The Architecture of Belief

John Fitzpatrick

A Thesis Submitted
for the
Degree of Doctor of Philosophy
of
The Australian National University

October 1989
I wish to thank my present supervisors Frank Jackson and Philip Pettit for their
detailed and helpful comments. Special thanks must go to Kim Sterelny, who has
continued in a supervisory role even though his leaving the ANU, for tenure
across the Tasman, in no way obligated him to do so. Thanks and best wishes
should also go to other past and present colleagues of the Philosophy Department
R.S.S.S. whose interests have overlapped with mine viz. my co-author David
Braddon-Mitchell, Paul Griffiths and Karen Neander. Thanks also to all my
other colleagues and friends, both within and without the department, of the last
three years—they know who they are. This work has also benefited from the
stimulating intellectual environment of the Australian National University:
there was never a shortage of interesting visitors and Fellows with whom to
interact philosophically.

In the last weeks of writing this thesis, I was saddened to hear of the
death of Don Mannison, one of my Honours year supervisors at the University of
Queensland. Don was a considerable influence upon my deciding to get back into
philosophy, and to undertake a Ph.D. Although he would probably not have
approved of much of the content of this project, the project would not have been
started, much less completed, without him.
Abstract

This thesis examines the theoretical underpinnings of intentional realism—the view that cognitive science and psychology will quantify over states such as beliefs and desires. Intentional realism contains the following three elements. First, it attributes a certain type of state to a system whose behaviour it is attempting to explain. I introduce the distinction between Level One and Level Two analyses of systems. Intentional realism postulates Level Two states of the system. Second, the states intentional realism quantifies over are representational states of a certain sort. In chapter 5 I distinguish relatively coarse-grained representational states, those that represent states of affairs, from fine-grained ones, and it is the former which intentional realism quantifies over. Thirdly, in chapter 6, we see that these Level Two coarse-grained representational states are deemed to be the causes of, at least, non-verbal behaviour. I go on to argue that intentional realism is weak on two fronts. On the first front, I claim that it is under Level One analysis that one individuates the properties of a system that are criterial of something's being a cognitive system, whereas The-One-True-Cognitive-Psychology seems to postulate Level Two states of a system. I claim that the potential for the supporters of intentional realism making a Level mistake is very real. The second front has to do with whether or not there will be Level Two representational states of the sort that intentional realism requires. I argue in chapter 9 that there is a body of evidence which seems to count against intentional realism.
Contents

Acknowledgements iii

Abstract iv

Prologue vii

Chapter 1
The Topography of Intentionality 2
1. Intentional Psychology 3
2. The Taxonomy 5
3. Intentional Psychology and Functionalism 9
4. Intentional States 23

Part I: Levels

Chapter 2
Complex Systems 29
1. The Two Levels 30
2. Level Mistakes 37
3. Level Two 40
4. Why Complexity? 52

Chapter 3
Autonomy 54
1. Descriptive Autonomy 56
2. Methodological Autonomy 58
3. Developmental Autonomy 64
4. Confirmation Autonomy 67
5. Function and The Neurosciences 68

Chapter 4
Cognitive Systems 72
1. Representation and Cognition 73
2. Cognitive Systems 79
3. Connectionism 92
4. Inexplicit Representation 95
Contents vi

Part II: Modularity

Chapter 5  
Function and Domain Specificity 99  
1. Domains 101  
2. The Grain Problem 107  
3. Proprietary Codes 110  
4. Domains and Function 113  
5. How Much of a Distinction? 121

Chapter 6  
Cognitive State Realism 124  
1. The Nature of Cognitive States 125  
2. Abstraction 132  
3. Cognitive State Realism at Work 134

Part III: Intentional Realism

Chapter 7  
What’s Wrong with Functionalism? 144  
1. Chauvinism or Liberalism? 145  
2. Schiffer 153  
3. Putnam and Multiple Realisability 156

Chapter 8  
Broad Content 163  
1. Can “Beliefs Be In The Head”? 165  
2. Individuation vs Constitution 172  
3. Supervenience 174

Chapter 9  
Two Theories of Cognitive Architecture 182  
1. Vertical Faculty Theory 183  
2. Implications of VFT 190  
3. A Priori Arguments Against VFT 191  
4. Empirical Arguments 201

Epilogue  
Dispensibility 212  
1. Is Folk Psychology Too Fine Grained in Its Explanations? 213  
2. A Broader Conception of Psychology 218  
3. The Social Sciences 220

Appendix A  
Historical Antecedents to VFT 222

Appendix B  
Neuropsychology 239

Bibliography 251
The province of the philosopher, Bertrand Russell once claimed, is the unknown: philosophers speculate about the things we don’t know about. The difference between science and philosophy is that science is what we know whereas philosophy is what we don’t know (1960 p.1). While I don’t think that this is an adequate description of all philosophical practice, I do think it nicely captures the practice of philosophical speculation—what Fodor (1975) called “speculative psychology”—about the future development of scientific psychology, what I call below The-One-True-Cognitive-Psychology. The-One-True-Cognitive-Psychology is a subdiscipline of what has become known as cognitive science, a generic, and as yet embryonic discipline taking as its source, disciplines as seemingly diverse as philosophy, linguistics, computer science, psychology and, perhaps, even neuroscience, depending upon who one reads.

