remains an open question whether Gardner does indeed give due weight to all possible interpretations of the evidence for and against modular theories. I shall

argue in subsections 5.4 - 5.4.5 of this chapter that Gardner overlooks the fact that there are strong reasons for declaring that there is empirical underdetermination of the evidence between his theory and rival views.

#### **5.4 Modules: intelligent giants or ignorant dwarves?** **Gardner versus Fodor**

##### **5.4.1 Central cognitive processes: the status of the myth**

Howard Gardner makes this comment upon Fodor's stance with respect to central cognitive systems:

"Fodor ultimately reaches a conclusion that, though pessimistic from a scientific point of view, aligns his position somewhat closer to my own. Fodor concludes that scientific investigation should be able to illuminate the modules, because they are relatively distinct and so can be subject to controlled experiment, but that the central processor is probably immune to study because its lines of information are at once unlimited and totally interconnected. As a practical matter, then, the science of cognition reduces to the study of the individual matters. Even if the central-processing view is valid, says Fodor, we will not be able to incorporate it meaningfully into our science of cognition" (Gardner, 1985, p. 284).

This lengthy quotation is important, for it indicates that Gardner blurs a fundamental distinction between his own view and that of Fodor. The Gardnerian position is that a modular theory may yield a complete explanation of our mental activities, in which case we will be able to dispense with the entire concept of central cognitive processes. By contrast, Fodor thinks that the concept of a central processor has a meaningful place in our science of cognition, but that the very nature of this concept presents us with the empirical fact that we will be able to discover very little about how it is actually instantiated within the mind/brain or about its functional mechanisms. In discussing Fodor, Gardner fails to distinguish between the conceptual and the empirical. This raises the question of whether he treats all Fodor's statements with the same cavalier approach.

Two other points should be made at this juncture. The first is that Gardner's theory postulates a small number of large modules which perform the functions traditionally attributed to central cognitive processes. By contrast, Fodorean modules are plentiful, but minute, and perform a wholly different function from central cognitive capacities. Consequently, Fodorean modular theory would have a much larger explanatory gap to fill if it did not postulate the existence of central cognitive processes within the mind/brain.

Secondly, it is questionable how Fodor can justify his insistence upon the existence of central cognitive processes in the face of his assertions that we will never be able to discover a great deal about their nature.<sup>6</sup> This casts doubt upon the scientific status of the concept of a central cognitive process within his theory, but this does not exonerate Gardner's misunderstanding of the position that Fodor wishes to adopt.

#### 5.4.2 Empirical underdetermination and the criteria for modularity

Fodor's criteria for modularity are, quite correctly, criticized by Gardner as "idiosyncratic" (see Gardner, quoted in Fodor, 1985, p. 13), because they are a mixture of the logical (for example, modules are domain-specific) and the empirical (for example, modules react speedily). Gardner argues that his own criteria are far more testable and therefore acceptable, since they are all based upon the evaluation of wholly empirical evidence from eight different sources. Yet although Gardner does not include any logical definitions among his criteria for a modular intelligence, his interpretation of the empirical evidence shows a distinct tendency to ignore the fact that in some cases the evidence is equivocal between his own view and that of Fodor. In other words, in cases where the choice of theory is grossly underdetermined by the evidence,

Gardner finds in favour of his own *a priori* belief that he is probably right and Fodor is probably wrong.

For example, one source of Gardner's body of evidence  $D_2$  is the potential isolation of an intelligence by brain damage. Gardner cites (Gardner, 1985) numerous examples of how particular lesions in the brain may affect one particular capacity such as language or music and yet leave others such as logical and mathematical reasoning or spatial awareness completely unimpaired. Moreover, very localized brain damage may destroy just one part of an intelligence leaving the rest of it totally intact. This leads Gardner to conclude that each of his intelligences is autonomous in that it has a distinct neurophysiological basis, and he takes further proof of this to be the fact that autistic children may have one ability or intelligence that functions normally or even extraordinarily well, although they perform badly in other spheres. Equally a child prodigy may have one very great talent without being gifted in any other fashion.

We have already noted that Gardner's bootstrap argument is weakened by the citation of the existence of autistic children and prodigies as evidence both for the inadequacy of traditional theories of intelligence ( $C_1$ ) and for the existence of autonomous intelligences ( $C_2$ ), so that the two hypotheses are not supported by wholly independent sources



of evidence. (It would of course be permissible for Gardner to claim that the evidence produced from studies of brain damage supports his initial hypothesis that traditional theories of intelligence are inadequate and hence explains why autistic children do not fit well into such theories - this is precisely the way in which the bootstrap strategy is supposed to work.)

However, Gardner is faced with a second difficulty. His interpretation of the evidence from studies of autistic children is not the only one possible. For example, in the case of the autistic artist, Nadia, Gardner writes that:

"her drawings stand as an eloquent demonstration of the dissociability of spatial intelligence from other intellectual strengths and of its potential for a singularly high degree of development" (Gardner, 1985, p. 190).

However, Gardner fails to specify precisely what it is about Nadia's performance that is intelligent. Nadia's remarkable ability could equally well be explained by Fodorean modular theory. If Nadia's visual input analyzers are more finely tuned than those of many people then they may yield a better representation of the world. She may then be able to reproduce this representation using her motor skills, which Gardner acknowledges may be among the

core "dumb mechanisms" of a spatial intelligence. He writes that

"central to spatial intelligence are the capacities to perceive the visual world accurately, to perform transformations and modifications upon one's original perceptions, and to be able to re-create aspects of one's visual experience, even in the absence of relevant physical stimuli" (Gardner, 1985, p. 173).

Nadia is unable to classify or organise even the content of her drawings. Gardner comments that

"[S]he lacked the conceptual knowledge requisite to her drawing skills. She could not perform sorting tasks where she had to put together items of the same category. Moreover, in her own drawings, she would show little regard to the particular object being depicted. Sometimes she would cease to draw an object right in the middle of its contour or continue to draw right off the page, as if slavishly transcribing a form that she had committed to memory" (Gardner, 1985, p. 190).

Hence we have no reason to believe that Nadia has higher-level skills in the spatial sphere any more than she has them in any other sphere of intelligence. Gardner

overlooks the fact that there is no more reason to read the evidence as support for his theory than there is to regard it as support for the Fodorean modularity theory. On the Fodorean view, Nadia's visual input analyzers are highly developed, but her general cognitive organizational processes are not, so that she remains unable to utilize the information received from these input analyzers in a more productive fashion. This may explain the difference between the limited development of the autistic child and the performance of the genius who is ordinary in every other way - the general processes of the latter are more able to cope with the output of the highly efficient core mechanisms. This indicates that the data relating to autism results in empirical underdetermination between Gardner's hypothesis (C<sub>2</sub>) relative to his modular theory of mind and the account provided in Fodor's theory, T.

There is a further difficulty with Gardner's use of the prodigy and the autistic child as illustrations of the autonomy of his intelligences. No account has been taken of the role played by motivation in the development of the intelligence. Does the child composer show a marked musical ability because he is interested in the subject or is he interested in the subject because he has the ability in the first place? One would imagine that the truth probably lies somewhere between these two extremes. However, Gardner does recognize that there is a gap in his

theory at this point.

Nor does the neurophysiological evidence cited by Gardner support his theory unequivocally. For example, Gardner admits that it is very rare to find a person whose linguistic abilities have been impaired whilst the remainder of their intellectual capacities remain intact. This casts doubt upon the identification of language as an autonomous intelligence. However, Gardner glosses over this possible objection to his hypothesis by claiming that language satisfies enough of the other criteria for a separate intelligence for it to be classified as such.<sup>7</sup>

This is an example of the advantage that Gardner has by not attempting to formulate an algorithm for the classification of intelligences - he can happily shift his criteria enough to allow him to neglect systematically those sources of empirical evidence that contradict his desired result. Given this method, he has no need of the *a priori* criteria for modules for which he criticises Fodor. If an candidate intelligence is excluded by one criterion, but readily included by other criteria, then Gardner can include precisely those intelligences in his list that best suit his theory just as easily as if he had postulated their existence on *a priori* grounds. Neither methodology pays much attention to the important question of systematic testability.

It should be said in Gardner's defence that his discussion of the problems encountered in the identification of intelligences does at least indicate that he is not wholly unaware of the subjectivity of his approach. However, he maintains that his method is scientific by dint of the very fact that he makes public the sources of his evidence so that readers may judge for themselves whether he is right or wrong (ibid, p. 62). It is my belief that, although Gardner should be praised for citing his evidence, this does not, and should not, remove the onus upon him to discuss rival interpretations of the evidence within his text, thereby achieving a more balanced outlook and illustrating the relative confirmation of his theory. Gardner's failure to do so weakens the strength of his bootstrap argument.

#### **5.4.3 Memory and modularity: is Gardner's theory remotely tenable?**

It has already been noted that Gardner pays insufficient attention to the relationship between motivation and intelligence. He also neglects to investigate the question of meaning with respect to memory in sufficient depth. Gardner comments (ibid, p. 195) that a chess master will remember the layout of pieces on a board after just a few seconds of exposure to it, provided that the pieces are arranged in a meaningful fashion. If they are placed

indiscriminately on the board, then he will recall their positions no more accurately than a mere novice. Gardner concludes from this that the memory of the chess master is influenced by the meaningful configuration of the pieces upon the board, which his highly developed spatial intelligence allows him to see instantly. In other words, his spatial memory derives its excellence from his spatial intelligence, whilst his memory in other domains, such as foreign languages, may be appalling.

This view is obviously problematic. Firstly, if Gardner's claim that the chess master can remember the layout on the board because his spatial intelligence tells him that it has meaning, what are we to say of an architect to whom the strategies of chess mean nothing, but who has an excellent memory for the details of structural plans? Does he have the same spatial intelligence as the chess master or a different one? Just how specialized can an intelligence be?

Here Gardner appears to be faced with a dilemma. His first option is to admit that the architect and the chess master have different spatial intelligences, in which case he is paving the way for a flood of candidate intelligences, not only within the spatial domain, but in all the other spheres that he mentions too. I can play polo very well, but I am a terrible swimmer. Do I have the same bodily-

kinaesthetic intelligence as Duncan Goodhew or a completely different one? I am an accomplished journalist and novelist, but cannot master a word of French. Do I have the same linguistic intelligence as Sartre or a different one? The list is endless. Fodor is making a similar point when he declares that

"I don't think that there are specialized modular cognitive systems corresponding to most of the things that we know how to do" (Fodor, 1985, p. 36).

However, the other horn of the dilemma is no more attractive for Gardner. If he claims that the architect and the chess master have the same spatial intelligence, then his attempt to explain their abilities by reference to spatial intelligence is bound to fail, since we have just seen that there is no reason why the architect should have any memory for what to him is a meaningless arrangement of chess pieces upon a board.

Either way, Gardner is forced to relinquish his modular argument. If there are many different intelligences, then the spatial intelligence of the chess master bears little resemblance to that of the architect and so the general notion of spatial intelligence does not explain a great deal. In fact, the modular theory risks becoming trivially true, if Gardner is reduced to claiming that the chess

master's ability to understand and remember the arrangement of chessmen on a board is explained by his superior understanding of and memory for the spatial arrangement of chessmen. On the other hand, if there is only one form of spatial intelligence, then Gardner's theory still loses much of its explanatory force, since it is impossible to explain why the architect remembers building plans well but chess arrangements badly by reference to a superior spatial intelligence.

Moreover, any attempt to rescue Gardner by reducing his argument to the claim that our memories function better with increased understanding of what it is that we are trying to remember will also fail, since even if this is correct, it does not in itself imply that our memory is not a horizontal cognitive process. Once more, Gardner is faced with the problem that Fodorean modular theory is equally capable of explaining the facts available. Our memory may utilise information from Fodorean input analyzers in different ways in different contexts and may yet be a general process. Fodor highlights a comparable mistake made by Gall who inferred that, as Fodor puts it

"the same faculty cannot be both weak and strong, so if it sometimes happens that mathematical memory is weak and musical memory robust, then the memory that mediates mathematics cannot be the same as the memory



that mediates music" (Fodor, 1983, p. 17).

However, Fodor continues

"All that can be inferred, strictly speaking, is that mathematical memory  $\neq$  musical memory; which, though patently true, is quite compatible with mathematical memory and musical memory being exercises of the self-same faculty with respect to mathematics in the one case and music in the other" (ibid, p. 17).

Hence not only does Gardner's evidence relating to memory function (again, part of  $D_2$ ) fail to provide unequivocal confirmation of  $C_2$ ; it is equally unable to lend further support to Gardner's general hypothesis ( $C_1$ ) that a general factor of intelligence does not adequately account for our abilities. Once more, the bootstrap strategy has not been supported by strong enough unequivocal evidence.

#### **5.4.4 A modular account of general processes**

There is one final criticism of Gardner that must be made before this section is concluded. This concerns Gardner's attempt to give a satisfactory modular account of those processes which we usually regard as general, such as common sense, originality, wisdom and the ability to use metaphor (Gardner, 1985, chapter 11).

These processes are interpreted by Gardner as amalgams of talents from more than one sphere of intelligence - perhaps only two or three, as in common sense, which Gardner declares not to be very general at all upon closer examination, since it is made up of interpersonal, bodily and spatial intelligence (Gardner, 1985, p. 289). Others such as metaphor and originality develop in one domain and then spread into others, whilst still others such as wisdom are amalgams of these amalgams. Wisdom, says Gardner, is a term applied to those who have

"considerable common sense and originality in one or more domains, coupled with a seasoned metaphorizing or analogising capacity" (Gardner, 1985, p. 295).

On Gardner's account then, the wise person must be very gifted in many spheres! Of course, Gardner's modules are not informationally encapsulated like Fodor's and on his view, frequently work together, but the force of his claim that modular intelligences exist must be weakened if so-called general processes must be explained in terms of ability in so many of these domains, unless the bulk of the population is to be regarded as lacking in common sense and wisdom! This impression is reinforced by factors such as Gardner's discussion of originality in which he seems to be referring to the standards that we would normally expect only of an expert (ibid, pp. 289 - 91).

It is clear that the empirical evidence which he cites in support of his theory does not justify the strong claim he makes (Gardner in Fodor, 1985, p. 13) that his brand of modularity theory has no need of central cognitive processes, so that he is

"more Fodorean than Fodor".

Gardner's account of general processes is unconvincing. This is of course more grist to Fodor's mill, since the Fodorean claim is that we know little about such processes precisely because they do not have a modular structure. Indeed, Gardner's evidence that general processes are modular does rather more to refute  $C_1$ , his general modular hypothesis, than it does to support his theory.

#### **5.4.5 Summary of the Gardner versus Fodor debate**

It is not possible to decide which of the two theories is correct purely on the evidence cited by Gardner, and so we are left with the difficulty of how to decide between his bootstrap argument and that of Fodor (see above, subsection 3.3.2.). We have already noted that some of the evidence ( $D_1$ ) used by Gardner to argue that the traditional notion of a general intelligence factor is incorrect ( $C_1$ ) is also employed in support of his specific hypothesis ( $C_2$ ) that intelligence is modular. This means that we may be

able to prefer the Fodorean model on one of the abstract grounds suggested by Glymour, that is, that we should prefer a theory in which each hypothesis is supported by independent sources of evidence. However, we have not yet examined the pitfalls of the Fodorean argument, and therefore it is as yet unclear whether the Fodorean theory will satisfy abstract criteria any better than Gardner's theory<sup>8</sup>. We must therefore reserve judgement about whether it is possible to employ Glymourian abstract criteria to decide between the two theories.

It is important to see that much of the preceding discussion of the Gardnerian position has been concerned not with showing that Gardner's theory is incorrect (although, inevitably, doubt has been cast upon some aspects of his view), but rather with a demonstration that the evidence cited by Gardner in favour of his theory does not support him in the unequivocal fashion that he often implies that it does. It has been argued that much of this evidence would be equally consistent with a Fodorean account of the mind/brain. Indeed, Gardner's abject failure to account for many everyday capacities such as common sense and wisdom may lead us to maintain that the only tenable version of his theory would be one in which the cores of each intelligence (Gardner's dumb mechanisms) are modular, but higher level processing is performed in a general, nonmodular fashion. If this is the case, then the

Gardnerian theory would collapse into the Fodorean position, and his own bootstrap strategy will fall.

Even if some form of Gardner's theory may be maintained which is distinct from the Fodorean view, then it is not wholly clear as yet how we may decide between them. How much weight should we attach to empirical evidence? Does our lack of knowledge of central cognitive processes mean that they do not exist or that we just do not know very much about them? How important are Fodor's functional criteria for modularity? This is a new area of multi-disciplinary study, in which both Fodor and Gardner are employing the bootstrap strategy to confirm their own model of mind/brain organization. There are few guidelines or precedents about what might constitute an acceptable strategy for evaluating evidence from a variety of sources. As a result, the preference of one model over another may well be based upon a *decision* as to what evidence is relevant and what kinds of tests it should be subjected to. We have already seen that Gardner allows his own view of intelligence to influence his interpretation of the evidence, and it will be demonstrated later that Fodor is no less guilty of this.

The modularist is in a position akin to that of Nordby's detective (Nordby, 1989) (see Chapter 2, subsections 2.5.1, 2.5.2). His choice of model may lead to, rather than arise

from, his interpretation of the evidence. Moreover, as the arguments of this chapter illustrate, the provisional nature of his conclusions is frequently forgotten in the course of discussions of important and contentious points. Neglect of alternative interpretations of the evidence is all too often a recurrent feature of the modularist debate. As we have seen in the case of Gardner, this neglect of the issue of empirical underdetermination results in the making of unwarranted assumptions. Although Gardner is an instrumentalist about his modules, his theory exhibits the same failure to ground his hypotheses upon unequivocal sources of evidence as we noted in realist theories.

Gardner is aware of some of the difficulties involved in the scientific pursuit of a theory of the mind/brain. (See earlier, subsection 5.4.2). A further example occurs in his discussion of the statistical analysis of intelligence testing. He comments that

"[w]hen it comes to the interpretation of intelligence testing, we are faced with an issue of taste or preference rather than one on which scientific closure is likely to be reached" (Gardner, 1985, p, 17).

It is unfortunate that Gardner does not adopt this attitude when he considers the evidence that he cites in favour of his theory.

## 5.5 Shanon versus Fodor - further failures to evaluate the evidence

### 5.5.1 Introduction

Benny Shanon suggests (Shanon, 1988) that there is evidence in support of other modular divisions within the mind/brain than those postulated by Fodor. Whilst I agree with this claim, I believe that Shanon, like Gardner, makes the fundamental error of failing to evaluate the evidence in a non-biased fashion. Whilst there may be good arguments that can be evinced against Fodor's view (see subsection 5.6 below), Shanon's arguments do not number amongst them. I shall substantiate this claim in the course of the following four subsections. The discussion will have the following structure.

Shanon suggests that the Fodorean notion of modularity (Fodor [1983]) as "structural and fixed" should be replaced by

"one which is *dynamic*, (and) context-dependent"  
(Shanon [1988], p.331).

I propose to comment briefly upon the major points of Shanon's attack on Fodor in subsections 5.5.2 and 5.5.3, which will deal with the structure of the modules and

methodological arguments respectively. It is in these two sections that Shanon's failure to evaluate the evidence for the Fodorean view is most clearly displayed. In subsection 5.5.4, I argue that, whilst Shanon claims victory on both counts, such a victory is at best Pyrrhic: Shanon's revision of the modularity thesis deprives it of its capacity to account for the discrepancy between our knowledge of input systems and our comparative ignorance of cognitive organisation within the brain. If Fodor is wrong about this division, then better arguments are needed to show it. Consequently, Shanon's modular theory of mind is less explanatory than that of Fodor. In Section 5.5.5 I shall make the further claim that this attempt to revise Fodorean modularity lends unintentional support to Fodor's despondency about our chances of providing a scientific account of the central cognitive processes. So, Shanon's whole attempt to provide a scientific account of cognition by revising the definition of modularity fails, because it cannot withstand the methodological criticisms he levels at Fodor. This means that even if it can be shown by other arguments that the Fodorean theory of modularity does not meet the requirements of science, Shanon's replacement theory will not fare any better.



### 5.5.2 Remodelling modularity

Shanon begins his paper with an attempt to summarize Fodor's theory of the modules. He agrees with Gardner's criticism (cited above, subsection 5.4.2) that Fodor's descriptions of the modular properties of input analyzers are derived not merely from functional considerations, but also based upon procedural and neurophysiological features. Of prime importance to Fodor are the functional characteristics of domain specificity (described in Fodor 1983 as "Gall's idea that there are distinct psychological mechanisms - vertical faculties - corresponding to distinct stimulus domains", p.48) and informational encapsulation (defined by Fodor 1983, p.69 as "the claim that the data that can bear on the confirmation of perceptual hypotheses includes, in the general case, considerably less than the organism may know"). However, Shanon, like Gardner, points out that Fodor employs many empirical considerations to enlarge upon these defining attributes of his mental modules. He writes that

"[e]vidently, such a characterization of the modularity of mind is not one that could have been suggested on mere *a priori* grounds" (1988, p.334).

Fodor's (1985, p.3) implies that, whereas informational

encapsulation is *crucial* to the concept of modular systems, it is a *contingent fact* that such systems also tend to have other psychological features in common. Shanon attempts to maintain the Fodorean insistence upon the criteria of informational encapsulation and domain specificity in his revised account of modularity. Indeed, there is little or no disagreement between Fodor and Shanon as to what is logically possible. Their conflict centres around the contingent, but actual nature of modularity. Both do rely heavily upon empirical evidence in formulating their hypotheses.