What direction will this rising discipline take? We commonly explain and predict agents’ actions by attributing states such as beliefs, desires, hopes, fears, etc. to agents. Many theorists about cognitive science and psychology think that The-One-True-Cognitive-Psychology will quantify over such states. I will call this view intentional realism.

This work hopes to lay some theoretical foundations which can, I hope, provide the basis of an answer to the question: will states such as beliefs and desires feature in The-One-True-Cognitive-Psychology? I hope this work will make intelligible the kinds of evidence that could count in deciding an answer to this question. That is because, ultimately, the answer to this question is an empirical one. Many philosophically interested theorists have claimed that we can have the answer to the question now rather than when the empirical data is in, because of certain conceptual problems inherent in the intentional realism doctrine. The foundations I present in this work count against this armchair answer to the question. That is not to say that I will be arguing for intentional realism; the foundations presented below will, I hope, allow one to examine some empirical data with an eye to perhaps in the future answering the question. In the final, highly speculative chapter, I examine some attempts at empirically answering our question. It is this speculative nature of any nonarmchair based attempt at answering our question where Russell’s
description becomes relevant: it's where philosophical psychology meets serious empirical psychology. Having said that, I should stress that the substantial portion of this work is theoretical, in fact highly so.

The plan is this. Chapter 1 provides a taxonomy of positions regarding states such as beliefs and desires and their relation to The-One-True-Cognitive-Psychology. It's here that I begin outlining the thesis of intentional realism.

In Part I the laying of the theoretical foundations begins. If beliefs and desires are states of an agent then we need some account of how to go about attributing states to a system such as a cognitive agent. In chapter 2 I therefore argue that there are at least two ways that we ordinarily go about attributing states to a system, one based on input-output regularities of the system, and those based upon decompositions of that system. This difference amounts to that of speculating about the properties of a system construed either as a black box or as a black box which has been opened. I argue here that The-One-True-Cognitive-Psychology attributes states according to this latter mode. I also introduce in this chapter the idea of a level of explanation-description. On attributing states to a system according to the latter method there will be many attributions depending upon the level of explanation-description in which one is interested. If one decides upon the latter type of state attribution, then one should decide upon the relationship between these levels. Do they affect each other? Might some be redundant given the completion of a description of a system at some lower level? Does one level reduce to another? It is issues such as these which are tackled in chapter 3. It's important to have some views about such issues because the kind of empirically based arguments outlined in Part III depend upon these issues turning out a certain way. Chapter 4 attempts to identify the properties of systems which endow those systems with their cognitive status. This cognitive status is gained by possessing certain kinds of representational capacities. I argue that to the extent that The-One-True-Cognitive-Psychology wants to quantify over representational states of a system identified as a result of opening up the black box, it is open to making what I call a Level mistake.

Part II moves from the general issues regarding levels to specific cognitive issues. Chapter 5 attempts to identify the underlying principles by which state attributions to the now opened cognitive black box are made. In other words, this chapter is about how we individuate cognitive mechanisms. Some cognitive mechanisms are here deemed to manipulate representations which are fine-grained, in that they take as their inputs a very narrow range of information from the system's environment, while others are coarse-
grained. Having decided about the ways of taxonomising cognitive mechanisms, chapter 6 identifies a thesis which claims that some of the cognitive mechanisms over which The-One-True-Cognitive-Psychology quantifies will be coarse-grained. This thesis, which I call cognitive state realism, is an essential component of the intentional realist programme. It is this thesis which I think is problematic for intentional realism. It is the burden of Part III to show that it may well be false.

Chapter 9 is the main component of Part III. In it I argue that cognitive state realism may well be false, given certain empirically based arguments stemming from neuropsychology and cognitive psychology. I claim that there is an alternative view of representational cognitive structure, that is supported by this evidence. I call this view vertical faculty theory (VFT). If VFT turns out to be true, then cognitive state realism is false.

However, that chapter 9 might well be considered redundant, if some traditional arguments against intentional realism were sound. So, some traditional objections are considered in chapters 7 and 8. Chapter 7 looks at Functionalism, a theory of the mind upon which intentional realism depends. There are many objections to the functionalist programme getting off the ground. I don't think those objections hold, primarily for reasons which have become evident in the earlier chapters. Hence, the foundations I offer have some weight in the philosophy of psychology.

Chapter 8 tackles the thorny issue of the semantic properties of intentional states and their relation to the states likely to be postulated by The-One-True-Cognitive-Psychology. One strongly philosophical thesis to which intentional realism is committed, through its reliance upon functionalism, is that of mind-brain supervenience. If the arguments from the content of states, such as beliefs and desires are sound, then mind-brain supervenience will fail, and hence, so does intentional realism. Needless to say, I don't think these arguments work.

The debates surrounding cognitive state realism and VFT have a history stretching back to the last century. Appendix A describes earlier incarnations of that debate. Appendix B describes in detail neuropsychological evidence which supports the empirical component to the arguments of chapter 9. Ideally, both Appendices should be read before chapter 9.
The Architecture of Belief
Chapter 1

The Topography of Intentionality

The chapters which follow are designed to provide the basis for answering the question, posed in the Prologue, regarding the development of the cognitive sciences. That question, remember, concerns the traditional mental concepts such as belief, desire, hope, etc., and their role in a mature science of the mind—The-One-True-Cognitive-Psychology. There are various views as to the role of these states with respect to cognitive science, only one of them being the view I am calling *intentional realism*. It is the view that folk psychological states, processes and generalisations, or, at least, approximations of them, are going to feature in The-One-True-Cognitive-Psychology. An affirmative or negative answer to our question will, respectively, lead to either acceptance or rejection of this view.

By way of introducing our task, we can look at the kinds of entities over which folk psychology quantifies, and what kind of account one might give of them. That will lead us into the various views regarding the relation between folk psychology and cognitive science, thus enabling us to isolate the position held by intentional realism. So, the plan is this. Firstly, we can look at the kinds of entities postulated by folk psychology, and examine the various philosophical accounts of those entities. Then I want to outline a taxonomy of the various positions regarding the status of intentional psychology and the future development of The-One-True-Cognitive-Psychology, and, hence, determine to what these views are committed. This will allow us to see where intentional realism falls in the terrain of intentionality. Having done that we are then in a position to examine in Parts I and II some of the theoretical assumptions underlying the intentional realist programme.
1 Intentional Psychology

Suppose we want to explain why Ronald Reagan ordered the destruction of some Iranian gun boats. One explanation might run thus:

Reagan believed that the Iranians were pirating Hollywood westerns and that the boats were smuggling copies across the Gulf. Reagan desired that the pirating of Hollywood westerns stop. He also believed that the ordering of the destruction would be an effective means towards ending pirating.

Explanations of the above form attribute mental states of the form of

(1) A φs that p

where A is replaced by some agent, such as Reagan, φ is replaced in some folk psychological verb, such ‘believes’ or ‘desires’, and p is replaced by a sentence representing some actual or possible state of affairs. Sentences of the form of (1) are intriguing because of the problems they pose for quantification, substitutivity of identicals and the nature of propositions. Sentences similar to (1) also seem to have psychological salience. But sentences such as (1) are used to attribute mental states to agents. Such attributions are thus of interest to the psychologist and the philosopher of mind/psychology because they sometimes feature in the explanation of agents’ actions.

Such explanations are ubiquitous. They’ve been given by the folk for centuries as explanations of agent actions. Hence the labels common sense or folk psychological explanations. Characteristic of each of these explanations is the reference to some state of affairs to which the agent bears some relation. Or alternatively, the state of affairs may be some possible state of affairs the agent wants to become actualised. States of affairs can be represented by propositions, with those actualised states of affairs being represented by true propositions. So we can take that to which the agent is related to be the proposition (representing a state of affairs) expressed by the sentential complement in the sentences explaining the agent’s behaviour.

The relation to these propositions representing states of affairs is expressed by the verbs (believes, desires, thinks, remembers, etc.) in the sentences of the explanations, the sentential complements of which contain the phrase expressing the proposition. Because the agent thinks or wants the propositions representing the states of affairs to be true, we can say that the agent has an attitude to those propositions; Reagan believes or fears that
Hollywood westerns are being pirated by the Iranians and desires or wants that state of affairs to stop, and, therefore, the proposition expressed by 'Hollywood westerns are not being pirated by the Iranians' to be true. Hence, the philosopher's term (Russell's) 'propositional attitude' for the relation between the agent and states of affairs that these verbs express.

It follows, then, that common sense or folk psychological explanations are propositional attitude psychological explanations. Because propositional attitudes involve relational properties of agents, that is, agents are related to states of affairs and hence propositional attitudes exhibit an "aboutness" or a "directedness towards objects"—what Brentano called intentionality—I call propositional attitude based psychology intentional psychology.\(^1\)

Intentional psychology seems to work. First of all, we get predictive power. If we decide that Reagan has the beliefs and desires that he does, then, ceteris paribus, nuking the Iranians will result in the cessation of the pirating of the Hollywood westerns. Part of the reason why intentional psychology has the predictive power it does is that it generalises to counterfactual instances. Propositional attitude ascriptions work across any individuals in any context you like—your closest associates or absolute strangers, in the laboratory, or in the field.\(^2\) What attitudes we decide to attribute is determined, in part, by inferences from behaviour and personal history. The pirating of Hollywood westerns is something Reagan is likely to care about given his professional history and quirks of personality. We also attribute attitudes on the basis of an agent's stated intentions and assertions. When Reagan tells us that he believes that the Iranians are movie pirates we take that to be a good indicator of what he believes and how he might act given his other beliefs and desires. This assumes, of course, that the agent is sincere, a competent language user etc.. But the ceteris paribus clauses take care of such contingencies—more on ceteris paribus clauses below, however.

As well as these third person propositional attitude attributions, intentional psychology works in the first person. I potted the black because I

---

\(^1\)Strictly speaking, intentionality need not only be associated with propositional attitudes. Presumably, non-propositional entities such as concepts can exhibit intentionality. I call propositional attitude psychology intentional psychology in keeping with the terminology of the cognitive literature which, fairly justifiably, takes the propositional case to be the most interesting case of intentionality.