However, Shanon commits a serious methodological error in declaring (1988, p.334) that a concrete discussion of the empirical evidence cited by Fodor lies beyond the scope of his paper. There are three main ways in which Shanon's neglect of Fodor's use of empirical evidence and willingness to allow for future empirical developments has unfortunate consequences for his arguments.

Firstly, the gap between the views of Fodor and Shanon is thereby made to appear wider than it really is. Shanon has a tendency to misinterpret Fodor's stance upon potential instantiations of modularity. For example, Shanon writes that the direction of information flow concentrated upon in Fodor 1983 runs from the external world to the internal cognitive faculties of man. He goes on

"there is, however, the converse course: the one that proceeds from within to without. This course presents itself in motor performance and in the production of language, as well as in processes of recollection, retrieval and problem-solving" (Shanon, 1988, p.334).

This implies that Shanon takes Fodor to be restricting the application of the term modular to input analyzers only. We will consider the arguments with respect to problem-solving and memory in subsection 5.5.4. Suffice it to say for the moment that Fodor explicitly allows that his tripartite division of the mind into transducers (which present information from the outside world as raw material for the modules), input analyzers (or modules) and central processes may not cover all the workings of the mind/brain. Indeed, he writes that

"it is left wide open that there may be modular systems that do not subserve any of these functions. Among the obvious candidates would be systems involved in the motor integration of such behaviours as speech and locomotion" (Fodor, 1983, p.42).<sup>9</sup>

Secondly, it is disconcerting that Shanon ignores Fodor's empirical considerations with regard to modular systems

when his argument against Fodor in section 3 is concerned with precisely the fact that Fodor does not adduce such considerations when discussing the nature of our central cognitive processes.

Thirdly, Shanon's own modular hypothesis rests heavily upon empirical evidence gleaned from various psychological experiments. As will be argued in subsection 5.5. below, Shanon makes the same mistake as Howard Gardner in assuming that the empirical evidence he discusses may be interpreted unequivocally in favour of his theory. It is therefore regrettable that he does not examine Fodor's evidence more thoroughly, since there is ample scope for criticism (as demonstrated in the peer commentaries of Fodor (1985)<sup>10</sup>. Shanon's failure to do so, coupled with the problems entailed by his own views make the Fodorean account seem more satisfactory than it really is.

Shanon agrees with Fodor that at least some of the central cognitive processes must be nonmodular, but he claims that, whilst all central cognitive processes may in principle operate in a nonmodular fashion, exhibiting the characteristics of isotropy (the property a system has when it can look at anything it knows) and Quineanity (a term applied to systems within which any given hypothesis is sensitive to properties of the entire system) ascribed to them by Fodor, in practice they may display instances of

modularity. Modules occurring within the central processes exist only temporarily, and are not associated with any fixed neural architecture. Nevertheless, Shanon claims that during their ephemeral existence they manifest the functional features of domain-specificity and informational encapsulation of the Fodorean modules (Shanon (1988), p.347).

The advantage of viewing the mind/brain in such a modular light is, in Shanon's view, that we will be able to study the patterns of encapsulation within the mind/brain and perceive important psychological generalizations that the stricter, Fodorean definition of the module overlooks. If Shanon is correct, then his arguments would cast serious doubt upon Fodor's bootstrap argument, since his specific hypothesis  $H_2$  regarding the modular structure of input analyzers and the non-modular structure of the central cognitive processes could no longer derive support from the body of empirical evidence  $E_2$ . Consequently, no further confirmation of the general Fodorean hypothesis ( $H_1$ ) that our central cognitive processes are functionally distinct from linguistic and perceptual processes would be available. However, Shanon's views will be disputed in subsections 5.5.4 and 5.5.5 of this chapter.

### 5.5.3 Methodology under the microscope - Shanon on Fodor

Shanon makes two major criticisms of the Fodorean methodology, which are closely linked to one another. The first is that Fodor's criteria for the modularity of input analyzers are very strict, whereas his reasons for holding the central cognitive processes to be nonmodular are less stringent. Shanon writes that

"in other words, for Fodor a few examples that an input system is not encapsulated are not sufficient for characterizing it as non-modular, whereas any instance of even potential nonencapsulation is sufficient for the characterization of the central processes as non-modular" (Shanon (1988), p.335).

He accuses Fodor of interpreting crucial empirical evidence relevant to such issues as what constitutes an individual system, the nature of representational output, etc. in the light of his own theories about the make-up and functional divisions of the mind/brain.<sup>11</sup> If Shanon's allegations are correct, this would clearly be problematic for the Fodorean bootstrapping argument, since his evidence  $E_2$  for the specific modular hypothesis of the mind he puts forward in  $H_2$  would not necessarily uphold  $H_2$  when examined objectively, and this would have repercussions for Fodor's

more general claim ( $H_1$ ) that the perceptual and linguistic systems within the mind/brain are functionally distinct from the central cognitive processes.

It may be said in Fodor's defence that he does acknowledge the problem of analyzing empirical evidence and observations without any prejudice. Shanon, too, notes this point (Shanon, 1988, p. 336). For example, Fodor writes that

"I claim only that, contrary to the textbook story, the empirical evidence for the continuity of perception with cognition is not overwhelming when contemplated with a jaundiced eye" (Fodor, 1985, p. 5).

Fodor's failure to review evidence impartially will be further discussed in subsection 5.6 below. However, it will be seen in subsection 5.5.4 that Shanon has equally scant regard for the need for strong evidence to support his hypotheses. Whilst this does not obviate his criticism of Fodor, it does cast doubt upon the question of whether he fully understands the relationship between evidence and hypothesis.

Shanon's comments concerning the impossibility of Fodor's

arriving at his modular conception of mind on purely a *priori* grounds implies that he regards Fodor's supplementation of his functional definitions with empirical data as responsible for his rigidity with respect to the modularity of the input systems. This point is similar to that made in a more critical fashion by Howard Gardner (noted above, subsection 5.4.2). The criticism is not wholly unwarranted, but as I shall argue in subsection 5.5.4, it is one which might well be applied to Shanon too.

The second criticism of Fodor's methodology made by Shanon is that Fodor operates from a mixture of two perspectives throughout his (1983). The first is psychological and, as mentioned above, concentrates upon the actual mechanisms of the input analyzers as well as citing their functional characteristics, whilst the second is philosophical and discusses the central cognitive processes in terms of isotropy and Quineanity. Shanon writes that

"the considerations of principle Fodor employs in *The Modularity of Mind* are philosophical to an extent that takes them outside the scope of psychology proper. The philosopher's .. perspective of principle is atemporal; the psychologist, by contrast, cannot avoid the consideration of the actual, and for him there is nothing more essential than time. Just as the



atemporal study of language is linguistics but not psycholinguistics, the atemporal study of mind is epistemology but not psychology. Surely, the study of scientific discovery need not be the paradigm for the study of standard, everyday human reasoning" (Shanon (1988), p.336).

Shanon is, however, putting the cart before the horse here. Fodor is taking a philosophical stance because, in his view, there is a lack of current empirically based scientific evidence about the way that the central cognitive processes function, hence he is forced to retreat to a more theoretical and philosophical stance. Here is how he describes his task (Fodor (1983), p.104):

"The fact is that there is practically no direct evidence, pro or con, on the question whether central systems are modular....When you run out of direct evidence, you might just as well try arguing from analogies, and that is what I propose to do".

Furthermore, although the approach shifts from a psychological to a philosophical one, the points made with respect to the central processes are exactly the opposite of those made about the peripheral modules in the light of psychological considerations. Isotropy and Quineanity are

the exact reverse sides of the coins of informational encapsulation and domain specificity respectively.

However, it should be noted that the Fodorean attempt to combine a philosophical study of the mind/brain with considerations taken from psychology and neurology is partially responsible for the difficulties he has in using the bootstrap strategy. This point will be discussed in more detail below.

Shanon's point against Fodor has a sequel, namely, that Fodor's strict definition of modularity forces him to overlook the possibility that there are some psychological groupings even amongst the central processes which exhibit the functional characteristics of domain specificity and informational encapsulation so vaunted by Fodor in his discussion of the peripheral modules. Study of these groupings, in Shanon's view, can be useful to scientific study of the mind/brain. Shanon's argument amounts to the accusation that Fodor's definition of modularity governs his interpretation of the evidence, so that his evidence  $E_2$  in favour of his own version of the modularity thesis ( $H_2$ ) is only obtained by judicious selection of examples of modules that support his hypothesis and systematic omission of those that do not.

However, subsection 5.5.4 (below) argues that Shanon's

methodology in defining these groups is faulty, since he helps himself indiscriminately to what is, in effect, a totally mixed bag of psychological phenomena. In doing so, he displays the same disregard for any distinction between functional and empirical criteria as Fodor. His argument therefore provides further support for Fodor's opinion that we will not be able to have any satisfactory study of the central processes which will incorporate both neurophysiological and psychological considerations.

#### 5.5.4 Blurring the focus: Shanon's modules

Shanon is aware that there are differences between his various modular groupings, and that they do not exhibit all the secondary qualities of the Fodorean modules (as described in subsection 5.5.2 above). However, I think that the dichotomies between them are far wider than Shanon admits, since they do not even all exhibit domain specificity or informational encapsulation in the same way. Shanon is therefore as guilty as Fodor of interpreting his criteria for modularity so that they pick out only those putative examples that support his own view.

We shall begin by considering the way in which Shanon claims that we develop isolated habits, beliefs or skills. This category includes examples of isolated fields of

competence like the ones exhibited by chess masters and *idiots savants*. Shanon also argues (Shanon, 1988, p. 342) that modularity is demonstrated when we perform a task satisfactorily in one context, yet fail to carry out an essentially identical task in a different situation. Moreover, he notes that (*ibid*, p. 341), our memories may function better in some contexts than they do in others. However, it is clear that the level of modularity under consideration here is by no means as fine-grained as that described by Fodor (cf. note 1 above). The difficulties inherent in these examples of modularity have already been dealt with at length in the course of my discussion of Howard Gardner (above, subsections 5.4.2 and 5.4.3).

However, not all of Shanon's modules bear such a strong resemblance to those cited by Gardner nor can they be so easily reinterpreted in line with Fodorean theory. For example, Shanon claims that information may be sheltered locally through prejudices, which arise through the sheer impossibility of checking our beliefs against all other beliefs and knowledge represented in our mind/brains. Nevertheless, it does not follow that prejudices are therefore either domain-specific or informationally encapsulated in any interesting or even remotely Fodorean sense. For a start, our inability to check one set of beliefs against all our other background beliefs is, in Fodor's opinion, an example of epistemic boundedness, which

occurs when

"our cognitive organization imposes epistemically significant constraints on the beliefs that we can entertain" (Fodor, 1983, p.120).

However, Fodor regards the idea that our mind/brains are simply not equal to such a task because of lack of memory or computational power as uninteresting, because it is equally valid irrespective of our stance upon the question of modularity.

Secondly, the functional definition of domain-specificity implies that it operates in a vertical direction, whereas it seems that prejudices may appear irrespective of the content or function of our beliefs.

This point about informational encapsulation and prejudices also applies to Shanon's claim that the ad hoc sheltering of information during thought experiments or during the consideration of specific problems constitutes a form of modularity. Whilst Fodorean modules are only concerned with one particular function and operate in a bullheaded and mandatory fashion, this cannot be said to apply to Shanon's modules. Consider his analogy of the ad hoc sheltering of information with the use of specific tools on a workbench for repairing an object. Shanon writes that

"it is well to concentrate all required materials and tools on a workbench, and to separate them from other items so as to be able to direct one's resources. Experiencing a state of affairs as quite real and at the same time maintaining the knowledge that it is, in fact, not real is a fundamental ability of human intelligence" (Shanon (1988), p.345).

The analogy serves only to highlight just how different Shanon's view is from the Fodorean notion of informational encapsulation and domain-specificity that he claims to follow so faithfully in broadening his criteria for modularity. In considering our problems we may well revise our beliefs in the light of new knowledge, hypotheses will become prejudices etc. The point is that they may be interchangeable, just as we may replace workbench tools in different places, or hunt for another if we find our current selection to be inadequate. The fact that we will not always succeed because, in the case of beliefs, the holistic system is too vast, or in the case of the workbench, our tools are in someone else's garage, does not make for informational or tool encapsulation in any interesting sense. Rather such faculties as thought experiments and treatment of prejudice appear to be horizontal and unencapsulated processes. (Remember that Fodorean domain specificity is vertically defined.)

Compare this with Fodor's description of the horizontal faculty of memory according to Plato (Fodor, 1983, p.12). Socrates says that we should think of a man

"hunting once more for any piece of knowledge that he wants, catching, holding it, and letting it go again" (ibid, p. 12).

#### 5.5.5 Shanon, Fodor and the prospects for science

I do not wish to deny that the psychological groupings described by Shanon exist within the mind/brain. Indeed, Shanon, like Fodor, combines empirical evidence (particularly in his slightly more plausible example of dreaming, which displays a neurophysiological basis) with his supposedly functional ascriptions of domain specificity and informational encapsulation. I simply wish to reject Shanon's conclusion that such groupings form a continuum in any important sense of modularity since many of them violate even those aspects of the Fodorean modular thesis that Shanon claims to endorse. If we were to follow Shanon's example, then literally any informational grouping would have to be modular. Clearly, this use of the term *modularity* is too broad. Shanon admits this problem, but does not deal with it adequately.

The patterns of modularity which Shanon regards as so

important for scientific study are in fact so heterogenous that there is no advantage over simply studying the groupings in isolation. Shanon's multiplicity of modules does nothing to dispel Fodor's gloom in holding that

"to the extent that the existence of form/function correspondence is a precondition for successful neuropsychological research, there is not much to be expected in the way of a neuropsychology of thought" (Fodor (1983), p.127).<sup>12</sup>

#### **5.6 Some better criticisms of the Fodorean modularity thesis**

The failures of Shanon and Gardner to provide any telling arguments against the Fodorean modularity thesis do not entail that no such arguments are available. Indeed, much of the initial plausibility of the Fodorean view may be attributed to two major sources. Firstly, we have seen that both Shanon and Gardner omit any consideration of alternative interpretations of the evidence that they cite in support of their views. It is easy to show that much of the evidence can be interpreted in a manner which supports the Fodorean theory equally well. Moreover, a close examination of Shanon's arguments yields the conclusion



that his claim to have demonstrated instances of modularity overlooked by Fodor which nevertheless satisfy the Fodorean criteria of informational encapsulation and domain specificity fails, because he has simply reinterpreted these criteria so that they will apply to his chosen examples. He is therefore arguing at cross-purposes with Fodor.

The second reason for the apparent survival of the Fodorean thesis lies within Fodor's own writing. It seems more acceptable than it really is because the range of evidence discussed is so limited. Fodor declares that a comprehensive view of all relevant evidence lies outside the scope of his work, and that any criticisms of his theory along these lines are therefore unwarranted. For example, he comments that

"Modularity was not, however, an attempt to make the case for modularity; or to weigh the evidence; or to summarize the literature; or to dissect the alternatives.....Modularity provides a sketch .... together with a smattering of supporting evidence and an occasional indication of how to get around some of the data that were alleged to support the earlier story. But, as Forster says, the evidence is appealed to 'essentially for illustrative purposes'; so, for

that matter, are most of the arguments" (Fodor, 1985, p. 33).

However, it may be argued against Fodor that it is irresponsible to write an entire proposal for a radically different theory of the mind/brain, whilst systematically ignoring so many important pieces of evidence which weaken his theory, even if he admits to doing so. Only the most critical of readers are likely to heed his warnings about the provisional status of his theory.<sup>13</sup>

Fodor creates the impression that his theory is more coherent than it really is. This imbalance should and could have been redressed by a more thorough review both of the evidence that Fodor thinks does support his theory and that which might refute it. His critics have not misunderstood the purpose with which he writes. Rather, they are more aware than Fodor that even a tentative account of the mind/brain, particularly one with such extensive ramifications for the future direction of scientific research into this topic, should regard the evaluation of evidence as of paramount importance even at an early stage.

The sources of evidence that are pertinent to the Fodorean modularity thesis which are neglected by Fodor have been well documented by other critics (Shallice, 1984, 1988;

Marshall, 1984; Putnam, 1984; Fodor, 1985), and a brief outline of them is all that is necessary here.

Several of the peer commentaries in Fodor (1985) comment upon the sparsity of the evidence cited by Fodor or upon the way in which Fodor allows for reinterpretation of the evidence to suit his own theory to the extent that he runs the risk of rendering his theory unfalsifiable. For example, Robert J. Sternberg writes that

"[t]o some, Fodor's weighing of various sources of evidence might seem odd, if not downright bizarre" (Fodor, 1985, p. 33).

Sternberg's view is that the division between modular processes and non-modular ones is not a purely functional one, which separates perception from cognition. Rather, automatic processes are modular and controlled ones are not. The chess expert's skills are modular, on Sternberg's account, not for the reasons cited by Shanon or by Gardner, but because they have become automatic, unlike the early chess moves of the novice. If Sternberg is right, then his arguments cast doubt upon the strength of Fodor's evidence ( $E_2$ ) for his specific modular hypothesis  $H_2$  that modules actually form a natural kind. Moreover, although the functional division envisaged by Fodor between language and perception on the one hand and cognition on the other (the

general Fodorean hypothesis  $H_1$ ) is not directly affected by Sternberg's remarks, it will gain little confirmation from  $H_2$  if  $E_2$  can be shown to be either incorrect or incomplete.

It is clearly impossible to say whether or not Sternberg's arguments do weaken Fodor's case without a more thorough examination of the empirical evidence. Such an examination would be out of place here however, as it is not my aim to correct Fodor's omissions, but merely to demonstrate how easily his bootstrapping argument can be destroyed by the consideration of alternative interpretations of the evidence. As Sternberg himself comments

"Fodor is a persuasive advocate of his position, but he does not seriously consider alternative hypotheses, or he dismisses them cavalierly" (ibid, p. 33).

Some unsatisfactory arguments for central cognitive modules have already been noted in my discussion of Shanon in the preceding subsection (5.5.4). However, there are better arguments which cast doubt upon the validity of Fodor's inference from his general hypothesis that there are functional divisions within the mind/brain to the more specific view that his own particular modular hypothesis is the only viable one. Many of these arguments are linked to criticisms of Fodor's failure to provide a thorough survey

of the literature that is relevant to his hypothesis. For example, Marshall (1984) agrees with the importance of the modular concept and that it would be advantageous from an evolutionary point of view if input systems were to be modular. He writes that

"it would certainly make biological sense to have a mechanism that restricted the access of input systems to central processes ... provided, of course, that loud noises and sudden movements could automatically overcome that inhibition" (Marshall, 1984, p. 220).

However, he goes on to remark that Fodor's attempts to defend his theory against the potential counterexamples raised by the possibility that language modules may be influenced by context are impoverished by

"the very narrow range of literature he considers" (ibid, p. 223).

Putnam expands upon the difficulties associated with the modular nature of language by arguing that even if Fodor succeeds in escaping from the objection that word recognition is contextually facilitated, he will still be caught upon the opposing horn of the dilemma when required to account for how linguistic and visual modules can ever

acquire concepts when they are supposed to be informationally encapsulated.<sup>14</sup> Difficulties occur when Fodor attempts to explain how we can relate the output of our visual and linguistic modules to the information contained within our cognitive processes. How can we know that the modules are yielding reliable information? Even though they are informationally encapsulated this does not preclude the possibility of their being systematically wrong. We are forced to consider seriously the problems raised by the traditional veil-of-perception issue, namely, that we might genuinely be mistaken as to our perceptions and have no way of knowing it. Yet Fodor is attempting to put forward a scientific theory of the mind, and should surely pay more heed to such empirical questions.

Shallice comments that

"[e]legant and powerful theories based on sketchy data can potentially be dangerous as well as very valuable. Questions can receive a premature answer when what is most needed is that they merely be adequately stated" (Shallice, 1984, p. 244).

### 5.7 Fodor's modular theory and the veil-of-perception issue

Fodor's failure to present a critical account of this evidence means that fundamental flaws within his theory may remain undiscovered until we have progressed a long way down the wrong path or, worse still, that by eschewing the need for evidence, he is thereby rendering his theory immune from testability. Once more, the difficulties encountered by the metaphysical realist in supplying satisfactory support for his theory recur within a more scientific approach to the mind/brain, despite the use of the bootstrap strategy.

Fodor's theory contains a semantically inconsistent tetrad. Fodor takes the following four claims to be true:

I modules exist within those parts of the mind/brain that are responsible for language and perception

II these modules play a causal role in the process of conveying information from the environment to our central cognitive system

III we know that these modules exist because of our capacities and abilities, but we are not

directly aware of their existence

IV       nonetheless, there is ample evidence of their nature from neuroscientific and psychological sources.

The semantic inconsistency of these four statements is clear. First, Fodor asserts the existence of the modules (as described in statements I and II) on the basis of our acknowledged capacities and abilities (statement III). Yet, as my consideration of Putnam in the previous subsection suggests, many of our linguistic and perceptual capacities are not well accounted for by Fodor's modular hypothesis. This indicates that we do not have unequivocal evidence for the existence of a modular system within the mind/brain. Fodor may be vulnerable to the strong or ontological version of the veil-of-perception problem. Indeed, since he systematically ignores sources of evidence that cast doubt upon his hypotheses (see previous subsection), then he may be guilty of rendering his modular theory of the mind untestable.