\(^2\)In chapter 6 we will briefly look at cognitive dissonance studies and work in attribution theory. If the theoretical implications drawn from these studies turn out to be correct then intentional psychology might well break down in the laboratory, to the extent that agents might get the causes of their behaviour wrong. That, I take, though, does not undermine intentional psychological explanations as such (since there might be other beliefs and desires at work), but rather brings into doubt the issue of introspective access to the causes of behaviour.
desired to maximise my snooker score, and knew that this was a way of achieving that desire.\textsuperscript{3}

In sum, intentional psychology quantifies over the propositional attitudes, and postulates law-like generalisations of the following forms:

if an agent fears that \( p \), then the agent desires that not \( p \)

if an agent hopes that \( p \) and discovers that \( p \), then the agent is pleased that \( p \)

if an agent believes that \( p \) and if \( p \) then \( q \), then the agent comes to believe that \( q \)

if A desires that \( p \) and believes that doing \( x \) will bring about \( p \), then the agent does \( x \)

with \textit{ceteris paribus} clauses inserted in the relevant positions.\textsuperscript{4}

Granted that intentional psychology in some sense works, just how it is \textit{taken} to work and how it \textit{ought} to work are central to determining the extent to which one buys into an intentional psychology. I intend these questions of how intentional psychology is taken to and should work to be asking how intentional psychology is to be related to a future, mature, scientific cognitive psychology—The-One-True-Cognitive-Psychology.

\section*{2 The Taxonomy}

We now turn to our taxonomy. Where one ends up in the taxonomy depends upon one's views about three theses regarding the relation between intentional psychology and The-One-True-Cognitive-Psychology. It is to these theses we now turn.

\subsection*{2.1 Consistency}

If someone adheres to the consistency thesis, they believe that even when The-One-True-Cognitive-Psychology is completed and deemed to be \textit{true}, we will still be able to use intentional psychology in order to explain intelligent

\textsuperscript{3}For a summary of the ways in which intentional psychology works see just about any of Fodor (1975) through (1987), in particular (1987 Ch. 1).

\textsuperscript{4}For a discussion of the generalisations employed by intentional psychology see for example Paul Churchland (1981) and Fodor (1988).
behaviour. So, adherents of the consistency thesis range from Fodor and Lycan to Dennett, Davidson, Jackson and Pettit. As there are vast differences between the views of these theorists, the reasons why they accept the consistency vary. We will return to their cases shortly. Those who reject this thesis will most obviously be the eliminative materialists such as the Churchlands and Steve Stich.

The Churchlands take intentional psychology to be an empirical theory. It's a theory because it purports to explain and, with a fair degree of success, predict actions and behaviour by postulating states of agents, such as the propositional attitudes, in conjunction with generalisations and rough and ready laws, as is the case with all empirical theories. According to the current view, however, this theory is grossly inadequate. Intentional psychology has nothing to say about the cognitive capacities psychology ought to be able to tell us about, for example, facial recognition, long and short term memory, speech production and comprehension, motor control, etc. It is also a stagnant theory, not having been revised for over two thousand years—except, perhaps, for the inclusion of subconscious beliefs, desires and drives into the intentional ontology. Intentional psychology is therefore a degenerating research program (Paul Churchland 1981).

Steve Stich finds this line of criticism congenial. On his view, as with the Churchlands, the main problem with intentional psychology is that the kinds it quantifies over (the propositional attitudes) aren’t going to feature in a scientific psychology, but for very precise reasons. Intentional states such as beliefs and desires are individuated on the intentional psychological story by their respective contents. Taxonomy by content, however, appeals to the semantic properties of states such as truth conditions and the like. But to get a handle on these properties we need to go beyond the boundaries or skin of the agent. The attribution of content is also, according to Stich, a highly graded and pragmatic enterprise. But The-One-True-Cognitive-Psychology is going to be interested only in things internal to the agent (Stich's principle of autonomy) which are attributable in a precise scientific manner. Hence, intentional psychology is not going to be consistent with The-One-True-Cognitive-Psychology (Stich 1983). We will encounter some of these issues again in chapter 8.

---

5For criticism of these putative shortcomings see Patricia Kitcher (1984). She argues that no current psychological, biological or neuroscientific accounts exist of the phenomena Churchland cites. If intentional psychology is grossly lacking in these areas then so are many other more supposedly respectable disciplines. If there an argument against the usefulness of a theory to be had in Churchland's examples then his beloved neurosciences will also be caught by its net.
2.2 Singularity of Explanation

The crucial point about the eliminativists who reject the consistency thesis is that they do so because they believe in the *singularity of explanation* thesis. Adherents of this thesis claim that intentional psychology and The-One-True-Cognitive-Psychology pursue the same explanatory ends; that their *explananda* are the same. On this view folk psychology is literally taken to be proto-cognitive psychology.

Since the eliminativists also believe that The-One-True-Cognitive-Psychology will not quantify over intentional states, then they conclude that there are no intentional states. If there are no intentional states then, assuming the singularity of explanation thesis, intentional psychology cannot be consistent with The-One-True-Cognitive-Psychology.

Some of our theorists who accept the consistency thesis also accept the singularity of explanation thesis. We'll now take a closer look at those who buy into the consistency thesis.

2.3 Reducibility

The reducibility thesis is simply the thesis that The-One-True-Cognitive-Psychology will quantify over the kinds which reduce the kinds postulated by intentional psychology. On Fodor's view of the reducibility thesis, it will turn out that The-One-True-Cognitive-Psychology will be: "a respectable science whose ontology explicitly acknowledges states that exhibit the sorts of properties that common sense attributes to the attitudes" (Fodor 1987, p. 10).

The eliminativists reject the consistency thesis due to their acceptance of the singularity of explanation thesis and their rejection of the reducibility thesis. Those who accept the reducibility thesis are the likes of Fodor and Lycan. Fodor, in particular, accepts the consistency thesis because he also believes the singularity of explanation thesis. It is in accepting these theses that Fodor becomes committed to intentional realism. More on this later.

Our other theorists who accept the consistency thesis reject the reducibility thesis. These include the likes of Dennett, and Jackson and Pettit. They can accept the consistency thesis and reject the reducibility thesis because they deny the singularity of explanation thesis. Why will we still be utilising intentional psychological idioms in the explanation of agents' actions when The-One-True-Cognitive-Psychology is complete? Because intentional psychology is really a different explanatory enterprise from The-One-True-Cognitive-Psychology. In accordance with Dennett's *intentional systems theory* (1987), for instance, we adopt what he calls the *intentional*...
stance towards creatures complex enough to be deserving of intentional (intentional psychological) explanations of their behaviour (1978 Ch. 1). In adopting the intentional stance an agent is treated as a black box. How the behaviours are produced is irrelevant to the explanation since the agent must be treated as an agent on a personal (rather than subpersonal) level. On the other hand, what The-One-True-Cognitive-Psychology is interested in is what Dennett calls the design stance, in which one examines the inner workings of the black box. To this extent, The-One-True-Cognitive-Psychology is a subpersonal enterprise, and so cannot possibly hope to treat the agent qua agent. Dennett's claim is that intentional states are like centres of gravity and the Equator. Following Reichenbach he calls the referents of theoretical terms such as centres of gravity and intentional states abstracta—abstract, calculation bound logical constructs—as opposed to illata—actualised theoretical entities (1987 p. 53). In the next chapter I will employ a similar distinction in order to detail various ways of attributing states to systems such as intentional systems.

On the view such as Dennett's, the mistake of both the intentional realists and eliminativists is that they take intentional states not to be abstracta but illata. If intentional states really are abstracta then it is a mistake to go looking inside an agent to find some sort of structure as if they were illata. Hearts are internal structures that might have been pre-anatomical theoretical entities used to explain pulses. It might have turned out that there were no hearts. Maybe the blood was pumped through the body by veins in the same way that the oesophagus and intestines move material. In order for there to be a pulse, some structure is needed to be postulated in order to generate the pulse. In the case of centres of gravity, the inference to some internal structure is required only to the extent that the objects have a mass. Depending upon the shape of the object and the distribution of gravitational forces, the centre of gravity will vary. The fact that there is some structure present in objects with centres of gravity is relevant only to the extent that it affects the distribution of mass within the object. There is no requirement that there has to be any type of structure present; organisms, gold bars, and automobiles all have centres of gravity.

Jackson and Pettit outline what they call programme explanation in their (1988). This style of explanation is contrasted with process explanation. The chief difference between these styles of explanation is that process explanation cites causally efficacious properties of events, organisms, or whatever in their explanantia. Programme explanations, on the other hand, do not require the citing of causally efficacious properties, although they do "programme" for such explanations. The example used to introduce the idea
of programme explanation is that of two electrons being acted upon by two independent forces $F_A$ and $F_B$, such that both the electrons accelerate at the same rate. The explanation of this is that the magnitudes of $F_A$ and $F_B$ are the same. While this is, intuitively, an acceptable explanation, the sameness of magnitudes across $F_A$ and $F_B$ is causally irrelevant to the actual acceleration of A. A accelerates because of $F_A$, and not that magnitude's relation to some other force. As Jackson and Pettit put the point: "the equality per se of the forces acting on the electrons does not do any causal work. The work is all done by the individual forces acting on the electrons" (1988 p. 393). And yet, the explanation which appeals to the equality of the forces explains their identical accelerations.

On the current view intentional psychology invokes programme explanations, whereas the The-One-True-Cognitive-Psychology will invoke process explanations. The mistake of those theorists who accept both the reducibility and singularity of explanation theses is that they believe that we need to cite causally efficacious properties in our psychological explanations. Given these views, how the completed and correct scientific story turns out after the intellectual wash is irrelevant to the explanation of actions in terms of propositional attitudes. If the status of intentional psychology is irrelevant to how The-One-True-Cognitive-Psychology turns out then, of course, they are consistent. 6

We may call anyone who accepts all three of the consistency, singularity of explanation, and reducibility theses an intentional realist. Since I am arguing against intentional realism in this work, I have no axe to grind with anyone who denies one or more of these theses. Even though Stich (along with the Churchlands) denies the consistency and reducibility theses, I do have an axe to grind with him. Why will become obvious in chapter 6.