Moreover, even if we grant that there is some kind of modular system within the mind/brain, we still cannot claim to know its nature on the basis of the neuroscientific and psychological evidence cited by Fodor (statement IV). As we have seen, there are strong suggestions that the



evidence strongly underdetermines theory selection. Note, for example, Sternberg's arguments in favour of a modularity thesis that focuses upon the expert/novice shift rather than upon a functional division within the mind/brain.<sup>15</sup> Fodor's failure to discuss such potentially disconfirming sources of evidence renders him vulnerable to the weaker, epistemological version of the veil-of-perception issue (as identified in Chapter 1, subsection 1.2.4).

### **5.8 Conclusion**

My contention is that Fodor bases his arguments on inadequate data so that his bootstrapping argument is not well confirmed, largely because there is insufficient consideration of the evidence in support of and against his specific modular hypothesis ( $H_2$ ). We have seen that there are many telling examples of specific disorders within the central cognitive systems, which cast doubt upon Fodor's denial that such systems are modular and that we will ever know a great deal about them (eg: Shallice, 1988, acalculia). I believe that Fodor's failure to consider the evidence is partly explained by the difficulties inherent in bootstrapping within a new field of science discussed in Chapter 2, subsections 2.5.1 - 2.6.

Here, Nordby's conclusions concerning the difficulties

involved in assessing rival theories by reference to formal criteria or abstract criteria seem very relevant to the modular debate (see Chapter 2, subsections 2.5.1, 2.5.2, 2.6). Fodor decides which evidence to include and which to exclude but, in the absence of any systematic theory, his decisions are less influenced by scientific considerations than by his own model of the mind/brain. As a result, Fodor risks either refutation or rendering his theory untestable. His modular theory, like his Representational Theory of Mind, fails to overcome the difficulties of metaphysical realism.

1. Much of this chapter has been published in two papers, Bennett (1988b) and Bennett (1990).
2. See, for example, McGinn's discussion of blindsight in Chapter 3, subsection 3.2.5.
3. Franz Gall's (1758-1828) phrenology was a forerunner of the current modularist attempt to reduce the role attributed to horizontal faculties of intelligence operating across all content domains and to isolate the existence of vertical faculties each of which would process information relating to one specific sense or subject, eg: music, or numbers. For a comparison of Fodor and Gall, see Gross (in Fodor, 1985).
4. In what follows, the letter C will be used to represent a hypothesis postulated by Gardner, his theory will be denoted M and the bodies of evidence in support of his hypotheses will be referred to as  $D_1$  and  $D_2$ . There is no significance attached to this change in my usual notation: it is purely introduced to avoid confusion with the notation used to represent the hypotheses and evidence within Fodor's theory, which will be discussed in conjunction with the Gardnerian view.

5. However, to avoid confusion with those theories that adopt an attitude of realism towards their postulated ontology of the mind/brain, I shall refrain from presenting Gardner's argument in the form of a semantically inconsistent tetrad.

6. This difficulty in Fodor's theory resembles those encountered in McGinn's arguments for the existence of a deep hidden structure within consciousness (see Chapter 3, subsection 3.2.4).

7. Gardner cites eight possible signs of an intelligence. These are: potential isolation by brain damage; the existence of idiots savants, autistic people and prodigies; an identifiable core operation which is sensitive to certain inputs; a distinctive developmental history; a plausible evolutionary history; support from experimental psychological tests; support from psychometric tests and susceptibility to encoding in a symbol system.

8. For example, I shall be arguing below (sub-section 5.6) that it is only by virtue of judicious selection of his sources that Fodor creates the impression that his modular theory is well confirmed by a variety of evidence. A different bibliography might well secure a very different result. Consequently, Fodor's modular hypothesis like that of Gardner, may not be supported by independent evidence, and therefore the bootstrapping fails.

9. It is also true that Fodor (1983) does not go on to develop this point further. However, explicit reference is made in Fodor (1985), p.4 to "the computational systems that deal with the perception/production of language" (emphasis mine) and to aphasias and agnosias, which may be located either in comprehension or in production. Shanon's valuable comments concerning modularity in production in Section 5 of his paper do not refute anything that Fodor has said, but rather flesh out a less adequate part of the Fodorean theory. Furthermore, many of the examples given there by Shanon, such as typing or playing a musical instrument, bear more of a striking resemblance to the modular intelligences described by Howard Gardner (1985) than they do to Fodorean input analyzers, since they involve a strong cultural influence upon a core processing mechanism. See also subsection 5.5.4, below.

10. See also section 5.6 (below) for other criticisms of Fodor's theory.

11. See also Achinstein's criticisms of the bootstrap argument in subsection 2.4.1 of Chapter 2, in which Achinstein claims that evidence can only support the constituent hypotheses of the bootstrap if we already accept the theory as true.

12. Fodor does not always rule out the possibility of modular effects within the central cognitive processes with such severity. He writes that "one might consider the possibility that Mother Nature, having tried peripheral modular mechanisms and found them good, then contrived, via the novice-expert shift, to simulate some of the effects of modularity at the level of

central systems" (Fodor, 1985, p.39). However, I do not think that this statement represents much of a concession by Fodor. In the preceding paragraph, he is scathing about the limited sense in which the novice-expert shift might give rise to modular effects, since it is always open to us to exercise conscious control over tasks usually conducted in an automatic, expert fashion. The novice-expert shift thus falls foul of objections along the lines of those I levelled against Shanon's workbench theory in subsection 5.5.4 above.

13. Compare Fodor's equally cavalier attitude towards criticisms of his Representational Theory of Mind, Chapter 4, subsection 4.4.2.

14. Analogous problems have been discussed with reference to Fodor's theory of how overt languages can be learned and can develop from an innate and therefore rigid inner mental language in the previous chapter (subsections 4.4.2, 4.4.3).

15. See also Shallice (1988, p. 273) for a suggestion that phenomena such as acalculia cast doubt upon Fodor's claim that the modular/non-modular distinction corresponds exactly with the boundary between the input systems and the central processing unit.

## Chapter 6 - Metaphysical realism in co-evolutionary theories of the mind/brain<sup>1</sup>

### 6.0 Introduction

In Section A, I established that the difficulties of metaphysical realism can only be avoided by a bootstrapping argument if there is a satisfactory relationship between the hypotheses contained within a theory and the evidence that is intended to support them (see Chapter 2, subsections 2.4, 2.4.1). It is not enough to cite evidence that is merely compatible with the hypotheses that we wish to confirm true relative to the overall theory; we must also ensure that this evidence is properly explained by the hypotheses under scrutiny. Moreover, we saw that circularity can only be avoided if the bodies of evidence used to support each hypothesis are arrived at by independent paths.

I also argued in Section A that philosophers and scientists who are breaking new ground in their attempts to formulate theories do not in fact make use of the formal criteria favoured by Glymour (1980) for deciding which theory is to be preferred in cases of empirical underdetermination. (See Chapter 2, subsections 2.4 - 2.5.2). Rather, they, like the detective discussed in Nordby (1989), "bootstrap

while barefoot". This means that, in the absence of any pre-existing overall guidelines concerning the relative importance of a variety of sources of evidence, they will make decisions about what counts as acceptable evidence in support of their hypotheses based upon their own preconceptions of what kind of theory they should be looking for.

Thus far in section B, I have demonstrated that the issue of realism should be regarded as of the same importance in the study of the mind/brain as it is in other branches of philosophy and science (Introduction to Section B, subsection B2). This is acknowledged on a superficial level by both of the philosophers whose arguments I have examined in detail in Chapters 3 - 5. McGinn and Fodor pay lip service to the ideals of scientific realism and urge the view that their own theories have been constructed using only evidence that is both testable and supportive of their hypotheses. However, I have contended that their use of the bootstrap strategy is flawed, since the bodies of evidence upon which they rely are clearly inadequate for the tasks that they are intended to perform. As a result, their theories may be reconstructed in the form of semantically inconsistent tetrads which correspond to that created by my reconstruction of the traditional veil-of-perception issue in Chapter 1 (subsection 1.2.4). Semantic inconsistency is a key characteristic of metaphysical

realism (see Chapter 1, subsection 1.2.4, and Chapter 2, subsection 2.1).

I have suggested that the elements of metaphysical realism within McGinn's theory of the structure of consciousness and in the Fodorean Representational Theory of Mind derive in the main from poor conceptual analysis. However, to the extent that they attempt to prop up their philosophical arguments with piecemeal evidence from neurophysiology and psychology, their efforts to construct a theory by bootstrapping also suffer from the complications inherent in the task of combining evidence from a variety of sources in the absence of any established criteria as to how this should be done.

It might be imagined that the recent trend of investigating the workings of the mind/brain by means of a combination of philosophical argument and scientific activity comprising both computational and neurophysiological research (the so-called co-evolutionary approach) would have resulted in the development of a bootstrapping methodology capable of avoiding these problems of evidence and realism.

However, we have already seen in Chapter 5 that, although all the contributors to the modularity debate made reference to neurophysiological data to bolster their conceptual arguments, this did not help to resolve their

dispute since I was still able to argue that their theories were underdetermined with respect to this evidence (see, for example, Chapter 5, subsection 5.4, in which Fodor's theory is compared with that of Gardner). Of course, it is perfectly possible that the problems that I have isolated in the treatment of neurophysiological data by this particular group of theorists are simply a result of a poor understanding on their part of what is required for evidence to support a theory. Others may have succeeded in producing a successful bootstrap methodology which is capable of the successful integration of philosophical arguments with information derived from neurological study of the mind/brain, and which may also allow for the incorporation of data obtained from additional scientific enterprises such as computational research and psychological experiments.

This chapter has two main purposes. Firstly, I intend to establish that the failure to develop such a methodology is not confined to the modularity theorists, by examining the bootstrap arguments of two further leading exponents of the co-evolutionary approach, David Marr and Patricia Smith Churchland<sup>2</sup>.

Secondly, I shall argue that, although the treatment of the evidence by co-evolutionary theorists of the mind/brain is undoubtedly haphazard and frequently biased in favour of



their own preferred hypotheses, this situation is a natural consequence of the very novelty of the study of the mind/brain using a plethora of different disciplines, each with their own criteria for testability and satisfactory evidence. This means that the difficulties of bootstrapping whilst barefoot are as integral to the co-evolutionary approach to the mind/brain problem as they are to the task of solving serial killings as described in Nordby (1989).

The discussion of Churchland's work in the second half of this chapter will also sustain my subsidiary claim that Churchland is quick to notice the need for hypotheses to be supported by satisfactory evidence when she is criticising theories of the mind/brain put forward by others, but slow to realise the implications of these issues when developing her own bootstrap argument. This deficit in her method greatly weakens her assertion that her co-evolutionary approach to the mind/brain is better supported by the available evidence than the Fodorean sentential paradigm.

This point will be discussed at length in the second half of this chapter. I begin with an examination of Marr's theory of vision. Marr is important not only for his contribution to our understanding of the process of vision, but also for his methodological arguments detailing the need to view the mind/brain as an information processor.

## 6.1 Marr and metaphysical realism

### 6.1.1 Marr's theory in outline

Marr conceives of vision in very much the same terms as Aristotle did. He comments that

"vision is the *process* of discovering from images what is present in the world, and where it is" (Marr, 1982a, p. 3, his italics).

Marr proposes a bifurcated study of vision. We must, as the above quotation indicates, regard vision as a kind of information-processing, but it is also imperative to investigate the nature of the system of internal representations that is a prerequisite of any information-processing<sup>3</sup>. As we noted in the Introduction to Section B (subsection B3), each species will employ a different representational system capable of emphasising the information most vital to its survival.

This might lead us to expect a heavy emphasis upon neurological details within Marr's theory. Surprisingly, however, Marr warns against excessive preoccupation with the investigation of neural activity on the grounds that

"neurophysiology and psychophysics have as their business to *describe* the behavior of cells or of subjects but not to *explain* such behavior" (ibid, p. 15).

Although Marr concentrates upon vision in his own work, he regards his methodology as the correct approach for exploring any aspect of mind/brain function. The fundamental tenet of this methodology is that no explanation can ever be complete if it does not include a consideration of both *what* task is being carried out by a process or system and *why* it is being carried out (ibid, p. 22). Marr has christened this part of an explanation the *Computational theory*<sup>4</sup>, but it is really the main hypothesis within a larger theory of how we should tackle the problems inherent in trying to understand the mind/brain, and specifically, the visual process<sup>5</sup>.

Marr not only believes that this Computational level of explanation is an essential element of any investigation into the workings of the mind/brain, he also thinks that it must form our starting point. Any explanation that proceeds too hastily to draw conclusions from research into computer modelling of mind/brain processes or from neurophysiological data runs the risk of formulating misguided hypotheses. The reason for the Computational element within Marr's theory is that

"[t]here must exist an additional level of understanding at which the character of the information-processing tasks carried out during perception are analyzed and understood in a way that is independent of the particular mechanisms and structures that implement them in our heads....Such analysis does not usurp an understanding at the other levels - of neurons or of computer programs - but it is a necessary complement to them, since without it there can be no real understanding of the function of all those neurons" (ibid, p. 19).

Once we have decided what information-processing process is being carried out, it is defined in abstract terms which indicate the precise manner in which one set of data is mapped onto another and the constraints within which this must be carried out (ibid, pp. 23-4)<sup>6</sup>. Only then may we proceed to an examination of the issues that relate to the second and third levels of explanation.

Marr's second level of explanation requires the selection of appropriate representations for the input and output of our brain activity and of an algorithm for the transformation of the informational content of the input into the informational content of the output. These selections may well be interdependent according to Marr;

algorithms may only be practical given certain representational schemata for input and output. Both choice of representation and choice of algorithm are, of course, dictated by the constraints delineated at the first, more abstract level of explanation. Other considerations may influence the choice of algorithm too, such as efficiency.

Finally, Marr incorporates neurophysiological information into the third level of explanation involved in a study of the mind/brain. Levels two and three are interdependent in the sense that the choice of algorithm may be dictated in part by the nature of the physical instantiation of the information-processing system as revealed by neurophysiological research.

This means that although Marr subscribes to a broadly co-evolutionary approach for the exploration of the mind/brain, believing that

"the real power of the approach lies in the integration of all three levels of attack" (ibid, p. 330),

he does not share Churchland's enthusiasm for incorporating neurological information into a theory of mind/brain function from its inception<sup>7</sup>. Rather, Marr is a top-down

theorist who emphasizes the importance of logical, digitalized information-processing. Indeed, he would agree with those whom Churchland characterises as "purists of the top-down persuasion" that

"the cardinal article of faith is that first you figure out what the mind-brain does, and secondarily you find out how it might implement the functions described" (1986a, p.419).

I shall now give a very brief summary of the main features of Marr's theory of vision. It should, of course, be borne in mind that the theory is actually far more complicated than it will appear in this description, but none of my remarks about the success of Marr's bootstrap strategy will hinge upon the details that have been omitted<sup>8</sup>.

The starting point of Marr's explanation of vision is, as we would expect, a general Computational hypothesis which describes the key stages of visual information-processing in terms of the tasks that they perform. The first stage of vision is the creation of the primal sketch from the retinal image to reveal important features of the two-dimensional image which, when combined according to particular rules, may indicate the positions of edges of objects in the external world. Secondly, the primal sketch is transformed into a  $2\frac{1}{2}$  dimensional sketch, which,

although still viewer- rather than object-centred, yields further information about the position and direction of surfaces. Here, the constraints formulated at the Computational level are of paramount importance, since the information contained in primal sketches is frequently compatible with more than one  $2\frac{1}{2}$ -D sketch. If there were no constraints upon the possible interpretations of this information, then our visual system would not be able to cope with such instances of empirical underdetermination.

Finally, it is important for the viewer to have representations of the world that may be used for purposes of object recognition. This results in a 3-D model representation with the task of describing

"shapes and their spatial organization in an object-centered coordinate frame, using a modular hierarchical representation that includes volumetric primitives (i.e., primitives that represent the volume of space that a shape occupies) as well as surface primitives" (ibid, p. 37).

These abstract characterisations of the three major stages of visual processing constitute what Kitcher (1988, p. 5) terms Marr's global Computational hypothesis of vision. Marr also postulates several local Computational theories

about the visual process, each of which delineates the contribution of a particular subprocess towards one of the three stages outlined above (Kitcher, 1988, p. 8). A local Computational hypothesis will attempt to identify the constraints that must operate to avoid empirical underdetermination of the output of the given subprocess by the input, by first assuming values for inputs and outputs, and then providing a mathematical confirmation that unique outputs are indeed derived for each input and the constraints postulated. Only then can we tackle the questions of algorithms and suitable representations for neurophysiological instantiation.

#### **6.1.2 The bootstrap strategy within Marr's theory**

It should be clear from the previous subsection that Marr is engaged in two main tasks. Firstly, he is concerned to develop a paradigm methodology for the investigation of all aspects of mind/brain function. Secondly, he argues for a specific theory of vision, which is derived from the rigorous application of the general methodological approach that he endorses. Marr sometimes runs these two strands of his work together in a rather confusing manner, but it is clear from much of the argument in the introductory chapter of Marr (1982a, see especially pp. 24-29) that Marr regards his discussion as relevant to the study of all information-



processing devices. For example, Marr declares that there are

"different levels at which an information-processing device must be understood before one can be said to have understood it completely" (ibid, p. 24).

Consequently, it is possible to discern more than one use of the bootstrap strategy within Marr's writings. Within Marr's overall theory of information-processing within the mind/brain, T, Marr's first hypothesis ( $A_1$ ) is the general Computational hypothesis that the best way to understand the processes and representations within the mind/brain is to use the three levels of explanation described in subsection 6.1.1 above, starting with an abstract, mathematically rigorous characterizations both of the task performed by the process or representation and of the constraints operating in the world outside the mind/brain which facilitate this performance.

Marr is emphatic that this general methodological principle is the only possible way of obtaining an satisfactory explanation. He argues that previous attempts to understand the workings of the mind/brain have failed precisely because they did not incorporate such a rigorous, highly abstract description of what tasks are being carried

out. No amount of computer modelling can adequately demonstrate that the same mechanism is actually employed by the mind/brain. Marr claims that

"the critical point is that understanding computers is different from understanding computations. To understand a computer, one has to study that computer. To understand an information-processing task, one has to study that information-processing task. To understand fully a particular machine carrying out a particular information-processing task, one has to do both things. Neither alone will suffice" (ibid, p. 5).

Nor can we really learn how a task is being carried out simply by isolating the neurons responsible (see above, subsection 6.1.1). Research projects on computer programming or on neuronal function cannot, either severally or together, yield all the information necessary for a full understanding of the processes and representations at work but, when guided by a Computational description of the tasks under scrutiny, they furnish the student with vital additional perspectives on the issues. We may term these considerations Marr's body of evidence  $B_1$ , which he regards as support for his general hypothesis  $A_1$ .

In the light of these considerations, Marr argues that we should tackle the specific question of vision using the methodology outlined in hypothesis  $A_1$ . Here we see that Marr, like Fodor and McGinn (see Chapter 3, subsection 3.2.2, and Chapter 4, subsection 4.2.2) makes use of the bootstrap argument to infer a specific hypothesis which is founded upon the same evidence as the general hypothesis. Let us call the entire issue of vision Marr's second hypothesis,  $A_2$ .

It is here that the situation becomes complex. Marr relies upon evidence that his treatment of vision is successful in order to provide further confirmation of his initial general methodological Computational hypothesis ( $A_1$ ). Let us call this required body of evidence  $B_2$ . Clearly,  $B_2$  must contain strong indications that Marr's handling of the question of vision is not only compatible with what we already know about vision, but is a better explanation of the phenomenon than any alternatives which do not make use of his Computational methodology. That is to say, as we would expect, we must seek the required evidence  $B_2$  by evaluating the hypothesis  $A_2$ .

However, Marr's account of vision is constructed using a second bootstrapping argument, very similar to the first. This is recognized by Kitcher who remarks that

"[t]he global and local Computational theories constrain each other - here "constrain" in the sense of preventing a theory from being wild speculation. The global theory describes the large stages, the local theories explain how those large Computations are carried out, and the two levels must fit together. So the local Computational theories are constrained from above, and they are constrained from below" (Kitcher, 1988, p. 21).

Kitcher thinks that the use of such a bootstrap strategy ensures that the overall theory of vision is testable. This view will be discussed below (subsection 6.1.3, part B).

Indeed,  $A_2$  is not only the second hypothesis of the bootstrap argument relative to Marr's general theory,  $T$ , of information-processing within the mind/brain, but is also Marr's specific theory of vision. Relative to this theory then, Marr postulates further hypotheses. The first,  $V_1$ , is the general Computational hypothesis that the entire question of vision must be approached using the three levels of explanation outlined in the previous subsection, beginning with an rigorous abstract characterization in mathematical terms of the tasks performed by each of the three key stages within the visual process and the

constraints at work in the world outside the perceiver which make those tasks possible. Moreover, Marr holds that vision is modular, that is to say, it is made up of a number of subprocesses, each of which can be studied independently.