3 Intentional Psychology and Functionalism

Intentional realism, thus far described, is the view that the kinds of intentional psychology are, more or less, going to feature in The-One-True-

---

6A caveat regarding Dennett, Pettit and the reducibility thesis. Of late, in his (1987 Ch. 2), Dennett claims that the reducibility thesis might turn out to be true (1987 pp. 34-5). In which case the kinds postulated by intentional psychology construed on the intentional stance will get realised by some sub-personal kinds of The-One-True-Cognitive-Psychology. Dennett does not think that the thesis is obviously true as does someone like Fodor. But he is willing to concede the possibility of such a realisation. Pettit also allows for the possibility that the kinds postulated by a programme explanation might map onto the kinds postulated by the process explanations of the The-One-True-Cognitive-Psychology. But he claims this eventuality to be very improbable.
Cognitive-Psychology. To that extent, the intentional realist is going to require some account of intentional psychological kinds—she needs an account of the propositional attitudes. So, we may examine how the intentional realist individuates (a) the class of intentional states (propositional attitudes) from other mental states, and (b) the different kinds of intentional states (beliefs from desires) and (c) beliefs that \( p \) from beliefs that \( q \).

### 3.1 Individuating Mental State-Types

Intentional realists employ a philosophical thesis called *functionalism* in order to give an account of mental state types. It is treated in more detail in chapters 3 and 7. For now, all we need to know is that according to the functionalist, generic mental states such as propositional attitudes, emotions and sensations are individuated *relationally*. It is a mental state's actual and counterfactual connections to an agent's inputs, outputs and other mental states—where these connections determine the state's *causal role*—that makes it the mental state that it is. The same type of story is going to be told in order to individuate more fine grained mental state types. So, something being a belief as opposed to a desire is determined by *its* causal role, *its* particular relations between inputs, outputs and other mental states-types. Fodor explicitly takes this tack. Functionalism says that: "psychological kinds are relationally defined; more specifically, it says that what makes something a belief state is certain of its actual and potential causal relations to such other mental particulars as beliefs, perceptions, desires, memories, actions, intentions, and so on" (Fodor 1987 p. 68). And again: "a belief state is by definition one that causally interacts with desires and actions in the way that your favourite decision theory specifies; and that causally interacts with memories and percepts in the way that your favourite inductive logic specifies; and so forth" (Fodor 1987 p. 69). And ditto for other intentional states such as desire.

Other non-intentional mental state types such as pain, for example, will have some other causal roles set aside for their individuation. The functionalist would appear to be able to supply both (a) and (b) with his functional definitions. The traditional problem with this style of definition,

---

7The acceptability of this characterisation given by Fodor depends upon the status one attributes to fields such as decision theory and inductive logic. One might think that these enterprises are *normative* in nature, and are ideals which agents tend to approximate. Functionalism, as a doctrine in the philosophy of mind/psychology, makes no such normative claims. I think Fodor thinks these fields are non-normative, in that they describe the actual working of an agent's psychological history. We should, at least, assume this reading to be correct, in order to avoid the normativity issue.
though, is its unashamed circularity. However, that problem can be disposed of with the help of the Lewis method for dealing with functional theories in terms of Ramsey sentences (Loar 1981 Ch. 3 and Lewis 1972).

In section 1, I said that propositional attitudes involved relational properties of agents, in particular, to possess a propositional attitude is to bear some relation to a proposition which represents some state of affairs. But what does that mean? If you’re inclined to intentional psychology and functionalism then you are going to have to detail this relation. Two options are usually accepted within the functionalist-intentional realist camp: intentional states may be (i) relations directly between an agent and some state of affairs, or they be (ii) relations between organisms, states of affairs and mental representations. First (i) then (ii).

3.2 Monadic Propositional Attitudes

The classical account of (i) is Loar (1981). The story goes something like this. Mental states are type-individuated by reference to their actual and potential causal roles. Suppose we were to map out the potential psychological history of an organism such that each mental state features as a node in a network of causal interrelations. The actual psychological history would be a path through the network. Corresponding to this network of causal interactions is a network of inferential relations among propositions. The idea, then, is that mental states take as their propositional objects certain propositions in the inferential network. This presumably is non-arbitrary since there are isomorphisms between the two networks. For consider the propositions that (a) Reagan believes in astrology, (b) if Reagan believes in astrology then he will fail to make correct presidential decisions, and (c) Reagan will fail to make correct presidential decisions. These three propositions are inferentially related: the first two imply the latter. But an organism’s mental state of believing both (a) and (b) tends to cause (ceteris paribus) the organism to believe (c). On this view, then, the causal role of the mental states mirrors the inferential role of the propositions which they take as their objects.8

Two points about this story. Firstly, this view takes propositional attitudes to be monadic functional states. What this means is that since the attitudes are functional states and we can determine the causal network, we know just about all there is to know about that mental state. Having a propositional attitude is just to have a node featuring in the causal network. Mental states are monadic, on this view, because the nodes are unitary: they

---

8 I draw this way of describing monadic functionalism largely from Fodor (1985).
have no structure. Propositional attitudes are still relations, in the sense that
to possess a node in the network, is to be related to a state of affairs, although
the mental state itself is monadic.