We saw earlier that Marr holds that a comprehensive understanding of any information-processing device necessarily involves explanation at the three different levels described in the previous section. I commented that he arrived at this view as a result of his realization that explanations which failed to exploit the idea that the mind/brain is used for information processing. Similarly, Marr chooses to defend  $V_1$ , his corresponding general hypothesis about how we should approach the study of vision, by reference to earlier accounts which have failed to understand that vision is simply a kind of information-processing. For example, Marr comments that, prior to his treatment of the problem, nobody had succeeded in understanding stereopsis because

"[n]one of them formulated the computational problem precisely at the top level, and almost all the proposed networks actually computed the wrong thing.....people became so entranced by the mechanisms for doing something that they erroneously thought they understood it well

enough to build machinery for it" (Marr, 1982a, p. 336).

Marr analyses Gibson's failure to provide a fully satisfactory account of the visual process in the same way. He comments that

"although some aspects of [Gibson's] thinking were on the right lines, he did not understand properly what information processing was, which led him to seriously underestimate the complexity of the information-processing problems involved in vision and the consequent subtlety that is necessary in approaching them" (ibid, p. 29).

Moreover, Marr is adamant that any satisfactory theory of vision must explain why our perceptions of the external world display an overwhelming tendency to be accurate. This is his justification for including the calculation of constraints upon the world as part of the explanatory device described in  $V_1$ . Marr writes that

"[t]he critical act in formulating computational theories for such processes is the discovery of valid constraints on the way the world behaves that provide sufficient additional information to

allow recovery of the desired characteristic [of a scene from images of it]" (ibid, p. 330).

Marr also cites evidence in support of his claim that vision may be thought of in terms of a number of modular subprocesses. The computer-generated work of Bela Julesz on stereopsis yields this conclusion, as well as lending further support to the strategy of beginning the investigation of any subprocess at the Computational level. Marr concludes that

"observations like Bela Julesz's are extremely valuable because they enable us to formulate clear computational questions that we know must have answers because the human visual system can carry out the task in question" (ibid, p. 102).

This also illustrates the importance of an integrated approach to the issue of vision. The computer generations of Julesz suggest that it is possible to find answers to the questions stimulated by viewing vision as information-processing.

These various considerations are regarded by Marr as evidence in support of his initial general hypothesis about how we should study vision ( $V_1$ ). They will be referred to collectively in what follows as  $E_1$ .

Marr's next step is to argue that, since vision is modular, we can apply the strategy of explanation on three levels to each of the subprocesses which contribute towards the construction of one or another of the three key stages of vision outlined above in subsection 6.1.1, the primal sketch, the  $2\frac{1}{2}$ -D sketch and the 3-D sketch. This means that he bootstraps from  $V_1$ , supported by  $E_1$ , to the possible construction of a whole host of specific hypotheses about the nature of a variety of visual subprocesses, such as stereopsis, directional selectivity, and apparent motion. These hypotheses are the local Computational hypotheses referred to by Kitcher (1988, p. 8), and each consists in a three-level account of a particular subprocess. Marr notes that

"general observations can often lead to the formulation of a particular process or representational theory, specific examples of which can be programmed or subjected to detailed psychophysical testing. Once we have sufficient confidence in the correctness of the process or representation at this level, we can inquire about its detailed implementation, which involves the ultimate and very difficult problems of neurophysiology and neuroanatomy" (ibid, p. 331).



I shall refer to these accounts collectively as hypothesis  $V_2$ . It will not be necessary to describe these accounts in great detail at this stage<sup>9</sup>. Marr is relying upon the success of these Computational accounts of the visual subprocesses to provide him with a body of favourable evidence  $E_2$ , so that further confirmation is lent to  $V_1$ . Such confirmation would amount to a vindication of Marr's preferred methodology for the study of vision.

The successful construction of a theory of vision through the use of the bootstrap strategy is thought by Marr to provide confirmation in its turn for his original hypothesis that the processes and representations within the mind/brain should be analyzed as information-processors. As noted above, Marr believes that an evaluation of his theory of vision ( $A_2$ ) will yield evidence that his general methodological approach as outlined in hypothesis  $A_1$  has been vindicated. The body of evidence  $B_2$  within his initial, more general bootstrap argument consists of the success of Marr's use of his three-tiered system of explanation to produce a more comprehensive study of vision than rival theories, and one which is compatible with other facts that we have about the nature of the mind/brain.

The next subsection comprises an evaluation of Marr's use of the bootstrap strategy to see if it can sustain his two

main claims: firstly, that his three-tiered approach to understanding the mind/brain is the optimum method for study in this field, and secondly, that this method has afforded a satisfactory and comprehensive theory of vision.

### 6.1.3 Criticism of Marr's bootstrap strategy

#### A - Introduction

Before I discuss the flaws within Marr's bootstrap strategy, it is worth reiterating the vital criteria that Marr's argument must satisfy if it is to succeed in providing us with the outline of a methodology which will enable us to explain not only the processes involved in vision, but also other processes within the mind/brain. Like any other explanation, Marr's account must supply us with new information in a testable format. If parts of his theory are not testable, then he is vulnerable to the criticism that he is resting his hypotheses upon unwarranted assumptions, with the result that his argument cannot be said to satisfy the demands of scientific realism as described in Chapter 2 (subsection 2.2.2). This would clearly be disastrous for Marr, who is a realist about the existence of representations and processes within the mind/brain<sup>10</sup>.

Although Marr does make some attempts to seek confirmation of the accuracy of his abstract task descriptions of visual processes and representations using computer simulations and neurophysiological research, his bootstrapping is rather lopsided, since at every stage in his argument he concentrates his discussion upon the issues raised by the

use of the information-processing concept<sup>11</sup>. The neuronal implementation suffers particularly from neglect. His priorities are clearly illustrated in the following excerpt from his treatment of stereopsis. Marr declares that

"A complete neural implementation of the second stereo matching algorithm just described has not yet been formulated. One reason is that *such a formulation was not worth the considerable work involved until we were reasonably certain from implementation studies and psychophysics that the algorithm works and is roughly correct*" (ibid, p. 152, my italics).

Although Marr sometimes demonstrates his awareness of the need to provide satisfactory evidence for his arguments wherever possible (see, for example, Marr, 1982a, p. 111), this is not necessarily enough. In Chapter 4 we noted a clear discrepancy between Fodor's ostensible acknowledgement of the onus upon him to provide reliable, testable sources of evidence in support of his sentential view of the mind/brain, and his actual failure to do so (see Chapter 4, subsections 4.3, 4.4).

A detailed examination of Marr's local Computational accounts of the subprocesses involved in the construction of the three key representational stages within vision is

not needed here. This is partly because many of the criticisms of the bootstrap strategy relate to the general methodology used, rather than to the nature of the individual subprocesses. Moreover, detailed criticisms of Marr's views by others have already revealed potential difficulties within his accounts of these subprocesses (see Kitcher, 1988 for a useful list of references). It will become clear in the following two parts of this subsection that such potential difficulties are enough to cast doubt upon the strength of Marr's evidence in support of his strategy.

Further, very similar hypotheses and evidence are relied upon at both levels of Marr's bootstrapping strategy, both levels consisting in an initial general methodological hypothesis licensing the extension of the three-level approach to the study of processes and representations to deal with more specific issues. This initial hypothesis is then confirmed to the degree that the approach succeeds in throwing more light on these issues. As a result of this similarity, objections made against the methodology within one level of the bootstrapping argument will cast doubt upon the validity of the entire approach.

There are two main lines of criticism of the strategy. The first relates to Marr's postulation of constraints upon the nature of the external world as necessary to solve the

difficulties created by empirical underdetermination. The second is concerned with the cogency of a methodology which takes information-processing as a starting-point for the study of the mind/brain, its processes and its representations.

I shall consider each of these criticisms in turn.

**B - The status of constraints within Marr's bootstrap argument**

Marr is aware that vision should provide us with a view of the external world that is, by and large, accurate. However, this is sometimes made difficult by the possibility of gross empirical underdetermination when information is being mapped from one representational stage to another. For example, the process of stereopsis, which is used to help extract information about surfaces from the primal sketch, exhibits exactly this problem. Similar difficulties arise in the final mapping of the information contained in a viewer-centred  $2\frac{1}{2}$ -D sketch onto an object-centred 3-D model (see Marr, 1982a, pp. 317-8). Constraints in the external world must therefore be identified, so that the number of possible mappings is reduced to one. Marr notes that in the case of stereopsis

"[t]he critical step .... is the discovery of additional constraints on the process that are imposed naturally and that limit the result sufficiently to allow a unique solution" (ibid, p. 104).

Indeed, the use of constraints is an integral part of the top level of Marr's Computational methodology. However, these constraints are based upon *a priori* judgements rather than upon empirical data. Marr admits this in his discussion of silhouettes. He notes that

"there must be some *a priori* [sic] assumptions in the way we interpret silhouettes that allow us to infer a shape from an outline" (ibid, p. 219).

It is this aspect of Marr's approach that inspires Boden's comparison of his theory of vision to Kant's transcendental deduction. Like Kant, Marr is claiming that we have some kind of predisposition to see the world in a certain way. Boden writes that Marr's constraints

"make universal claims about what the world must be like, and what see-ers must be like, if vision is to be possible" (Boden, 1989, p.41).

It is the presence of these *a priori* elements within Marr's bootstrapping arguments which renders Marr susceptible to the same kinds of difficulties relating to evidence and realism as those that were levelled at the theories of McGinn, Fodor and the modularity theorists earlier in Section B. Indeed, Morgan draws an explicit comparison by commenting that

"[t]his is just the sort of argument that Fodor and other language nativists use to assert that a child could never learn the meaning of even common nouns, unless they made strong assumptions about the nature of objects, their permanence, and so on" (Morgan, 1984, p. 160).

Before I go on to demonstrate that Marr's argument about the nature of constraints may be reconstructed as a semantically inconsistent tetrad, it is worth developing the criticism a little more.

Kitcher (1988, p. 20) comments that fault has been found by other critics of Marr's work (such as Todd, see Kitcher, 1988, p. 20) in the nature of the particular constraints used to develop the local Computational accounts of visual subprocesses that make up hypothesis  $V_2^{12}$  within Marr's theory of vision. These criticisms are not in themselves important for my argument: I am interested solely in the



fact that they raise the possibility that Marr's selection of these particular constraints might be wrong, and the consequences of this error for the remainder of the bootstrap theory.

Since the constraints operating upon the external world help to define the information-processing task being carried out by a particular subprocess, mistakes at this level could well result in the adoption of the wrong algorithm and the failure of many of the accounts within  $V_2$ . This would result in the undermining of the body of evidence  $E_2$  in support of  $V_2$ , and deprive Marr of the additional confirmation of his general hypothesis ( $V_1$ ), which advocates the use of a three-tiered approach to the study of vision. Kitcher is quick to see these implications too, and argues that

"[w]hether this result had further ramifications would depend on what the correct theories (or theory) of the process look(s) like. If they still had constraints of some sort, then it would seem that Marr was right about the form of a theory of vision, but wrong in detail. On the other hand, we might figure out that the rigidity constraint is wrong, and still have no very good idea about how the process does work" (Kitcher, 1988, p. 20-21).

However, Kitcher has failed to see an important difficulty with her analysis of the situation. She is right to argue from the specific points raised by Marr's critics to the need for a re-evaluation of Marr's general theory, but there is a more fundamental problem. For, if we are to take Marr at face value, the constraints governing the development of our explanation of visual subprocesses are not arrived at via a process of experiments and testing, but via a consideration of *a priori* notions about what the world must be like for vision to be possible at all. This means that they cannot be rejected by experimental evidence (and are thereby rendered immune from the arguments put forward by critics like Todd). The only way to criticize Marr's view is to argue that he is simply mistaken about the nature of the physical laws which constrain how the external world must appear to us and to all other creatures with the capacity for vision. However, this does not mean that there are no constraints at all, merely that they are not the ones that Marr first thought that they were. Since we can neither prove nor refute Marr's ideas by experiment, they are not testable. Boden summarizes the situation as follows:

"[t]ranscendental deductions are not immune to criticism, but they cannot be challenged experimentally. They are based on a *a priori*

argument, not on isolated facts gained through particular experience" (Boden, 1989, p. 42).

This places Marr at odds with the rest of science and lays him open to the accusation of metaphysical realism. It is easy to show that his argument about the constraints on vision may be reconstructed as a semantically inconsistent tetrad which displays the same problems with respect to evidence as those underlying the assumptions of Fodor, McGinn and the modularity theorists.

Marr holds the following four statements to be true:

I there is an external reality over and above our perceptions

II certain constraining features of this reality enable us to perceive it, hence they play a causal role in the construction of our perceptions

III we can only ever have direct knowledge of our constrained perceptions; we can never have any knowledge of the external world except through our perceptions, which are themselves mediated by constraints, the nature of which we can neither confirm nor refute by experiment

IV our perceptions furnish us with an outlook on the world which is by and large accurate because of these constraints

Since we only ever have direct knowledge of our perceptions, which we take to be representations of an outer reality, but lack any source of evidence for the character of the constraints which dictate the nature of our perceptions (statement III), we can have no grounds for asserting that there are such constraints upon reality which make vision possible (statements I and II). Here, once again, is another reformulation of the *ontological* version of the veil-of-perception.

Moreover, even if we simply assume that such constraints exist, we have no reason to believe that we will ever have any way of knowing what they are, and therefore no reason for believing that our perceptions of the external world are accurate (statement IV). This constitutes our weaker *epistemological* version of the veil-of-perception argument, as it applies to Marr.

Boden claims that

"Marr's 'transcendental deduction' is not absolutely *a priori*, since it takes for granted empirical knowledge of the world, as opposed to world-independent metaphysics" (ibid, p. 42).

It is for this reason that we may argue empirically that Marr has systematically misdescribed the nature of the world that we see. This weakens the accusation levelled against Marr that he falls into the strong version of the veil-of-perception trap. However, since we cannot test individual constraints by experiment, it is equally clear that Marr does suffer the difficulties associated with the weaker version of the veil-of-perception issue. We cannot know which postulated constraints really do exist and which do not.

Marr's theory of vision is therefore greatly weakened. If no evidence is forthcoming regarding the nature of the constraints he assumes in the formulation of his local Computational accounts of visual subprocesses ( $V_2$ ), then the evidence of their success ( $E_2$ ) is also rendered dubious, and no further confirmation of his general methodological approach to vision as described in hypothesis  $V_1$  can be obtained. Indeed, this consequence accords with our intuitions, since, if it is not possible to test the assumptions that Marr makes about the external world, then

this difficulty is one inherent in his entire methodology.

Moreover, it was argued in subsection 6.1.2 that increased confidence in Marr's general methodological principle ( $A_1$ ) was declared to depend upon its success in providing us with a satisfactory theory of vision. Clearly, the difficulties noted with regard to Marr's account of the constraints upon vision will therefore impinge upon his general theory of how the mind/brain should be studied. The existence of non-testable constraints is undesirable in the theory of any aspect of the mind/brain, and so Marr's general hypothesis ( $A_1$ ) may be declared unsatisfactory in the light of the body of evidence  $B_2$  derived from an analysis of the success of the bootstrap argument as it is applied within the theory of vision.

Thus far, we see that Marr's problems with realism and evidence arise as a result of poor conceptual argument, based upon assumptions about the nature of the external world that it is not possible to test. In Part C, I shall evince a second objection to Marr's theory which concentrates upon the attempt to combine abstract, information-processing considerations with neurophysiological evidence.

C - The value of the information-processing approach to the mind/brain problem

I shall commence this part of the discussion by seeing if it is possible to mitigate the criticisms of Marr evinced in Part B (above).

Boden's criticisms of Marr certainly make him sound like an external or metaphysical realist. She writes, for example, that

"Marr set himself the 'Kantian' task of specifying universal constraints, or ontological features, by virtue of which perception is made possible....these constraints set limits on the nature of visual systems, and also on the nature of visible worlds" (ibid, p.38).

This appears to be a clear-cut case of metaphysical realism. The external world has to take a certain form for vision to be possible.

However, it is also possible to argue that Marr is an *internal* and not a metaphysical realist. Some evidence that Marr's theory might be viewed along these lines occurs in Kitcher (1988) in a discussion of how Marr's local and

global theories may constrain each other. Kitcher writes that

"We can get positive confirmation .... by finding data that are best explained by the existence of the hypothesized representation" (Kitcher, 1988, p. 21).

This sounds like internal realism. Moreover, Marr is certainly not unaware of the problems inherent in his use of constraints within his theory of vision. He intimates in the following passage that he may be prepared to treat them, not as *a priori*, but as revisable in the light of empirical evidence. Marr declares that

"there is no hope of understanding the processes properly until some other means have been found for determining what is safe to assume [sic] about the world and what is not, together with the related question of the reliability of the different kinds of information" (Marr, 1982a, p. 266).

Perhaps Marr is saying that since we know vision to be possible, we must form theories about the world that will explain this in the best possible way.



But here we encounter the second objection to Marr's theory. Marr claims that "the best possible way" to study vision is, by adopting an approach that is, first and foremost, based upon viewing vision as an kind of information-processing and formulating a high level task description of what takes place within that part of the mind/brain that deals with vision. However, evolution cannot always be relied upon to solve problems using processes that are perfect from a design stance. For example, it might not be possible to integrate such processes with other processes within the system. Indeed, rough and ready solutions are often pioneered by nature. Morgan cites one example within Marr's account of visual subprocesses which is not well supported by the available neurophysiological and psychophysical evidence. This is the formation of zero-crossings, distinctive indications of a change in intensity within a visual image, which are useful in the mapping of information from the basic image onto a primal sketch. Morgan argues that

"[e]legant though the ZC solution is, Marr offers no evidence for its correctness" (Morgan, 1984, p. 162).

Of course, it would be perfectly acceptable for Marr to regard such solutions as mere hypotheses which are then open to refutation by neurophysiological and

psychophysiological data, or even by information from research into computation. The problems arise from the suspicion that Marr overemphasises the importance of the information-processing approach at the cost of ignoring possible evidence that his theoretical accounts of the visual subprocesses are wrong. Boden's characterizes Marr's methodology as

"a combination of a priori argument and empirical study (including computer-modelling). His theoretical primitives for each representational stage were suggested in the light of abstract computational considerations, tested by being embodied in computer models, and also judged by psychological evidence" (Boden, 1988, p.60).

This may well be an accurate description of Marr's intended methodology, but if Morgan is correct, then Marr may well be guilty of placing too much weight upon the abstract, high level descriptions of the visual processes. Consequently, he runs the risk of producing incorrect local Computational accounts of what is actually going on within the mind/brain, with the result that, once again, he is unable to maintain his claim that the success of such accounts constitutes a sound body of evidence ( $E_2$ ) which yields further support for his general approach to vision as embodied in  $V_1$ .

This criticism is supported by Kitcher, who comments that "[i]nsofar as Marr and his co-workers can devise experiments that reveal vision as an 'elegant contrivance', then the Computational approach of functional decomposition into optimally designed stages will seem attractive. If they are wrong in this very substantial assumption, however, then the unified theory of vision that Marr dreamt of will be impossible. The passages from global Computational theories to local theories, and to algorithms, and to biological hardware will not be at all smooth, or perhaps even possible" (Kitcher, 1988, p. 23).

This demonstrates that even if Marr can be shown to accept the position of a scientific realist with regard to the status of constraints upon the external world (at least some of the time), his problems are still not over. His decision to relegate the study of neurons to a relatively unimportant place in his theory may result in his rejection of nature's own solutions to the problems of vision.<sup>13</sup> Instead, Marr may be advocating accounts which merely describe a possible solution, rather than explaining what is actually happening within the mind/brain. If this should turn out to be the case, then his methodology no more leads to an explanatory account of vision than do the neurophysiological theories that he was so quick to

criticise (see above, subsection 6.1.1). This deprives his bootstrap argument of its very motivation!

Moreover, the failure of Marr's three-tiered approach to vision to provide accurate explanations will also reduce the amount of confirmation sought by Marr for his general hypothesis ( $A_1$ ) that the mind/brain question should also be tackled using a three-tiered approach. The difficulties noted within the theory of vision conflict with Marr's claim that the success of this theory will vindicate the use of the methodology to study other processes and representations within the mind/brain.

Worse still, we noted above that the hypotheses within the first, more general bootstrap argument mirror those contained within the theory of vision. This means that the objection to the information-processing approach is equally pertinent here. If it is used to increase our understanding of other capacities within the mind/brain, we may still obtain accounts that conflict with the information that we have about the behaviour and configuration of neurons. This point is addressed by Churchland, who comments that

"evolution proceeds by building on structures already in place; it cannot begin from scratch, even though considerations of optimal design

might favour this" (P.S. Churchland, 1988, p. 396).

Andy Clark also supports my argument against over-concentration upon the information-processing aspect of mind/brain function. He writes that if

"evolution proceeds as a tinkerer, each step in the evolutionary chain exploits a net historical opportunity whose nature is determined by whatever materials happen to be available to adapt to a new requirement...This historical snowballing effect, combined with the need to achieve some workable total system at each modification..., often makes natural solutions rather opaque from a design-orientated perspective" (Clark, 1989, pp. 70-71).<sup>14</sup>

It was noted in Chapter 2, subsection 2.5.1 that Glymour can be criticised for advocating the use of excessively abstract and formal criteria for deciding which bootstrapping argument should be adopted in cases where the choice of theory is underdetermined by the evidence (Nordby, 1989). Nordby holds that there are some bootstrapping scenarios in which such criteria are inappropriate, because there is no overall guiding theory to determine either what should count as relevant evidence

or the relative importance of information from a variety of different sources (each with their own standards of evidence). In Chapter 2, subsections 2.5.2, 2.6, I argued that Nordby's argument should be extended to cover attempts to develop new theories of the mind/brain which incorporate the findings of neurophysiological data and psychology into philosophical argument. Here, I contend that Marr's excessive reliance upon the concept of information-processing in the analysis of the mind/brain demonstrates similar deficiencies to those created by Glymour's reliance upon formal criteria.