The second point is that those who take propositional attitudes to be
monadic functional states tend to be functional-role semanticists. The story
given above obviously assigns contents to an organism's mental states on the
basis of functional-causal role. Fodor points out that the connection between
the claimed monadicity of the attitudes and functional-role semantics is not a
necessary one; it's quite possible to hold that the attitudes are monadic
functional states but nevertheless semantically evaluated on the basis of
something other than functional role (denotational semantics, say).

3.3 Mental Representations

Now to (ii). The issue divides into two: the Representational Theory of the
Mind and the Language of Thought.

3.3.1 RTM

If you are an intentional realist you might think that in order to bear a
relation to a proposition, an organism must possess some further psycho-
logical relation to a psychological entity which stands for the proposition. In
that way, the functionally defined psychological state itself would not be
monadic, but dyadic: to possess a propositional attitude is to be in a
psychological state composed out of a relation to something representing the
state of affairs to which one is related. On this view, to believe and desire is to
be related in some way to a mental representation. The most famous
proponent of this view is Fodor. He says:

What I'm selling is the Representational Theory of Mind (hence RTM).
At the heart of the theory is the postulation of a language of thought:
an infinite set of "mental representations" which function both as the
immediate objects of propositional attitudes and as the domains of
mental processes. More precisely, RTM is the conjunction of the
following two claims:

Claim 1 (the nature of propositional attitudes):
For any organism \( O \), and any attitude \( A \) toward the proposition \( P \),
there is a ("computational"/"functional") relation \( R \) and a mental
representation \( MP \) such that
MP means that $P$, and

\[ O \text{ has } A \text{ iff } O \text{ bears } R \text{ to MP...} \]

To believe that such and such is to have a mental symbol that means such and such tokened in your head in a certain way; it's to have such a token "in your belief box..."

Claim 2 (the nature of mental processes):
Mental processes are causal sequences of tokenings of mental representations. (Fodor 1987 pp. 16-17)\(^9\)

Notice here that what Claim 1 attempts to do is explicate the "structure" of the mental states which we describe as the propositional attitudes. Just as in the case of propositional attitudes being monadic functional states, having a propositional attitude is bearing a relation to some state of affairs. On this view, though, unlike the previous version of functionalism about the attitudes, the mental state has a structure. In order to have the PA one must also bear some relation to a mental representation. This is the reason why this account, unlike the previous account, is truly relational. So, the realist with RTM leanings is committed to the inclusion of the objects of the relations mentioned in the definition of propositional attitudes in the ontology of intentional psychology, where those objects are mental representations.

What differentiates believing from desiring (and from remembering etc.), on this account, is the different relations one can bear to MP. Think of these different relations as being realised by different boxes in one's head. A mental state counts as a belief if there is a tokening of some representation in the belief box. Similarly, a mental state counts as a desire just in case a mental representation gets tokened in the desire box. Of course, these boxes are "individuated" functionally.

Fodor concedes that the biconditional in Claim 1 is too strong. There are cases in which propositional attitudes are attributed in the absence of the relation between the organism and representation, and also cases when the relation is present but the attitude is not attributed.

Examples of the first type of case are inexplicitly represented attitudes. Consider Dennett's chess playing computer which acts according to the principle "get your queen out early" (Dennett 1981 p. 107). It's not the case that this rule is represented anywhere in the computer's memory, but

\(^9\)This formulation of the view is also given in (Fodor 1985 p. 88).
Chapter 1

The Topography of Intentionality

the machine acts according to this “belief”. We want to attribute this belief to the machine even though there is no tokening of a relation and a representation (Fodor 1987 pp. 21-3).

Examples of the second sort of case are those in which we get relations to representations without propositional attitude attribution. Let’s suppose that homunculism is true: that one’s cognitive economy is made up of homunculi of ever decreasing size and intelligence a la Dennett (1981) and Lycan (1981a). It might well be that at some lower, subpersonal (the term is Dennett’s 1981 pp. 216-20) or subdoxastic (Stich 1978) level there might be homunculi which they are related in the right causally efficacious ways to mental representations, in which case Claim I would be satisfied (Fodor 1987 pp. 23-4). That such states can count as representational highlights the point made in section 1 that intentionality is not solely a property of the propositional attitudes. Intentionality is exhibited by these states in virtue of their representational properties.

In the face of these objections, RTM is saved, according to Fodor, by interpreting Claim I to be a claim about core cases. Just as chemistry does not identify each water sample with H2O, those cases of pure water being core cases, so the RTM would not require every common sense attribution of propositional attitudes to be accompanied by a tokening of some relation to a mental representation, as per Claim I. This leaves one with the problem of determining which attributions constitute the core cases. This gets solved by Claim 2, according to which mental processes are causal sequences of tokenings of mental representations. From this it “follows that tokenings of attitudes must correspond to tokenings of mental representations when they—the attitude tokenings—are episodes in mental processes” (Fodor 1987 pp. 24-5). That is, if one thinks the thought that p then the RTM inspired intentional realist is committed to the tokening of a relation to a mental representation with the content “p”. In a nutshell: If an attitude features as part of a mental process then there had better be a tokening of a mental representation and it’s these tokenings which constitute the core cases.