In Marr's view, previous studies of vision have failed to single out any one account of visual processes from among many candidates on the grounds that it alone provides a comprehensive and correct explanation of these processes. This failure is, he argues, directly attributable to their neglect of the information-processing approach. This approach helps us to unify the information provided by neurophysiological research and by studies in computation into a rigorous explanation (Marr, 1982a, p. 19). Indeed, Marr's view amounts to the claim that the use of a three-tiered Computational approach provides the guiding theory for resolving the difficulties created by the integration of evidence from a variety of different disciplines that are, I have argued, as prevalent in the philosophy of the

mind/brain as they are in detective work (see Chapter 2, subsection 2.5.2).

However, it is clear that Marr has radically overestimated the power of his Computational approach. It is highly unlikely that evolution will furnish creatures with an elegant solution to the information-processing tasks that they have to address, yet Marr's methodology is liable to over-emphasise abstract considerations that are more likely to suggest such elegant solutions than are the neurophysiological data that he relegates to a much later stage of theory construction. Indeed, Marr's criterion for a satisfactory account of a visual subprocess is that it must begin with an abstract description of the information-processing task that the process is carrying out. This means that he automatically rules out accounts that do not set out from the same premises as incapable of providing a satisfactory explanation. It is likely that Marr is falling into precisely the trap outlined in Giere (1983) of incorporating constraints into our criteria for a suitable account so that only those accounts that he already finds acceptable will fulfil these criteria (see also my discussion of Giere, Chapter 2, subsection 2.5.2). This is another, more insidious way of rendering one's theories systematically immune from refutation.

Like Glymour's formal criteria for theory selection, Marr's approach is doomed to failure because it cannot adequately integrate data from a variety of empirical sources with more abstract considerations. We must *decide* in the absence of any guidelines which sources are most important. Marr has made a choice, but this choice is so clearly at odds with the empirical evidence available, that it must be rejected<sup>15</sup>.



#### 6.1.4 Summary of the lessons to be learned from Marr's bootstrap methodology

I have argued that, although Marr does adopt a co-evolutionary approach to the study of the mind/brain, his use of a two-level bootstrap strategy to vindicate this approach fails because there is not enough evidence to support his hypotheses that the approach will succeed.

Much of the plausibility of Marr's theory is lost once it is seen that his use of the concept of information-processing is doubly marred (no pun intended!). Firstly, I showed in the previous subsection (part B) that the constraints upon the external world postulated by Marr are not refutable by empirical evidence, with the result that Marr, like Fodor and McGinn, fails to satisfy scientific realism's requirement for testable evidence of the explanatory success of their hypotheses.

Secondly, although Marr claims that his three-tiered approach to explanation is the only acceptable methodology for the study of the processes of the mind/brain, I have argued that this claim is unwarranted. In the first place, evolutionary considerations suggest that nature does not necessarily seek (and perhaps could not) the elegant solutions to information-processing problems that Marr's methodology is liable to yield. Moreover, in making an

information-processing task description the essential and most important part of any explanation, Marr has simply eliminated theories that may be constructed in the future<sup>16</sup> without reference to such a description from being considered as possible explanations, thus excluding a potential source of objections to his own view. This renders his claim to have isolated the only possible methodology for the study of the mind/brain less testable than it might otherwise have been. Marr thus fails to resolve the difficulties associated with the integration of data from a variety of empirical sources into a philosophical account.

Marr is therefore no more entitled to claim that he has provided us with a bootstrapping argument that is based upon well-tested evidence than is McGinn or Fodor, despite his overtly more scientific approach. Once more, we find that the use of the bootstrap strategy has not resulted in a satisfactory account of the mind/brain. Marr's theory is blighted, because it incorporates unwarranted assumptions which render him vulnerable to the same objections as the metaphysical realists discussed in Chapter 2.

I now turn to an examination of Churchland's ability to deal with the very same issues of realism and evidence.

## 6.2 Churchland's place in the debate

Patricia Smith Churchland is a very important figure in the recent history of the philosophy of mind. Her recent work in co-evolutionary theory has effectively controlled much of the post-1986 debate (following the publication of P.S. Churchland, 1986a).

There are three central aspects to my discussion of the work of Patricia Smith Churchland<sup>17</sup>. Firstly, it is important to show that, despite the fact that her interpretation of the need for the co-evolution of philosophy, psychology and neuroscience in formulating theories of the mind/brain is radically different from that of Marr (see below, subsection 6.2.1), this does not prevent her from falling into the traps associated with metaphysical realism. A dispassionate consideration of Churchland's interpretation of evidence obtained from a variety of different disciplines, including psychology and neuroscience (subsection 6.2.3, below), will indicate that her inferences about the nature of the mind/brain, and about the methodology that we should use to learn about it, are frequently unwarranted.

Secondly, Churchland's failings in this respect provide further confirmation for my argument that Nordby's account of the problems encountered when "bootstrapping while barefoot" (Nordby, 1989, p. 383) is readily extended to describe the difficulties involved in using the bootstrap

argument to develop co-evolutionary theories of mind/brain function. Here, as in detective work, there are no fundamental guidelines governing the relative importance of diverse sources of information, each of which have their own criteria for the evaluation of evidence. It will be demonstrated in the remainder of this chapter that although Churchland, like Marr, attempts to produce such guidelines, these guidelines derive much of their initial plausibility from Churchland's exclusion of those sources of evidence which cast doubt upon her interpretation of the data. She therefore fails to solve the problem of empirical underdetermination of theories by the evidence as it occurs within the study of the mind/brain.

Finally, Churchland's neglect of some sources of evidence that tell against her theory of the mind/brain indicates that her use of the bootstrap argument is not significantly better than that of Fodor, which she criticises so roundly (see Chapter 4, subsections 4.4.2, 4.4.3). In fact, it has been argued that the imbalance in Churchland's argument is partly caused by her overestimation of the scope of her (rightly) critical attack upon the Fodorean sentential paradigm.

In reality, there are some startling similarities between Churchland's work and that of Fodor. Both employ arguments which make unwarranted assumptions that have been

systematically shielded from potentially damaging criticisms, and both are therefore vulnerable to the charge that their views are more akin to the metaphysical realist than the scientific realist, despite their very vocal protestations to the contrary.

I shall approach Churchland's bootstrapping argument in the same way that I dealt with Fodor's Representational Theory of Mind, and Marr's theory of vision. A short summary of Churchland's eliminative materialist position (subsection 6.2.1) precedes a breakdown in subsection 6.2.2 of the bootstrap strategy that she employs. This will be followed by a criticism of the evidence that she employs at each stage of the argument (subsection 6.2.3)<sup>18</sup>. I shall then demonstrate in subsection 6.2.4 that Churchland's co-evolutionary theory of mind may be reconstructed in the form of a semantically inconsistent tetrad which resembles the one associated with the traditional veil-of-perception issue, before concluding the chapter with a review of the importance of the work of Marr and Churchland for my overall argument.

### 6.2.1 A brief outline of Churchland's co-evolutionary theory of the mind/brain

Like Marr, Churchland pronounces herself dissatisfied with the traditional isolation of the philosophy of mind, psychology and neuroscience. She believes that insular reflections on the part of the philosopher can only lead to conclusions that are completely divorced from reality. For example, she writes that

"[f]or philosophers, an understanding of what progress has been made in neuroscience is essential to sustain and constrain theories about such things as how representations relate to the world" (P.S. Churchland, 1986a, p. 3).

Conversely, the neuroscientist may spend many hours toiling at his laboratory bench in complete ignorance of the wider implications of the data that his experiments produce. Churchland comments that

"neuroscience needs philosophy because ongoing research must have a synoptic vision within which the immediate research goals make sense. Such a synoptic vision, transcending disciplinary boundaries but informed by the relevant disciplines, testing the integrity of the

governing paradigm and investigating alternatives, is philosophy" (ibid, p. 482).

Thus far, it is unlikely that Marr would find anything disagreeable in Churchland's remarks, but it is clear that Churchland is prepared to accord a much greater significance to neuroscientific data in the actual *shaping* of a theory of the mind/brain. Recall that Marr relegates the examination of neuronal function to the final stages of his three-tiered top-down explanation of processes within the mind/brain. Churchland makes her opposing position clear in this passage:

"We want to know how we *actually* see, plan, learn, and so forth. If we commit ourselves to purely top-down approaches, we deny ourselves important constraints that would help narrow the search space. Ditto, of course, for purely bottom-up approaches" (P.S. Churchland, 1988, p. 394).

Churchland disagrees with the distinction made by top-down theories of the mind/brain between structure and function. Whereas the top-down theorist contends that

"at best, neuroscience can provide only a *structural* theory, as opposed to the *functional*

theory sought; at most, it will give us the engineering minutiae as opposed to the design configurations we want" (P.S. Churchland, 1980b, p. 186),

Churchland argues that no such rigid divisions can be postulated since, for example,

"[i]t is to the neural structure of the cerebellar cortex that we must look in order to determine more exactly what the cerebellum does and how it does what it does" (ibid, p. 205).

Moreover, Churchland believes that it is good practice within the philosophy of science (and, indeed, within science itself) to aim for unified explanations of phenomena. This often results in the reduction of a very specific explanatory theoretical framework to another, more basic theory, so that the basic theory explains everything that was previously accounted for by the specific theory, and succeeds in explaining phenomena that this specific theory could not deal with, whilst also accounting for the failure of the specific theory.

This may sometimes result in the simplification not only of explanations, but also of the ontologies that are thought to exist. It may be that a more basic theory will indicate



that the structures and entities postulated by the specific theory to account for a given phenomenon do not exist. Churchland cites the reduction of the laws of optics to the laws of electromagnetic theory as an example of such ontological simplification. Light simply is electromagnetic radiation, but until the end of the nineteenth century it had been thought that they were two different things (P.S. Churchland, 1986a, p. 280).

In more radical cases, unity of explanation is achieved not by the reduction of one theory to another, but by the total elimination of the more specific theory, and the ontologies that it postulates.

Churchland's co-evolutionary approach to understanding the mind/brain results in her adherence to a position known as eliminative materialism. The work of several key figures from the history of philosophy has had an influence upon the development of this view. For example, Churchland comments that it was Kant who first realised that Hume's remarks about the problems inherent in the traditional veil-of-perception issue could be extended to cover our knowledge of the mind/brain, so that

"there is nothing epistemologically unique or sacrosanct about introspectively based beliefs; they are not on an entirely different footing

from beliefs about the outer world, and they have neither more nor less need for justification" (P.S. Churchland, 1986a, p. 248).

Moreover, Kant's followers developed a second Humean insight, namely, that the mind must be studied by empirical means (ibid, pp. 247-8).

Ultimately, however, it was Feyerabend who suggested that Kant's negative hypothesis that our knowledge of our own minds is not direct, but mediated by our very concepts of mental and processes could be developed into the positive thesis that these concepts might themselves be revised in the light of scientific, empirical investigations of the mind/brain (ibid, p. 275). Thus two of Hume's central tenets fused into a naturalistic basis for eliminative materialism.

Briefly stated, eliminative materialism is the view that, since folk psychological theory<sup>19</sup> provides us with a very inadequate explanation of our mental life, it should ultimately be revised into or even replaced by a more scientific psychology. The categories of this scientific psychology are envisaged to reduce smoothly to mirror those divisions that neuroscientific research has isolated within the brain (see P.S. Churchland, 1986a, p.396). Many objections to eliminative materialism concentrate their

attack upon the impossibility of achieving neat correlations between the functional categories of psychology and the structures within the brain. Churchland counters that

"from the reductionist viewpoint, this possibility does not look like an obstacle to reduction so much as it predicts a fragmentation and reconfiguration of the psychological categories" (ibid, p. 365).

Churchland uses examples taken from neuroscientific research to support her view that our psychological theories should be developed in conjunction with results from the laboratory. I include a brief outline of one such example below, with the sole aim of indicating the way in which Churchland thinks that it casts doubt upon the established terminology of folk psychology and top-down theories of the mind/brain.

Churchland discusses the tensor network theory of Pellionisz and Llinás<sup>20</sup> and its implications for her wider thesis of eliminative materialism (ibid, pp. 412-458). This is an attempt to explain the mechanics of sensorimotor control based on a computer model of the structure of the cerebellum of a frog. Eye-limb co-ordination is then understood in terms of a vector-to-vector transformation

between the co-ordinate system of neurons providing the visual input (the visual phase space) and the co-ordinate system of the neurons that occasion motor output (the motor phase space). Each phase space is represented in the brain in the form of a grid, such that when the vector transformation of the co-ordinates in the visual phase space is complete, a deformed map or grid of this phase space is superimposed upon the grid of the motor phase space. The grids are linked by short vertical fibres. When a certain point within the sensory phase space is stimulated, the corresponding point in the motor phase space (according to the vector transformation) is also stimulated and sends a signal to the motor fibres which control limb movement. Churchland argues that the theory is consistent with the structures to be found in the central nervous system, since

"many structures abide by a principle of topographic mapping, whereby neighborhood relations of cells at one periphery are preserved in the arrangement of cells at other locations in the projection system. If we think of the neurons at the sensory periphery as forming a receptor sheet, then deformed versions of that sheet are represented in a large number of CNS regions" (ibid, pp. 119-20).

Churchland proceeds to argue that the tensor network theory suggests that our cognitive abilities may be based upon similar transformations between phase spaces, which themselves consist in representations that are definitely not sentential. Neuroscience, therefore, provides a way of countering Fodor's claim that representations must be sentential if computational processes exist within the mind/brain (ibid, pp. 451-2).<sup>21</sup>

I now turn to an examination of Churchland's attitude towards bootstrapping and an analysis of her use of the strategy.

#### **6.2.2 Churchland and the bootstrap method**

##### **A - Churchland's awareness of the difficulties of the bootstrap method**

I noted at the start of my examination of Churchland's criticisms of the Fodorean language of thought hypothesis in Chapter 4 (subsection 4.4.1) that Churchland is not opposed to the use of the bootstrap strategy per se. Her support for this method of constructing a theory of the mind/brain is equally apparent in her arguments for eliminative materialism. For example, she comments that

"what may happen is that neuroscience, initially making use of commonsense concepts, will make discoveries that transmogrify them. Neuroscience is able to bootstrap as any other science" (P.S. Churchland, 1980b, p. 193).

Moreover, Churchland refers to the use of the bootstrap method in terminology reminiscent of the traditional veil-of-perception issue. Folk psychological descriptions of our mental states and experiences may be caused by the structures within our mind/brains, but they are not an accurate reflection of the nature of these structures. We may conclude that Churchland shares Feyerabend's view that, just as we question the way that the external world appears to us, we should also question the way that our mind/brain appears to operate, lest we are equally misled here by appearances. Churchland remarks that this line of questioning may result in a situation in which

"[p]rimitive theories give way to sophisticated theories, and as the latter become the common coin of everyday life, they may then acquire the status of common sense. A remarkable thing about the human brain is that it can use those primitive theories to bootstrap its way to ever more comprehensive and powerful theories - to

find the reality behind the appearances" (P.S. Churchland, 1986a, p. 265).

Further, there are favourable indications that Churchland understands the particular difficulties of bootstrapping within a new area of scientific activity, into which the results of many different disciplines are incorporated (see Chapter 2, subsection 2.5.2).

For example, Churchland makes it abundantly clear that she is aware of the lack of an overall theory governing the co-evolutionary development of psychology and neuroscience. How can our theories ever get off the ground unless we know what questions to ask, and how can we know what questions to ask in the absence of any theory? Indeed, Churchland notes that

"[i]t may be that commonsense theory (folk psychology) is so misconceived, and its taxonomy so askew, that even the formulation of our questions thwarts our inquiry ...

....The difficulties here are such as to make one fear that we cannot get a theory until we have one - that acquisition of theory, like acquisition of wealth or breeding stock, is limited to those who already have it...." (ibid, pp. 152-3).

However, Churchland concludes that the situation here is no worse than in any other branch of science, and declares that

"any successful science got to where it did by heroic and *stubborn bootstrapping*" (ibid, p. 153, my italics)<sup>22</sup>.

Nonetheless, a cursory glance at P.S. Churchland's 1986a would suggest that we have every cause to be optimistic that she is capable of avoiding the problems that we saw afflict the bootstrapping arguments of McGinn, Fodor, and most recently, Marr. In other words, it appears that Churchland is well aware of the need for caution when evincing conclusions based upon a consideration of a variety of disciplines, each with their own sources of evidence, and even with their own preconceptions concerning the correct usage of key terms, such as "representation" and "information".

An example of this kind of preconception may be found in Hacker (1987). Hacker argues that attempts to construct theories of the brain using concepts such as "representation" are doomed to failure. The resulting theories will be meaningless, since they are



"born of the inadvertent misuse of language, misunderstandings of crucial concepts such as *language, code, representation, understanding,* and equivocation over, for example, *information, sign, map*" (Hacker, 1987, p. 491).

Churchland is scathing of such "category error" criticisms, believing that they serve only to restrict the development of psychological theory (P.S.Churchland, 1986a, p. 273). These criticisms are a legacy of conceptual analysis, and should be rejected on the grounds that

"one person's category error is another person's deep theory about the nature of the universe, and what is deemed appropriate or inappropriate in the application of categories depends tremendously on one's empirical beliefs and one's theoretical imagination" (P.S. Churchland, 1986a, p. 273).

Those philosophers who persist in retaining our traditional use of all concepts in spite of any empirical suggestions that they may be appropriately extended to new domains are simply guilty of rendering their theories immune to empirical refutation. Such rearguard action can only result in lack of testability, and hence in the charge that these theories are little better than those postulated by

the traditional metaphysical realist (see Chapter 2 subsection 2.2.1). A co-evolutionary methodology is able to avoid these difficulties, Churchland maintains, since one of its fundamental tenets is that

"psychology and neuroscience should each be vulnerable to disconfirmation and revision at any level by the discoveries of the other ..." (ibid, p. 376).

Neglect of this point can, in her view, only lead to disaster. She continues

"[t]he isolation of psychology from the disconfirmatory reach of neuroscience would be a mistake, because in general it is such susceptibility that keeps a science honest" (ibid, p. 376).

Still more encouraging is Churchland's emphasis upon the problems endemic in the use of data from a branch of study that is still in its infancy. She stresses the dangers of making unwarranted inferences from a small amount of information, the risks of obtaining conflicting results from a variety of methodologies and, most importantly, warns against the temptation to cite a hypothesis out of

context, as if it were a proven truth (P.S. Churchland, 1986a, pp. 147-8).

However, one of the major messages of this section of my thesis is that an apparent awareness of the constraints upon the scientific realist with regard to the need for testable evidence from a variety of sources is frequently belied by the nature of the assumptions that are actually made within the bootstrapping arguments of the philosophers under scrutiny. In subsection 6.2.3, I shall demonstrate that this dichotomy between an ostensible grasp of the need for satisfactory evidence upon which to base inferences and an actual neglect of this very same issue is also present within Churchland's co-evolutionary theory of the mind/brain. Specifically, she demonstrates a failure to expose her general hypothesis concerning the inadequacy of folk psychology to the possibility of refutation, thereby rendering it systematically untestable.

Churchland's use of the bootstrap strategy will be outlined in the second half of this subsection.

**B - The bootstrap in action - Churchland's co-evolutionary theory of the mind/brain**

Churchland, like the other theorists whose arguments I have examined, sees the potential of the bootstrap argument for

arguing from a general hypothesis to a more specific one, which will (when combined with additional evidence in its favour) eventually yield further confirmation of the general hypothesis. Her two major bodies of evidence are intended to disabuse opponents of two key misconceptions about the value of neuroscience for psychology. Churchland remarks that

"[t]he complaint that neuroscience cannot in principle do justice to the generalizations of psychology errs in two directions: it is overconfident about the integrity of folk psychology and it underestimates the value of co-evolution of theories" (ibid, p. 385).

Within her theory of eliminative materialism, which I shall call EM, Churchland postulates a general hypotheses,  $H_1$ , which may be summarized as follows:

$H_1$  We are accustomed to categorizing and discussing our mental states and experiences using the terminology and rudimentary generalizations of folk psychology. However, the concepts of folk psychology are incapable of describing or explaining many of these states and experiences, and so the theory should be regarded only as a temporary starting-point for the study of the mind/brain. It is ripe for

radical revision, and will perhaps be eliminated completely in time.<sup>23</sup> To this extent she agrees with Ryle (1949).

Some of Churchland's sources of evidence in support of this general hypothesis have been mentioned briefly in the first half of this discussion. However, it will be instructive to list all the sources here, so that they can be readily identified in the course of the critical examination of the bootstrap argument in subsection 6.2.3 below. These sources, which I shall refer to collectively as  $E_1$ , are as follows:

a) data accumulated from the study of brain-damaged subjects and those suffering from mental illness indicate that the categories of folk psychology do not correspond to the underlying mechanisms within the brain. For example, Churchland comments that

"[t]he brain undoubtedly has a number of mechanisms for monitoring brain processes, and the folk psychological categories of 'awareness' and 'consciousness' indifferently lump together an assortment of the mechanisms" (ibid, p. 321).<sup>24 25</sup>

b) folk psychology is a theory, and is therefore subject to correction or even elimination. Churchland declares that

"if we see that folk psychology has no right to epistemological privilege, and no immunity to revision and correction, then we can begin to see that its generalizations and categories can be corrected and improved upon" (ibid, p. 311)

c) folk psychology makes use of the sentential paradigm to characterize the representations and processes of the mind/brain. There are strong arguments against the plausibility of such a paradigm, including evolutionary considerations which suggest that it is unlikely that all representations within the mind/brain are sentential (see Chapter 4, subsections 4.4.2, 4.4.3 for details).

d) an argument by analogy with other folk theories, such as folk physics or chemistry. Churchland comments that a discussion of the fate of folk physics yields

"several points that will be useful in the discussion of reducing mental states to brain states: first, that in reductive developments one theory can displace and falsify another; second, that sometimes what is displaced and falsified is a *folk theory* within which those who hold it make their observations; third, that despite the self-evidence of the folk theory, it can be demonstrated to be misconceived; and fourth, that

as a newly acquired theory becomes familiar, it can be as routinely and casually used as the old folk theory" (ibid, p. 291).

These are the main sources of evidence cited by Churchland in support of the possibility that folk psychology may be revised or eliminated<sup>26</sup>. They are used, in conjunction with the general hypothesis itself ( $H_1$ ), to derive her more specific hypothesis ( $H_2$ ), which comprises the methodological assertion that we should seek to replace folk psychology with a more scientific psychology. It is envisaged that the categories postulated by this scientific psychology will correlate neatly with those uncovered in the mind/brain by neuroscientific research. Churchland pronounces that

"[a]s neurobiology and neuropsychology probe the mechanics and functions of the brain, a reconfiguring of categories can be predicted" (ibid, p. 321-2).

Like Marr, Churchland is perfectly prepared to countenance the use of the bootstrap strategy within individual co-evolutionary research projects. For example, she notes with regard to the tensor network theory of Pellionisz and Llinás (see above, subsection 6.2.1), that

"[a]s more is discovered about the neuronal basis, the basic hypothesis may be corrected and elaborated, and thus theory and experimental research co-evolve" (ibid, p. 437).

As we might expect, Churchland derives her final body of evidence,  $E_2$ , from the success of specific attempts to develop co-evolutionary accounts of processes and representations within the mind/brain. In her opinion, such accounts provide us with further support for the view that the structural divisions within the mind/brain correlate with functional divisions that are not well represented by the concepts of folk psychology. Consequently, she believes they yield further confirmation of Churchland's general hypothesis ( $H_1$ ) about the poverty of folk psychology as an explanatory theory.<sup>27</sup>

I shall now move on to a critical examination of the plausibility of Churchland's argument.

### **6.2.3 Criticisms of Churchland's eliminative materialism**

This subsection will be divided into two parts. The first part will discuss the validity of Churchland's initial general hypothesis ( $H_1$ ) concerning the nature of, and outlook for, folk psychology, and the strength of the body of evidence ( $E_1$ ) that she cites in support of this theory.



In the light of these criticisms of  $H_1$  and  $E_1$ , I shall then argue that Churchland's inference to  $H_2$ , the need for a co-evolutionary approach to the mind/brain, is unconvincing.

The central claim of the second part of this subsection will be that the data obtained from the co-evolutionary studies of the mind/brain cited by Churchland do not constitute sufficient evidence that this approach is the only one likely to yield success. In fact, they derive much of their plausibility from a prior acceptance of hypothesis  $H_1$  and its supporting body of evidence, and therefore cannot provide further confirmation of  $H_1$  in a non-circular fashion.

**A - Folk Psychology and foul play - the first half of Churchland's bootstrap argument**

We saw in the previous subsection that Churchland relies upon a variety of arguments to support her general hypothesis  $H_1$  that folk psychology yields an inaccurate view of the mind/brain and therefore should not be allowed to remain as the backbone of explanations of our mental states, processes and experiences. It seems as if Churchland has understood the need for a bootstrap argument to employ evidence from a variety of sources to avoid accusations of circularity (at least thus far).

However, I shall argue that this part of her bootstrap strategy is flawed on three counts. Firstly, many of the points that Churchland cites in support of her hypothesis about folk psychology are simply wrong. Secondly, although her remaining observations about the nature of folk psychology, and about the difficulties inherent in the sentential paradigm, are correct, they do not entitle her to the radical view that folk psychology must be radically revised, if not eliminated. Thirdly, and perhaps most damningly, Churchland renders her hypothesis immune to falsification by systematically avoiding discussion of some sources of evidence that might tell against her hypothesis. She is therefore guilty, like McGinn, Fodor and Marr, of basing her theory upon unwarranted and untestable (hence metaphysical) assumptions.

I shall divide my criticisms of Churchland's argument into two groups, each of which deals with one of the first two fundamental errors identified in the preceding paragraph. These two types of error overlap in some cases with the third flaw because the deliberate and systematic exclusion of some evidence has resulted in either faulty assertions or unwarranted inferences from acceptable premises. I shall point out these cases where appropriate.

I begin with the evidence that is just plain wrong. In fact, many of Churchland's critics have argued that she is

attacking a straw man, because she has misunderstood the true nature of folk psychology and is therefore either expecting it to explain phenomena that it was never supposed to explain [for example, claim (a) in my summary, above], or attributing to it qualities that it does not possess [claims (b) and (c)]. One key objection is made by Wilkes (1984, 1986). Wilkes argues firstly, that folk psychology is not a theory, since it does not consist of universal laws like those of science, but rather of generalizations that will only hold true in given circumstances, and with many caveats. Moreover, she continues, folk psychology is simply not intended to explain phenomena that are caused by exceptional circumstances such as brain damage, or mental illness. Both of these objections are encapsulated in the following passage. Wilkes comments that

"Churchland stresses the failure of common-sense psychology to provide terms that are suitable as parts of scientific explanations. What she does not emphasize enough is that common sense also has wildly different explananda most of the time" (Wilkes, 1986, p. 170).

These two points need not necessarily be linked. I think that Churchland may well be entitled to the view that folk psychology is a theory [claim (b)], given that the laws of

science may themselves be very complex and may require very specific conditions for them to hold. However, even if Churchland is right to call folk psychology a theory, this does not obviate the objection that she is not entitled to claim that it is inadequate simply because it cannot account for every single mental state imaginable [claim (a)]. Thus Horgan and Woodward declare that

"Churchland's argument seems to impose the a *priori* demand that any successful psychological theory account for a certain pre-established range of phenomena, and do so in a unified way. Arguments of this general type deserve to be treated with scepticism and caution" (Horgan and Woodward, 1985, p. 200)<sup>28</sup>.

I think that Horgan and Woodward are correct to criticise this particular source of "evidence" for the poverty of folk psychology as an explanatory force. Moreover, Churchland's use of an a *priori* constraint upon the variety of phenomena that a satisfactory theory of the mind/brain must account for results in a measure of immunity for her hypothesis against falsification. If a satisfactory explanation of the mind/brain must, *ex hypothesi*, account for a number of specified phenomena, then any potential rival that does not do this is automatically excluded from consideration, thereby reducing the number of plausible

alternatives to eliminative materialism. Here, we see that one source of evidence [(a) above] for the poverty of folk psychology rests upon an unwarranted assumption that may be described as metaphysical, because it has been rendered untestable.

There is a great deal of evidence that Churchland regards the view that all representation within the mind/brain is sentential as an integral part of folk psychology. For example, she comments that

"there are substantial reasons for predicting that at best inference and sentence-like representations will have a small role in the theory of information-processing, and for predicting quite radical revisions in folk psychology" (P.S. Churchland, 1986a, p. 386).

However, several critics have argued persuasively that it is perfectly possible to endorse a psychology that makes reference to the intentional states of folk psychology without being committed to the view that all representation is sentential. For example, Kitcher claims that

"[i]n associating cognitive psychology with sententialism, Churchland appears to have gone

seriously astray.....she has uncritically accepted Fodor's claim that propositional attitudes are the constructs in terms of which cognitive theory elaborates its typical elaborations" (Kitcher, 1984, p. 85).

In other words, Churchland's penetrating criticisms of Fodor's own representational theory of the mind are acceptable, but they do not entitle her to the further claim that all theories that postulate the intentional concepts used in folk psychology are tarred with the same Fodorean brush. They are not. Thus, claim (c) in support of Churchland's general hypothesis  $H_1$  is quite simply wrong.

It was also noted in subsection 6.2.2 that Churchland uses an argument from analogy to show that folk psychology is an unnecessary and unhelpful theory [claim d]. Comparisons are made with folk physics which has now been rejected in favour of more scientific theories. Churchland maintains that discussion of such examples from other branches of science can only be advantageous for an area of study that is still in its infancy. She asserts that

"by surveying dispassionately sciences with long histories, mature theories, and a rich theoretical evolution, it is to be hoped that analogies and disanalogies can be discerned that

will be instructive in confronting the issues at hand" (P.S. Churchland, 1986a, p. 8).

However, even a cursory survey of the available literature yields the conclusion that Churchland has overstressed the similarities of folk physics with folk psychology and ignored the differences. This lends her hypothesis about the negative value of folk psychology greater credibility than it deserves. Stoljar, for example, argues that the argument from analogy fails because it is important for us to examine the merits of folk psychology as a theory in its own right, rather than simply assuming that it will suffer the same fate as folk physics (Stoljar, 1988)<sup>29</sup>. Indeed, it is clear in the light of the arguments already put forward against the remainder of Churchland's evidence in favour of her initial hypothesis that, not only does the argument by analogy beg the question, it also does folk psychology a grave disservice (assuming Churchland to have correctly summarized the history of folk physics).

So far, we have seen grounds for doubting three out of four of Churchland's sources of evidence in favour of hypothesis  $H_1$ . Only claim (b), that folk psychology is a theory, and is liable to revision in the light of empirical evidence, has not yet been countered by any very telling objections. However, even this source of support for Churchland's general hypothesis is tenuous, and I shall now present

arguments which demonstrate that Churchland's radical conclusions are not justified.

The first point borders upon the trivial, but is nevertheless worthy of a brief mention. Even if Churchland is correct to claim that folk psychology, as a theory, should be open to correction or elimination, this does not mean that the theory is automatically in need of revision. Churchland is, in principle at least, aware of this point, hence her remark that

"I have not argued that [displacement] will be the fate of folk psychology, but only that once we have seen that folk psychology has no more epistemological privilege than folk physics,... we can rid ourselves of a huge weight of argument that tempts us by appeal to the obviousness and the certain inviolability of folk psychology" (P.S. Churchland, 1986a, p. 312).

As a result, Churchland can only justify the replacement of folk psychology if there is empirical evidence that it is dissatisfactory. So far, as we have seen, Churchland has provided no solid evidence that this is actually the case.

Moreover, even if the categories of folk psychology cannot be reduced smoothly to those thought to exist within the



mind/brain, Churchland is not entitled to move from this general hypothesis ( $H_1$ ) to the bald assertion of her more specific hypothesis  $H_2$ , that the best methodology for understanding the mind/brain is a co-evolutionary one. There are, as her critics point out, other possibilities that Churchland has simply overlooked, even though the evidence that she has cited so far in favour of her position radically underdetermines the choice between the various alternatives. This is the message put forward by McDonough:

"The assumption that the replacement could be something like neurobiology begs the question about the *limits of the revisability of our self-conception*" (McDonough, 1991, p. 273).

There are a variety of other hypotheses that follow equally from the situation as Churchland has so far described it (allowing her for the moment that there is at least some support for her negative views upon folk psychology). For example, the failure of folk psychology to reduce to neuroscience has also been remarked upon by Donald Davidson (1980). Unlike Churchland, however, Davidson does not hold that we must revise folk psychology in order to accept that mental events are caused by and ultimately identical with events and processes within the brain. Instead, he adopts a position known as anomalous monism, which is

characterized by the view that there can be no laws relating the description of these states in the concepts of the mental and their description as physical events. This is because laws do not hold between events themselves (unlike causal connections) but between events under certain descriptions. For Davidson, mental events form a holistic, open-ended system which is incapable of sustaining nomological connections, and which cannot therefore be reduced in any lawlike manner to a nomological, closed system like that of the physical world. Davidson, like Churchland, is a physicalist, but he accepts neither the need for or the possibility of a smooth reduction of the concepts of the mental to those of the physical. Horgan and Woodward comment that P.M. Churchland fails to discuss the Davidsonian view<sup>30</sup>, and this balance is not redressed in the work of P.S. Churchland<sup>31</sup>. Horgan and Woodward's conclusion applies with equal force to her views. She is

"just mistaken to assume that [folk psychology] must be reducible to neuroscience in order to be compatible with it" (Horgan and Woodward, 1985, p. 204).

Horgan and Woodward have suggested that even if folk psychology cannot account for the totality of our mental experiences, it may still be possible to develop more

satisfactory explanations without resorting to an ultimate reduction of psychology to the categories of neuroscience. They argue that

"cognitive psychologists have developed extensive and detailed theories ... that employ concepts recognizably like the folk- psychological concepts of belief, desire, judgment, etc" (ibid, p. 200).

Similarly, von Eckardt criticizes Churchland for failing to realize that cognitive psychology is equally able to bootstrap from the starting-point of folk psychology (von Eckardt, 1984, p. 68).

Churchland has simply failed to provide us with enough evidence that the basic concepts of folk psychology are so inherently flawed that only a reduction of psychology to neuroscience can provide a satisfactory explanation of the processes and representational capacities of the mind/brain. Greenwood criticises the Churchlands' enterprise as follows:

"the Churchlands [do not] attempt ... to document in detail the supposedly de facto inaccuracies of everyday folk-psychological explanations, nor do they, for example, engage in an extended

methodological critique of the evidential basis of generally accepted theories in cognitive, social, and clinical psychology that make essential reference to intentional psychological states" (Greenwood, 1991a, pp. 15-16),

As a result, Churchland's inference from the limitations of folk psychology to the specific hypothesis that we must endorse a co-evolutionary methodology for the study of the mind/brain is unwarranted. She has, once more, systematically ignored the possibility of rival interpretations of the empirical data, thereby securing the relative immunity of her argument from refutation. This is, as we have already noted, an unacceptable move, since it results in a theory which is grounded upon untestable assumptions<sup>32</sup>, and in the charge that Churchland is no more capable of dealing with the constraints placed upon evidence by scientific realism than McGinn, Fodor or Marr.

It is now time to assess the evidence provided by Churchland in further support of her more specific hypothesis  $H_2$  that the co-evolutionary approach to the mind/brain is the right one to use. This evidence ( $E_2$ ) consists in the data obtained from the use of this approach so far. Churchland believes that these data indicate that the approach is a fruitful one and that they provide

further confirmation of the poverty of folk psychology's explanatory apparatus.

**B - Neurophysiology and new faits accomplis - a critique of Churchland's co-evolutionary data**

I shall not be concerned in this discussion with the precise details of the co-evolutionary studies that Churchland uses to advocate the concentration by researchers upon such a methodology. It is sufficient to demonstrate that the results of these studies do not entitle her to this conclusion, nor do they provide any further substantiation of her view that folk psychology is likely to be revised or eliminated.

Many of the points to be made in this part of my argument are really just the natural corollaries of the criticisms that I have just made of the first part of Churchland's bootstrap argument. Firstly, Churchland holds that a more neuroscientific approach to the mind/brain will result in the explanation of mental processes and phenomena that folk psychology is not capable of accounting for. Yet, given the view put forward in the previous subsection that folk psychology was never intended to explain every single mental state, experience, or process, including those that result from rare occurrences like brain damage or mental illness, the ability of the co-evolutionary approach to

provide such explanations cannot be taken as evidence that folk psychology must be superseded by a more neuroscientifically orientated view.

Secondly, and more generally, the success of neuroscience in providing explanations of some of our mental phenomena does not entail that it should be allowed to dominate the investigation of the mind/brain. We could allow for two different levels of explanation, a psychological theory which makes use of intentional concepts, and a neuroscientific account which concentrates upon providing explanations in terms of the structures to be found within the brain. The success of some neuroscientific work does not entitle Churchland to infer that all other methodologies fail. Further, no more confirmation is provided of her general hypothesis that folk psychology and its concepts are doomed to failure.

It has been suggested that Churchland misinterprets the results of some neurological studies so that they give the appearance of supporting her argument more than they actually do. For example, Churchland is keen to refute the notion that all apparently intentional states require the existence of mental representations (which take a sentential form) so that the requisite information processing can take place (P.S. Churchland, 1980b, pp. 194-7). Consequently, she cites studies of the sea slug which

are intended to demonstrate that habituation to a stimulus is possible even in the absence of representations, because it can be accounted for in purely neurophysiological terms (ibid, p. 200). Similar neuronal accounts are given of the ability of monkeys to direct their attention towards a given object. Eckardt argues that both examples fail to support Churchland's view. In the case of the sea slug, she notes that it is not clear that habituation is an intentional phenomenon in the first place, and she comments that Churchland may not have succeeded in providing a non-representational account of visual attention, since the monkeys may still be using some kind of representational system (von Eckardt, 1984, p. 88). This means that unless Churchland is prepared to stipulate *a priori* that some uses of the term "representation" are acceptable whilst others are not (thereby rendering her view untestable), she is not entitled to the inference that these examples provide unequivocal support for her theory.

Similarly, Churchland suggests that the vector transformation theory put forward by Pellionisz and Llinás provides us with an alternative view of representation from that encapsulated in Fodor's sentential paradigm. However, since, as we saw in the previous part of this subsection, it is incorrect to assume that an irreducible psychology must incorporate sentential representations, the vector transformation theory does not provide us with any clear

grounds for pursuing the co-evolutionary approach and no other. Once more, it is equally clear that good neuroscientific data fail to lend further confirmation to the general hypothesis that folk psychology is inadequate. Hence, McDonough comments of the Churchlands that

"when they say that neuroscience may lead us to eliminate beliefs, what they really mean is that it may lead us to eliminate internal states that picture a sentence. *Beliefs* have already been eliminated on metaphysical grounds before the empirically motivated revision is even considered" (McDonough, 1991, p. 279).

Indeed, McDonough has succeeded in identifying a fundamental difficulty within the second half of Churchland's bootstrap argument. I have traced a distinctive pattern in her use of neuroscientific studies to support her strategy. Churchland cites evidence in support of a co-evolutionary methodology, which she hopes will also provide further confirmation of her general hypothesis that folk psychology is inadequate, yet this evidence cannot provide unequivocal support for either of her hypotheses unless her arguments against folk psychology have already been accepted. In other words, the body of evidence ( $E_2$ ) rests tacitly upon acceptance of evidence  $E_1$ , which has already been discredited. Churchland's



apparently empirical argument is thereby exposed as both circular, and as based upon unwarranted, untestable and metaphysical assumptions. She falls readily into the trap of bootstrapping identified by Achinstein that her evidence is only supportive of the hypotheses within her theory, if we assume the truth of the theory to begin with (see also Chapter 2).

These comments will be developed in the following subsections.

#### 6.2.4 Churchland and the veil-of-perception issue

Churchland, like the other theorists that I have discussed, presents an ostensibly impeccable understanding of scientific realism, but is then incapable of transferring this understanding into a satisfactory bootstrapping argument. Indeed, we can put Churchland's claims into the format of a semantically inconsistent tetrad. Churchland takes the following four statements to be true:

I there are natural neurological categories within the mind/brain that can be embodied in a scientific psychology using a one-to-one mapping of concepts

II these neurologically isolable categories give rise to our overt mental experiences, and will ultimately provide us with the only acceptable explanation of these experiences

III we are not directly aware of these categories, as defined by neuroscience and scientific psychology: their effects are generally interpreted in the light of our (largely inadequate) folk psychological theory of mind/brain processes

IV nonetheless, we can be sure that scientific psychology and neuroscience do postulate the right kinds of categories, thereby providing the only satisfactory means of understanding the mind/brain, both because of the systematic breakdown of folk psychology which demonstrates that a more neurological account is appropriate, and because of experimental evidence.

These statements are related to Churchland's bootstrap argument in the following manner. Statements I and II comprise the essence of hypothesis  $H_2$ , that a scientific psychology which will reduce smoothly to neurologically defined categories is achievable. Statement III is effectively a combination of hypothesis  $H_1$  and the body of

evidence  $E_1$  - that is, it embodies the initial premise that we are not usually aware of neurological categories, because we interpret our mental experiences in the light of a folk psychological theory, together with an implicit summary of  $E_1$ , the evidence that this theory is inadequate. Finally, statement IV amounts to an attempt to support hypothesis  $H_2$  on the basis of the evidence contained in  $E_1$  and  $E_2$  that folk psychology is inadequate and that neuroscience has provided some useful clues to the nature of the mind/brain.<sup>33</sup>

It is now easy to see why the tetrad is semantically inconsistent. Given that we are only ever aware of our folk psychological views of the mind/brain, how can we ever assert that the categories postulated by neuroscience and scientific psychology will provide us with an alternative explanation? This is the strong or ontological version of the veil-of-perception issue once more. The body of evidence  $E_1$ , upon examination, provides us with no firm grounds for doubting the adequacy of folk psychology. I think it is unlikely, however, that anyone would want to argue that Churchland has provided us with no reason at all for examining the possibility of a second, more co-evolutionary approach to the mind/brain problem<sup>34</sup>.

However, this does not mean that Churchland is out of the woods. It is clear that she falls foul of the weaker,

epistemological version of the veil-of-perception issue. Even if we grant her the possibility of a second, more basic level of explanation, she has failed to provide us with the evidence we require to be sure that it has a unique explanatory status or that it is capable of performing the tasks that she claims it can. Statement IV does not support statements I and II, because the evidence cited neither demonstrates that folk psychology is incapable of playing some explanatory role, nor that any potential replacement must incorporate neuroscience. This is a clear case of empirical underdetermination that Churchland has failed to address. The initial plausibility of her claim that a co-evolutionary approach is to be preferred is chiefly attributable to her systematic neglect of evidence that casts doubt upon the possibility that folk psychology is fundamentally flawed and should be eliminated. Once Churchland's initial evidence is exposed as resting upon such unwarranted metaphysical assumptions, her argument is greatly weakened. The co-evolutionary approach therefore loses its much vaunted superior explanatory status, thereby depriving eliminative materialism of its chief motivating factor. The problem is exacerbated by the fact that, as we have seen, her second body of evidence can only support her co-evolutionary approach given a prior acceptance of the paucity of folk psychology.

Churchland's bootstrap argument is essentially circular in precisely the manner declared unacceptable within scientific realism by Glymour (see Chapter 2, subsection 2.3.1). Yet Churchland claims to understand the need for constraints upon evidence. How has this happened?

#### **6.2.5 Analysis of Churchland's failure**

Churchland presents a mixture of poor conceptual arguments and unwarranted inferences from empirical evidence in support of her bootstrapping strategy. These flaws in an superficially well-structured argument demonstrate that the use of the bootstrap argument cannot simply prevent empirical underdetermination by the adoption of abstract criteria such as those suggested by Glymour and P.M. Churchland (see Chapter 2, subsection 2.3.1). Patricia Smith Churchland has presented us with an argument which attempts to provide a unified explanatory system and claims to provide a variety of evidence in support of her hypotheses. Both of these strategies would be applauded by P.M. Churchland and Glymour. However, Patricia Smith Churchland has failed to appreciate the rigours of bootstrapping in an area of science which must take account of data from radically different disciplines. We saw in Chapter 2 that Nordby's description of the difficulties for the detective who must decide what facts constitute relevant evidence for the crime that he is investigating is

equally applicable in this kind of scientific scenario. The decisions as to what evidence is relevant, and what is not, and the weight that should be accorded to evidence from a variety of disciplines are not abstract, formal decisions, but ones which are frequently influenced by the preferred view of the theorist.

It seems clear that this is what has occurred within Churchland's bootstrap argument. Churchland's use of neuroscientific data appears acceptable if we assume what she sets out to prove, but this is clearly not a scientific decision. Churchland is predisposed to regard eliminative materialism as the only satisfactory theory of the mind/brain, and this results in her imposing very strict criteria for the success of folk psychology (and ignoring evidence that folk psychology might have a role to play in understanding the mind/brain), whilst employing very lax criteria for the validity of the more neuroscientific approach, so that any data are interpreted as supporting evidence for her own view (see also my discussions of Giere in Chapter 2, subsection 2.6)<sup>35</sup>.

Churchland's theory therefore fails to constitute a radical improvement upon the Fodorean sentential paradigm. It is interesting that she castigates anti-reductionist arguments on the grounds that they sometimes

"simply work fuzzy intuitions about what is and is not imaginable, or even about what is and is not desirable" (P.S. Churchland, 1986a, p. 327).

The burden of my argument in the second half of this chapter has been to demonstrate that Churchland's own theory rests upon little more.

### **6.3 Conclusion**

Both Churchland and Marr present co-evolutionary theories of the mind/brain, but emphasise different elements within this broad framework. I have demonstrated that both of them nevertheless fall into difficulties akin to those of the traditional metaphysical realist, since their decisions about the evidence that supports their hypotheses are not impartial, but are rather made in the light of their own preferred views of the mind/brain and the appropriate ways to study it. This means that their work does not, at present, constitute a significant advance upon the more traditional theories of mind I discussed earlier.

In the final section of this thesis, I shall summarize my conclusions and make some suggestions for future work in this area.

1. Some of the content of this chapter formed part of the paper I read at King's College, London in November, 1990.
2. The term "co-evolutionary" was coined by Churchland to denote those research strategies which are a synthesis of philosophical, psychological and neuroscientific methods. It is in this spirit that she writes that: "top-down strategies (as characteristic of philosophy, cognitive psychology, and artificial intelligence research) and bottom-up strategies (as characteristic of the neurosciences) for solving the mysteries of mind-brain function should not be pursued in icy isolation from one another. What is envisaged instead is a rich interanimation between the two, which can be expected to provoke a fruitful co-evolution of theories, models, and methods, where each informs, corrects, and inspires the other" (P.S. Churchland, 1986a, p. 3, my italics). Although Marr's approach to the nature of vision falls clearly within this very broad definition of a co-evolutionary methodology, there is fundamental disagreement between Marr and Churchland regarding the emphasis placed upon the role of neurophysiological data within the creation of a theory. See subsection 6.1.1 for further discussion.
3. This is exactly analogous to the claim made by Fodor that there can be no computation without representations (see Chapter 4, subsection 4.2.2).
4. Marr employs the term "computational" to denote a specific kind of hypothesis or theory with distinctive features that are not ordinarily associated with the term as it is more generally used. I have therefore decided to adopt Patricia Kitcher's (1988) excellent device of distinguishing between the two uses of the term by writing "Computational" with a capital "C" when I am employing it in Marr's sense. When the term is used in the wider sense, it will begin with a lower case "c". For a brief discussion of the distinction between the two uses, see also my Introduction to Section B.
5. To avoid confusion in the later stages of this chapter, in which the way in which Marr constructs an overall theory of vision by bootstrapping between hypotheses, I propose to refer to Marr's Computational theories as Computational hypotheses. Since Kitcher uses the same terminology as Marr, I shall make the same substitutions when referring to her work. This change in vocabulary is purely pragmatic and does not destroy the meaning of anything that either Marr or Kitcher has written.
6. It is worth noting that part of the task of the Computational level of explanation is to determine which constraints in the external world will be strong enough to permit the definition of the information-processing task (Marr, 1982a, p. 23). We will return to this point in subsection 6.1.2.
7. However, I shall demonstrate in the course of this chapter that the different emphasis placed upon the relative importance of neurophysiological information by Marr and Churchland has



absolutely no effect upon the difficulties they encounter in their use of the bootstrap strategy.

8. It will become apparent in later subsections that, where more details of Marr's theory are needed to understand the thrust of my argument, I have endeavoured to provide them.

9. Further details of the specific theories will be given where necessary in later subsections. I have decided to refer to them collectively in this subsection to avoid over-complicating the general outline of Marr's bootstrap argument.

10. In view of the problems inherent in Fodor's Representational Theory of Mind highlighted in Chapter 4, Marr's choice of phrase to declare that he is a realist with regard to representations and processes within the mind is perhaps rather unfortunate. He writes that "[f]rom a philosophical point of view, the approach that I describe is an extension of what have sometimes been called representational theories of mind....Modern representational theories conceive of the mind as having access to systems of internal representations; mental states are characterized by asserting what the internal representations currently specify, and mental processes by how such internal representations are obtained and how they interact" (Marr, 1982a, p. 6).

11. Note that although Marr takes a top-down stance on the methodological issues involved in studying visual processes, he argues that visual processes *themselves* operate from the bottom-up in that each representational stage is constructed from the combination of primitive elements (see Marr, 1982a, p. 52). These are very separate points and should not be confused.

12. See subsection 6.1.2 for an explanation of these labels.

13. Marr does in fact admit this possibility (see Marr, 1982a, p. 339).

14. See also Chapter 4, subsection 4.4.2.

15. It is surprising that Marr makes this mistake, since he does recognize that the impetus to construct theories may originate from many sources. In fact, Marr confesses that "there is no real recipe for this type of research - even though I have sometimes suggested that there is - any more than there is a straightforward procedure for discovering things in any other branch of science" (Marr, 1982a, p. 331).

16. It is true that Marr has provided some justification for the development of his unique methodology based upon the failure of other theories to provide a full and rigorous explanation of processes within the mind/brain. However, by elevating the description of a task in information-processing terms to the status of a necessary and primary factor in developing explanations of these processes, Marr has effectively precluded

the possibility that his analysis of the situation is mistaken. In other words, he is ignoring the possibility that these earlier theories failed because of some other difficulty and that subsequent attempts may succeed even in the absence of a Computational account.

17. I have chosen to concentrate my discussion upon the work of Patricia Smith Churchland so that I can compare her incisive negative criticisms of Fodor's argument with her neglect of the same issues of testability, realism and evidence in her positive attempt to formulate a satisfactory strategy for the development of a theory of mind/brain function. References to Paul M. Churchland are made only when I am convinced that Patricia Churchland shares his view. Although they both adopt an eliminative materialist stance, the emphasis within their arguments is slightly different, and space precludes a lengthier treatment of all aspects of their position. I have been unable to examine Stich's views for the same reason.

18. Many criticisms of Churchland will be directed against the conclusions that she draws from psychological and neuroscientific data that has obtained by research rather than against these data themselves.

19. Opinions vary as to the precise definition of "folk psychology", particularly between the eliminative materialists, who regard it as a theory and their opponents, who think that it is not. See P.M. Churchland (1991), and Wilkes (1984, 1986) for further discussion. Broadly speaking, however, "folk psychology" is the collective term for the everyday expressions that we use to describe our mental experiences, and states. It therefore includes such concepts as "beliefs", "desires", and "consciousness".

20. Further examples may be found in P.S. Churchland, 1980b, as well as elsewhere in her 1986a.

21. It is not my intention in this thesis to provide detailed criticism of the neuroscientific theories that Churchland endorses. My general line of attack will concentrate upon the conclusions that Churchland thinks such theories justify. For my arguments against the Pellionisz/Llinás tensor network theory, see my 1988a.

22. I shall argue in subsection 6.2.4 below that Churchland's bootstrapping is indeed stubborn in a very negative manner, since it involves a refusal to acknowledge two major points. Firstly, the successes of neuroscientific research projects do not necessarily entail the revision or elimination of folk psychology and secondly, the inability of folk psychology to explain all our mental experiences and states need not imply that it should be deemed wholly inadequate either.

23. This hypothesis is effectively embodied in the following passage from P.S. Churchland, 1986a, p. 374: "Some initial theory is essential to get the whole enterprise going, and broadly

speaking, folk psychology is that initial theory. We have already gone beyond folk psychology, and as neuroscience and psychology co-evolve, the likelihood is that the initial theory will by inches be revised, lock, stock, and barrel".

24. See also P.S. Churchland's examination of blindsight (1980b, pp. 191-3), which she says forces us to question either the commonsense belief that our ability to report upon a feature of the perceived world prerequisites that we have experienced this feature, or the equally basic assumption that we can only experience something if we are conscious of doing so.

25. Churchland also comments that if "diseases such as schizophrenia do have a basis we can describe in biochemical terms, this invites the idea that we might enhance our knowledge generally, of the sane as well as the insane, should we acquire knowledge of the biochemical aspects of emotions, moods, and desires and of cognitive development and organization" (P.S. Churchland, 1986a, p. 88). This is an extension of Paul M. Churchland's point that mental illness is not well accounted for by folk psychology (P.M. Churchland, 1981, p.73).

26. Paul M. Churchland cites still more reasons for rejecting the categories of folk psychology, apart from its failure to account for mental illnesses, and its inability to develop or to extend its explanations into other areas (see, for example, P.M. Churchland, 1981, pp. 72-76).

27. See above, note 20, for relevant references. Further details of these co-evolutionary accounts of mental processes will be given where necessary in the course of my critical discussion of Churchland's bootstrap strategy (subsection 6.2.3).

28. Horgan and Woodward's attack was originally directed at the arguments of P.M. Churchland, but it applies with equal force to the views of Patricia Smith Churchland.

29. See also Kitcher, 1984, pp. 98-100 for criticism of P.M. Churchland's argument by analogy with alchemy.

30. A fleeting reference to Davidson is found in the introduction to P.M. Churchland, 1989, p. xii. Here, he describes the Davidsonian view as conceptual dualism, but dismisses it as a view that is liable to render folk psychology immune from scientific refutation, because it does not adequately allow for the requirement that science provide us with unified explanations. However, it is clear that it is Churchland who is jumping the gun here, and not Davidson, because it is impossible to simply assume that unified explanations will be achievable, when bootstrapping in a very new area (see my comments in Chapter 2, above, subsection 2.5.2).

31. P.S. Churchland (1986a) includes a reference to Davidson in her *Bibliography*, but does not discuss him in her text. See my 1988a for further comparisons of the two views. It is perhaps worth differentiating Davidson's view from that of the property

dualists, which is discussed in P.S. Churchland, 1986a (pp. 323-335). Property dualists could accept a type-type identity of mental and physical events, but refuse to accept that a physical description can adequately convey the nature of our mental experiences. Davidson, on the other hand, does not accept that any such type-type identification is possible.

32. Churchland might well be inclined to argue that her hypothesis is testable, since she has provided empirical evidence in support of her view. However, this in itself does not refute my criticism. Testability is not simply a matter of citing possible sources of evidence which confirm a hypothesis. Rather, it involves exposing this hypothesis to those sources of empirical evidence that could falsify it. See my discussion of Giere, Chapter 2, subsection 2.6.

33. This is an unusual version of the semantically inconsistent tetrad. Most versions claim that it is the veracity of the overt experience that guarantees the existence and nature of the inner processes (or, in the case of the traditional veil-of-perception issue, of the external world): here, it is the failure of folk psychology that provides the guarantee, combined with the success of current neuroscientific research. However, the relationship between the hypotheses and the supporting evidence in this tetrad is nonetheless semantically inconsistent in the same way as those of previous tetrads.

34. A possible exception would be McDonough (1991).

35. cf. Shanon's criticisms of Fodor's modularity thesis in Shanon, 1988, p. 335, also above, Chapter 5, subsection 5.5.3.

**Section C - Conclusion****Chapter 7 - Evidence or Ignorance? Some final reflections****7.1 Introduction**

This concluding chapter will be comparatively brief. It is in two parts. In the first, I shall summarize the aims, strategy and findings of my argument. In the second part, I make some suggestions for the future of the bootstrapping strategy in the philosophy of mind.

**7.2 Summary and conclusions****7.2.1 Aims of the thesis**

As I remarked in my general introduction to this thesis, my discussion of realism and evidence in the philosophy of mind had four central objectives. The first, and most fundamental, of these was the evaluation of a variety of theories of the mind/brain to establish the validity of the co-evolutionary theorists' claim that their approach to the study of the mind/brain constitutes a significant advance over the more traditional theories constructed using conceptual analysis.

This primary objective could only be achieved once an appropriate criterion for the comparison of a number of radically different theories of the mind/brain had been selected. Theories of the mind/brain are no less constrained by the principles of scientific realism than any other theories. This means that, if they are to provide a good explanation of the ways that the processes and representations within the mind/brain generate our mental experiences and capacities, their constituent hypotheses must be well tested and supported by reliable sources of evidence.

Neglect of these constraints upon explanation will result in the difficulties associated with metaphysical realism, of which the veil-of-perception issue may be regarded as a paradigm case. I hoped to show that it is possible to isolate the characteristic features of metaphysical realism within the veil-of-perception issue, and thereby to draw attention to the wider implications of the veil-of-perception issue.

My remaining objectives related to the methodology employed by all the theorists whose views were examined in Section B of the thesis. I argued that all the theories discussed were constructed with the help of a bootstrapping argument which enabled them to derive both a general hypothesis and a more specific one from the same initial body of evidence,

before going on to cite a second independently obtained set of data in support of this specific hypothesis, with the subsidiary effect of providing further confirmation for the more general hypothesis. I went on to render explicit the bootstrap strategy implicit in each theory.

Further, I hoped to show that the account of the problems involved in bootstrapping to solve a crime put forward in Nordby (1989) is also relevant to the use of the strategy in the construction of co-evolutionary theories of the mind/brain.

### 7.2.2 Structure of my argument

It would clearly have been unwise to attempt an evaluation of the relative explanatory merits of the different theories of the mind/brain discussed in Section B without any preliminary discussion of some of the key factors that influence nature of explanation, notably motivation, the aim of constructing a realist theory and the question of satisfactory evidence. I therefore began Section A with a brief summary of how motivation can affect the development of theories (Chapter 1, subsections 1.1.1) and a short resume of how theories are said to explain phenomena. In subsection 1.2.3, I introduced the first central feature of my argument, the use of the veil-of-perception issue as a paradigm instance of metaphysical realism. I discussed the

problems inherent in this view in detail in Chapter 1 (subsections 1.2.4 and 1.2.5) and demonstrated their relevance to twentieth century psychology in subsections 1.2.6, 1.2.7 and 1.2.8.

In the second chapter of Section A, the difficulties involved in providing evidence capable of supporting multi-level theories were discussed further, and variations of a methodology for overcoming these difficulties were described. I argued that, even in science, there are grounds for rejecting the possibility of evaluating rival theories using the abstract criteria suggested by Glymour (1980) (see subsections 2.5.1 and 2.5.2).

In Section B, I proceeded to isolate and criticise the use of a variation of the bootstrap methodology to construct the theories of the mind/brain under view. Each theory, ranging from McGinn's theory of the hidden structure of consciousness to Churchland's eliminative materialism was evaluated in the light of the constraints upon the nature of evidence imposed by scientific realism. A summary of my main conclusions is given below.

### **7.3 Conclusions**

I contend that I succeeded in reaching all four of my primary objectives. It will be convenient to discuss my



main conclusions in the order that they appear in the body of the thesis.

Firstly, I was able to show that the veil-of-perception is readily reformulated as a problem about a metaphysical relationship between sources of evidence and the hypotheses that they are intended to support. It is possible to isolate a characteristic feature of the metaphysical realists' position within the veil-of-perception issue. This is the existence of four assumptions that, taken together, form a semantically inconsistent tetrad by virtue of the fact that the relationship between the evidence that is cited and the hypotheses is a metaphysical one and not a scientific one. Within the terms of the theory, the hypotheses cannot be refuted by conflicting evidence, and they are therefore untestable (see subsection 1.2.4).

Further, I showed that it is possible to isolate similar semantically inconsistent tetrads within the theories of the mind/brain discussed in Section B. This demonstrates that the veil-of-perception issue should be rehabilitated as a test for metaphysical realism within the philosophy of science, thereby vindicating my belief that it has been sadly undervalued and dismissed too quickly in recent philosophy. Moreover, my discussion of modularity in Chapter 5 demonstrates that the veil-of-perception issue may be relocated from its traditional position of that

between self and world with equal force.

Secondly, I believe that one of my major contributions to the current debate within the philosophy of mind has been to render the implicit bootstrapping strategies of the theories discussed in Section B explicit. So far as I know, this has not previously been attempted, with the consequence that the true relationships between the hypotheses and the evidence used to support them have been mistakenly treated as scientific, rather than metaphysical, even by the theorists themselves.

Moreover, my extension (2.5.2) of Nordby's arguments concerning the difficulties of bootstrapping in the absence of an overall theory about how to evaluate evidence has been shown in practice to apply, as I suggested in Chapter 2, to the co-evolutionary theories of mind. These theories exhibit particular difficulties in evaluating evidence from a variety of disciplines. Moreover, as I showed in my discussions of Marr and Churchland in Chapter 6, these problems are faced irrespective of the degree of emphasis given to neuroscientific data.

Of course, many of my conclusions have been at the level of detailed criticisms of the theories discussed in Section B, and it would not be appropriate for me to repeat them all here. However, I have succeeded in showing that the claim

made by the co-evolutionary theorists that they have provided a more satisfactory account of the mind/brain, and how it should be studied cannot be upheld. By making their bootstrap strategies explicit and demonstrating that, in every case, it is possible to isolate the semantically inconsistent tetrad of metaphysical realism, I have been able to show that the mere use of the bootstrap strategy is not in itself an adequate guard against the dangers of incorporating systematically untestable hypotheses within a theory. The bootstrap strategy has failed to screen out such metaphysical assumptions.

There is therefore no real difference in kind between the problems isolated with the more traditional theories of mind (such as McGinn's account of consciousness, Fodor's Representational Theory of Mind and Churchland's co-evolutionary theories), but merely a difference of degree. Indeed, I argued in Chapters 4 and 6 that there is a strong dichotomy between Churchland's attitude to the need for testable evidence within Fodor's Representational Theory of Mind (see Chapter 4 subsections 4.4.2 and 4.4.3) and her failure to observe these constraints upon her own theory as noted in Chapter 6 (subsections 6.2 - 6.2.5).

Although there is more use of empirical evidence in the more modern, co-evolutionary theories, this has not resulted in an increase in testability of the constituent

hypotheses within their bootstrap arguments. Recall that my discussion of modularity demonstrates a strong possibility of empirical underdetermination between the views of Fodor and Gardner (which Gardner systematically ignores). This indicates that the difficulties noted by Nordby of how evidence should be assessed are particularly relevant here. Moreover, my discussion of Churchland in Chapter 6 indicates that her evidence too, frequently lends unique support to her hypotheses only if the basic principles of eliminative materialism have already been accepted. Churchland relies upon metaphysical assumptions to render her empirical evidence relevant.

I therefore conclude that the factor of motivation plays an important role in determining the actual tests that theories undergo. The lack of a guiding theory controlling the use of evidence noted by Nordby makes it particularly easy for theorists to accord greater weight to their preferred views, but the tendency to provide only lax tests for their favoured arguments was also noted in the traditional theories of Fodor and McGinn in Chapters 3 and 4.

#### 7.4 Future Work

The bootstrap scheme seems sufficiently adaptive to illuminate the difficulties involved in producing adherent,

testable theories of the mind/brain without lapsing into metaphysics. It follows that the near fatal flaws that I have isolated within the theories discussed in section B could have been avoided, had their authors rendered their bootstrap strategies explicit as I have just done. Adoption of this method may yet succeed in producing better tested theories of the mind/brain, as it goes some way towards redressing the ineliminable bias caused by motivational factors. I contend that my contribution to this debate has been, as a philosopher, to illuminate the nature and place of the current deficiencies within a system of evidential reasoning and to suggest a possible solution. It is, of course, a decision for the mind/brain theorists themselves, whether they wish to follow, or whether they prefer to leave the assumptions upon which their theories are based behind a metaphysical veil.

## BIBLIOGRAPHY

- Achinstein, Peter (1983) "Concepts of Evidence", in Achinstein, Peter (ed) *The Concept of Evidence*, Oxford University Press, New York, pp. 145-174.
- Achinstein, Peter (1983) *The Nature of Explanation*, Oxford University Press, New York, Oxford.
- Akins, Kathleen A. (1990) "Book Review of Patricia Churchland: *Neurophilosophy: Toward a Unified Theory of the Mind/Brain*", *Journal of Philosophy*, 87/2: 93-102.
- Allen, Francis A. (1959) "Criminal Justice, Legal Values, and the Rehabilitative Ideal" in Jeffrie Murphy (ed), (1973) *Punishment and Responsibility*. Wadsworth, Florence, Kentucky, pp. 172-185.
- Allport, D. Alan (1980) "Patterns and Actions: Cognitive Mechanisms are Content Specific" in G. Claxton (ed), *Cognitive Psychology: New Directions*. Routledge & Kegan Paul, London.
- Arbib, M. (1972) *The Metaphorical Brain*, Wiley-Interscience. New York, Chichester.
- Ayer, A.J. (1973) *The Central Questions of Philosophy*, Weidenfeld & Nicolson, London.
- Bealer, G. (1987) "The Boundary Between Philosophy and Cognitive Science", *Journal of Philosophy*, 84/10: 553-555.
- Bennett, J. (1971) *Locke, Berkeley, Hume: Central Themes*. Oxford University Press, Oxford
- Bennett, L.J. (1988a) "How successful is Patricia Churchland's defence of the use of neurophysiological theory in the philosophy of mind?" (unpublished)
- Bennett, L.J. (1988b) "What is a mental module?", *Warwick Journal of Philosophy*, 1:175-85.
- Bennett, L.J. (1990) "Modularity of Mind Revisited", *British Journal for the Philosophy of Science*, 41/3:429-436.
- Bishop, J. (1988) "Is a Unified Science of the Mind-Brain Possible?" (review of Churchland, 1986a) and reply by P.S. Churchland) *Biology and Philosophy*, 3/3: 375-91

- Blakemore, C. (1988) *The Mind Machine*, BBC Books, London
- Blakemore, C. & Greenfield, S. (eds.) (1987) *Mindwaves: Thoughts on Intelligence, Identity and Consciousness*, Basil Blackwell, Oxford, New York.
- Boden, M.A. (1988) *Computer Models of Mind*, CUP, Cambridge, London, New York, New Rochelle, Melbourne, Sydney
- Boden, M.A. (1989) *Artificial Intelligence in Psychology: Interdisciplinary Essays*. MIT Press, Cambridge, Mass., London, England.
- Braithwaite, R.B. (1968) *Scientific Explanation: A Study of the Function of Theory, Probability and Law in Science*. Cambridge University Press, London, New York.
- Campbell, Keith (1986) "Can Intuitive Psychology Survive the Growth of Neuroscience?", *Inquiry*, 29/2: 143-52.
- Churchland, P.M. (1975a) "Two Grades of Evidential Bias", *Philosophy of Science* 42/3:250-259.
- Churchland, P.M. (1975b) "Karl Popper's Philosophy of Science", *Canadian Journal of Philosophy* 5/1:145-156.
- Churchland, P.M. (1979) *Scientific realism and the plasticity of mind*, Cambridge University Press, Cambridge, London, New York, Melbourne.
- Churchland, P.M. (1980a) "In defense of naturalism" [commentary on Fodor (1980)], *The Behavioral and Brain Sciences*, 3:74-5.
- Churchland, P.M. (1980b) "Plasticity: conceptual and neuronal" [commentary on Pylyshyn (1980)], *The Behavioral and Brain Sciences*, 3:133-4.
- Churchland, P.M. (1981) "Eliminative materialism and the propositional attitudes", *Journal of Philosophy*, 78/2:67-90.
- Churchland, P.M. (1984) *Matter and Consciousness - A Contemporary Introduction to the Philosophy of Mind*, MIT Press, Cambridge, Mass., London.
- Churchland, P.M. (1985) "The Ontological Status of Observables: In Praise of the Superempirical Virtues", in Paul M. Churchland and Clifford A. Hooker (eds.) *Images of Science*, Chicago

University Press, Chicago and London, pp. 35-47.

- Churchland, P.M. (1986a) "The Continuity of Philosophy and the Sciences", *Mind and Language*, 1/1: 5-14.
- Churchland, P.M. (1986b) "Semantic Content: In defense of a network approach", [commentary on Sayre (1986)], *The Behavioral and Brain Sciences*, 9/1: 139-40.
- Churchland, P.M. (1989) *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, MIT Press, Cambridge, Mass., London.
- Churchland, P.M. (1991) "Folk psychology and the explanation of human behaviour" in Greenwood (ed) (1991), pp. 51-69.
- Churchland, P.M. & Churchland, P.S. (1983a) "Stalking the Wild Epistemic Engine", *Noûs*, 17: 5-18.
- Churchland, P.M. & Churchland, P.S. (1983b) "Content: Semantic and information-theoretic" [commentary on Dretske (1983)], *The Behavioral and Brain Sciences*, 6: 67-8.
- Churchland, P.S. (1978) "Fodor on Language Learning", *Synthese*, 38/1: 149-159.
- Churchland, P.S. (1980a) "Language, Thought and Information Processing", *Noûs*, 14: 147-170.
- Churchland, P.S. (1980b) "A Perspective on Mind-Brain Research", *Journal of Philosophy*, 77/4: 185-207.
- Churchland, P.S. (1980c) "Neuroscience and psychology: should the labor be divided?" [commentary on Pylyshyn (1980)], *The Behavioral and Brain Sciences*, 3: 132-3.
- Churchland, P.S. (1981) "On the Alleged Backwards Referral of Experiences and its Relevance to the Mind-Body Problem", *Philosophy of Science*, 48: 165-81.
- Churchland, P.S. (1983) "Consciousness: The transmutation of a concept", *Pacific Philosophical Quarterly*, 64: 80-95.
- Churchland, P.S. (1986a) *Neurophilosophy: Toward a Unified Science of the Mind/Brain*



- (*Computational Models of Cognition and Perception*), A Bradford Book, MIT, Cambridge, Mass., London, England
- Churchland, P.S. (1986b) "Replies to Comments" (on Churchland, 1986a), *Inquiry*, 29/2: 241-72.
- Churchland, P.S. (1987) "Epistemology in the Age of Neuroscience", *Journal of Philosophy*, 84/10:544-553.
- Clark, Andy (1989) *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. MIT (Bradford Books), Cambridge, Mass.; London, England.
- Corballis, M.C. (1988) "Psychology's Place in the Science of the Mind/Brain?" [review of Churchland, 1986a and reply by P.S. Churchland {Churchland, P.S. 1988}], *Biology and Philosophy*, 3/3: 363-73.
- Davidson, Donald (1980) *Essays on Actions and Events*, Oxford University Press, New York.
- Descartes, René (1982) *Philosophical Works*, Vol. I, translated by E.S. Haldane & G.R.T. Ross, CUP, Cambridge, London, New York, New Rochelle, Melbourne, Sydney.
- Dennett, D.C. (1981) *Brainstorms: Philosophical Essays on Mind and Psychology*, Harvester Press, Brighton, Sussex.
- Dennett, D.C. (1988) "Psychosemantics: Fodor - review", *Journal of Philosophy*, 85/7: 384-389.
- Dretske, F. (1981) *Knowledge and the Flow of Information*, Basil Blackwell, Oxford.
- Dretske, F. (1983) "Précis of Knowledge and the Flow of Information", *The Behavioral and Brain Sciences*, 6:55-90.
- Dretske, F. (1986) "Misrepresentation", in R.J. Bogdan (ed.) *Belief, Form, Content and Function*, Clarendon Press, Oxford, pp.17-36
- Earman, John (ed) (1983) *Testing Scientific Theories*, Minnesota Studies in the Philosophy of Science, vol. 10, University of Minnesota Press, Minneapolis.
- Ellis, B. (1985) "What Science aims to Do", in Churchland and Hooker (ed) (1985) *Images of*

Science, University of Chicago Press, Chicago and London, pp. 48-74.

Fodor, J.A. (1976) *The Language of Thought*, Harvester Press, Sussex.

Fodor, J.A. (1980) "Methodological Solipsism considered as a research strategy in cognitive psychology", *The Behavioral and Brain Sciences*, 3: 63-109.

Fodor, J.A. (1981) *Representations: Philosophical Essays on the Foundations of Cognitive Science*, Harvester Press Ltd, Brighton, Sussex.

Fodor, J.A. (1983) *The Modularity of Mind: An Essay on Faculty Psychology*, A Bradford Book, MIT, Cambridge, Mass., and London, England.

Fodor, J.A. (1984a) "Observation Reconsidered", *Philosophy of Science*, 51: 23-43.

Fodor, J.A. (1984b) "Semantics, Wisconsin Style", *Synthese* 59: 231-250.

Fodor, J.A. (1985a) "Fodor's guide to Mental Representation: The Intelligent Auntie's Vade-Mecum", *Mind*, 94: 76-100.

Fodor, J.A. (1985b) "Précis of the Modularity of Mind (with peer commentaries and author's response)", *The Behavioral and Brain Sciences*, 8: 1-42.

Fodor, J.A. (1986a) "Why Paramecia Don't Have Mental Representations", *Midwest Studies in Philosophy*, vol.X (Studies in the Philosophy of Mind) ed: Peter A. French, Theodore E. Vehling Jnr., and Howard K. Wettstein, pp.3-23

Fodor, J.A. (1986b) "Banish disContent" in J. Butterfield (ed.) *Language, Mind and Logic*, CUP, Cambridge, London, New York, New Rochelle, Melbourne, Sydney, pp.1-23.

Fodor, J.A. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Bradford book (in co-operation with the British Psychological Society), MIT, Cambridge, Mass., London, England.

Fodor, J.A. (1988) "A Reply to Churchland's 'Perceptual Plasticity and Theoretical Neutrality'", *Philosophy of Science*, 55: 188-198.

- Gardner, H. (1985) *Frames of Mind: The Theory of Multiple Intelligences*, Paladin, London.
- Giere, Ronald N. (1983) "Testing Theoretical Hypotheses" in Earman (1983), pp. 267-298.
- Gillett, G. (1988) "Perception and Neuroscience", *British Journal for the Philosophy of Science*, 19.(page nos. needed).
- Gillett, G. (1989) "Representation and Cognitive Science", *Inquiry*, 32/3:261-76.
- Glymour, C. (1980) *Theory and Evidence*, Princeton University Press, Princeton, New Jersey, U.S.A. & Guildford, Surrey, England.
- Glymour, C. (1983) "Relevant Evidence" in Achinstein, Peter (ed) *The Concept of Evidence*, Oxford University Press, New York, pp. 124-144.
- Glymour, C. (1983) "On Testing and Evidence" in Earman (1983), pp. 3-26.
- Goldman, A.I. (1987) "Cognitive Science and Metaphysics", *Journal of Philosophy*, 84/10: 537-544.
- Greenwood, John D. (ed) (1991) *The future of folk psychology*, Cambridge University Press, Cambridge, New York, Port Chester, Melbourne, Sydney.
- Greenwood, John D. (1991a) "Introduction: Folk Psychology and scientific psychology" in Greenwood (1991), pp. 1-21.
- Greenwood, John D. (1991b) "Reasons to believe" in Greenwood (1991), pp. 70-93.
- Gregory, R.L. (1970) *The Intelligent Eye*, Weidenfeld & Nicolson, London.
- Hacker, P. (1987) "Languages, Minds and Brain", in C. Blakemore and S. Greenfield (eds.) (1987), pp.485-505
- Hacking, Ian (1983) *Representing and Intervening*, CUP, Cambridge, New York, New Rochelle, Melbourne, Sydney.
- Hofstadter, D.R. & Dennett, D. (eds) (1982) *The Mind's Eye*, Penguin Books Ltd, Harmondsworth, England

- Horgan, Terence & Woodward, James (1985) "Folk Psychology is Here to Stay", *The Philosophical Review*, 94/2:197-226.
- Horowitz, Amir (1990) "Dretske on Perception", *Ratio (New Series)*, 3/2:136-41.
- Hull, David (1974) *Philosophy of Biological Science*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
- Hume, D. (1978) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, 2nd edition, revised P.H. Nidditch, Oxford University Press, Oxford.
- Hundert, Ed (1989) *Philosophy, Psychiatry and Neuroscience: Three Approaches to the Mind*, Clarendon Press, Oxford.
- Hunt, G.M.K. (1990) "Laws: projectability and uniformity", *International Studies in the Philosophy of Science*, 4/3:241-6.
- Johnson-Laird, P.N. (1983) *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*, Cambridge University Press, Cambridge, London, New York, New Rochelle, Melbourne, Sydney.
- Johnson-Laird, P.N. (1988) *The Computer and the Mind: An Introduction to Cognitive Science*, Fontana Paperbacks, London.
- Kitcher, Patricia (1980) "Discussion: How to reduce a functional psychology?", *Philosophy of Science*, 47:134-140.
- Kitcher, Patricia (1984) "In Defense of Intentional Psychology", *J. Phil*, 81/2: 89-106.
- Kitcher, Patricia (1982) "Discussion: Genetics, Reduction and Functional Psychology", *Philosophy of Science*, 49:633-636.
- Kitcher, Patricia. (1988) "Marr's Computational Theory of Vision", *Philosophy of Science*, 55/1:1-24.
- Lachman, R., Lachman, J. & Butterfield, E.C. (1979) *Cognitive Psychology and Information Processing: An Introduction*, Lawrence Erlbaum Associates, Hillsdale, New York, London.
- Liberman, A.M. (1974) "The Specialization of the Language Hemisphere", *The Neurosciences, Third Study Program*, Francis O. Schmitt, Frederic G.

- Worden (eds.), MIT, Cambridge, Mass.; London, England, pp. 43-56.
- Locke, John (1981) *An Essay Concerning Human Understanding*, edited and abridged by A.D. Woodley, Collins Fount Paperbacks, Glasgow.
- McMullin, Ernan (1992) *The Inference That Makes Science: The Aquinas Lecture, 1992*, Marquette University Press, Milwaukee.
- McDonaugh, Richard (1991) "A culturalist account of folk psychology" in Greenwood (1991), pp. 263-288.
- McGinn, C (1982) *The Character of Mind*, Oxford University Press, Oxford.
- McGinn, C. (1990) *The Problem of Consciousness*, Basil Blackwell, Oxford.
- Madell, Geoffrey (1986) "Neurophilosophy: A Principled Sceptic's Response" [commentary on Churchland, P.S. {1986a}], *Inquiry*, 29/2:153-68.
- Marcel, Anthony J. (1983) "Conscious and Unconscious Perception", *Cognitive Psychology*, 15:238-300.
- Marr, D. (1982a) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W.H. Freeman and Company, New York.
- Marr, D. (1982b) "Representing and Computing Visual Information" in P.H. Winston and R.H. Brown (eds.) *Artificial Intelligence: An MIT Perspective*, MIT, Cambridge, Mass., London, England, vol.2.
- Marras, Ausonio (1983) "Taming the Wild Epistemic Engine: Reply to the Churchlands", *Noûs*, 17:21-22.
- Marshall, John C. (1984) "Multiple Perspectives on Modularity", *Cognition*, 17:209-242.
- Morgan, M.J. (1984) "Computational Theories of Vision", *Quarterly Journal of Experimental Psychology*, 36A: 157-165.
- Neisser, Ulric (1976) *Cognition and Reality*, W.H. Freeman & Co, San Francisco.
- Newton-Smith, W.H. (1981) *The Rationality of Science*, Routledge, London & New York.

- Nordby, J. J. (1989) "Bootstrapping while Barefoot (Crime Models vs. Theoretical Models in the Hunt for Serial Killers)", *Synthese*, 81/3: 373-389.
- Papineau, D. (1987) *Reality and Representation*, Basil Blackwell, Oxford.
- Place, Ullin T. (1989) "Towards a Connectionist Version of the Causal Theory of Reference", *Acta Analytica*, 3:71-97.
- Plato (1892) *The Republic*, tr. J.L. Davies and D.J. Vaughan, Macmillan & Co., London, New York
- Popper, Karl R. (1969) *Conjectures and Refutations*, Routledge and Kegan Paul, London.
- Preston, John M. (1989) "Folk Psychology as Theory or Practice? The Case for Eliminative Materialism", *Inquiry*, 32/3: 277-303.
- Putnam, Hilary (1984) "Models and Modules", *Cognition*, 17: 253-264.
- Pylyshyn, Zenon W. (1980) "Computation and Cognition: issues in the foundations of cognitive science", *The Behavioral and Brain Sciences*, 3: 111-169.
- Salmon, W.C. (1989) *Four decades of scientific explanation*, University of Minnesota Press, Minneapolis.
- Samet, J. (1986) "Troubles with Fodor's Nativism", *Midwest Studies in Philosophy*, 10: pp.575-594.
- Sereno, Martin (1986) "A Program for the Neurobiology of Mind" [commentary on Churchland, P.S. {1986a}], *Inquiry*, 29/2:217-40.
- Shallice, Tim (1984) "More functionally isolable sub-systems but fewer 'modules'?", *Cognition*, 17: 243-252.
- Shallice, Tim (1988) *From Neuropsychology to Mental Structure*, CUP, Cambridge, New Rochelle, Melbourne, Sydney.
- Shanon, B. (1988) "Remarks on the Modularity of Mind", *British Journal for the Philosophy of Science*, 39: 331-52.

- Simon, H.A. (1981) "Cognitive science: the newest science of the artificial" in D.A. Norman (ed.) *Perspectives on Cognitive Science*, Lawrence Erlbaum Associates, Norwood, New Jersey, USA.
- Skarda, Christine A. (1986) "Explaining Behavior: Bringing the Brain Back In" [commentary on Churchland, P.S. {1986a}], *Inquiry*, 29/2: 187-202.
- Sloman, A. (1985) "Why We Need Many Representation Formalisms" in M.A. Bramer (ed) *Research and Development in Expert Systems*, CUP, Cambridge, London, New York, New Rochelle, Melbourne, Sydney, pp.163-183.
- Sloman, A. (1989) "On Designing a Visual System", Cognitive Science Research Paper No. CSRP 146, School of Cognitive and Computing Sciences, The University of Sussex, Brighton.
- Smith, Aaron (1986) "Brain-mind Philosophy" [commentary on Churchland, P.S. {1986a}], *Inquiry*, 29/2: 203-15.
- Sternberg, R.J. (1985) *Beyond IQ: A Triarchic Theory of Human Intelligence*, CUP, Cambridge.
- Stoljar, Natalie (1988) "Churchland's Eliminativism", *Australasian Journal of Philosophy*, 66/4: 489-97.
- Suppe, Frederick (ed.) (1977) *The Structure of Scientific Theories*, University of Illinois Press, Urbana, Chicago, London.
- Tallis, Raymond (1991) "A Critique of Neuromythology" in R. Tallis & H. Robinson (eds.) *The Pursuit of Mind*, Carcanet Press Ltd, Manchester, pp. 86-109.
- Thorn, J.O. & Collocott, T.C. (1984) *Chambers Biographical Dictionary (Revised Edition)*, W.& R. Chambers Ltd, Edinburgh.
- Ullman, S. (1980) "Against direct perception", *The Behavioral and Brain Sciences*, 3: 373-415.
- van Frassen, B. (1983) "Theory Comparison and Relevant Evidence" in Earman (1983), pp. 27-42.
- von Eckardt, Barbara (1984) "Cognitive Psychology and Principled Skepticism" in *J. Phil*, 81/2: 67-88.

- Wasow, T. (1978) "Review of Fodor's Language of Thought", *Synthese*, 38: 161-167.
- Weiskrantz, L. (1986) *Blindsight: A Case Study and Implications*, Oxford Science Publications, Oxford University Press, Oxford.
- Wilkes, K.V. (1984) "Pragmatics in Science and Theory in Common Sense", *Inquiry*, 27: 339-61.
- Wilkes, K.V. (1986) "Nemo Psychologus nisi Physiologus" [commentary on Churchland, P.S. {1986a}], *Inquiry*, 29/2: 169-85.