Remember that in the case described by Dennett, there is no token intentional state represented in the computer’s memory. This question is not one of the machine’s potentially tokening the representation or rule ‘Get your queen out early’. Rather, the point is that the machine already works as if it’s following such a rule. Assuming, for the sake of argument, that there is a distinction to be had between rules and representations or between rules in a system’s memory and an instruction (a rule in the program which the system executes, see Cummins (1986 pp. 121-2)), nor is there any rule in either the machine’s program or hardwiring according to which the machine would
behave in order to get its queen out early. These intentional states seem to be emergent (the sense in which they are emergent will be made clear in the next chapter) states of the machine, in much the same way as all intentional states are in some sense emergent on the Dennett IST view.

The fact that there are such attributions of emergent intentional states is of not much concern to the realist. However, in distinguishing between emergent and real intentional states, the intentional realist is forced to grant that some of the candidate explanations on offer of a system's behaviour are referring to emergent intentional states. In effect, the intentional realist is committed, in principle, to a mixed taxonomy of intentional state attributions. There are the psychologically "real" intentional states quantified over by The-One-True-Cognitive-Psychology, and there are the other states on the model of IST.

This is not to say that the intentional realist quantifies over emergent intentional states. The intentional realist might, having determined which attributions are the core cases, eschew any mention of the emergent states in her theory and explanations. The point is that the realist must be able to at some point distinguish between the intentional attributions with an underlying psychological bite (those which feature in cognitive processes) from the states employed by IST. If there were no such way of distinguishing the core cases, then intentional realism is in trouble (Fodor 1987 p.24).

How might this distinguishing be accomplished? Fodor's motto "No Intentional Causation without Explicit Representation" alluded to above does not get us very far. From the behaviour of Dennett's chess playing computer we cannot decide whether that behaviour is generated by an explicit representation or not. We also do not know whether that behaviour is a result of intentional causation or not. We need some other way of classifying the real from emergent intentional states.

One problem with getting such a distinction is that for any functional description of a cognitive system which accounts for the behaviour of the system by postulating some set of representations plus program instructions, there is some other functional description which postulates some different set of representations and program instructions. In such cases, what count as core cases are going to differ even though the two functional descriptions are input-output equivalent.11

10For a discussion of why these attributions should not really be called intentional states, while at the same time emphasising the importance of inexplicit representation for cognitive theory see Cummins (1986).
11For more on this point, especially with respect to Fodor's enterprise see Matthews (1984). For a more general discussion of the "multiple realisability" of functional specifications see also Stabler (1983) where he applies this idea to the
To see this problem consider an example mentioned by Pylyshyn (1980 p. 122). It turned out that a certain AI “blocks world” vision system seemed to exhibit something similar to the famous Muller-Lyer illusion (the effect in which out of two lines of equal length, one having arrow vertices and the other having fork vertices, the latter appears longer than the former). For this reason the architecture of the vision system might provide a possible account of the illusion. This particular vision system possessed a line recognising procedure which used a scanner with a limited diameter. In order to detect the ends of a line the scanner would scan the line looking for terminating vertices. In the case of an arrow vertex the arrow lines enter the area of scan before the point of intersection at the end of the line, which leads to the procedure’s accumulating evidence for an arrow vertex earlier than in the case of a fork vertex. This is because the lines of the fork are scanned after the end of the line where the intersection of the fork lines occurs. The effect is that the system recognises the end of the line earlier in the case of arrow vertices than in the case of fork vertices.

Does this count as an adequate explanation of the illusion in the case of our human visual system? The answer would be affirmative only if in fact our vision system uses a limited diameter line scanner. Says Pylyshyn:

Thus, whether this particular account of the illusion is classed as a valid serendipitous finding or merely a fortuitous coincidence depends very much on whether the assumption concerning the mechanism, or the architectural property of the detector, can survive empirical scrutiny. (1980 p. 122)

In other words (in Pylyshyn’s terminology) the AI model only counts as a cognitive simulation of our visual system if it is strongly equivalent with the mechanisms and architectural properties we actually possess. The cognitive enquirer will only decide that matter when all the relevant empirical data is in and scrutinised.

What type of data does Pylyshyn have in mind? This is the same question as: under what conditions will we want to claim that a functional architectural model counts as a cognitive simulation (or: when will we attribute strong equivalence to the functional description and our actual psychological representation of grammars. The multiple realisability of functional specifications is crucial for chapter 7 below.

Pylyshyn identifies a weak construal of simulation where all that is requires is input-out equivalence. Thus you can simulate on a computer the motions of the planets and all one is interested in is that the “coordinate values listed on the printout correspond to the ones that will actually be observed under the specified conditions” (1980 p.120). The generation of such values can be had by using any of a number of algorithms. In a cognitive simulation the actual algorithm used is crucial.
architecture)? Firstly, one would have to guarantee that the two architectures are input-output equivalent in the relevant ways. Any model of the human visual system would have not only to exhibit the Muller-Lyer illusion, but also, for example, have to enable the perception of depth both binocularly and monocularly.

Having satisfied ourselves of input-output equivalence, Pylyshyn suggests that measures such as reaction time may provide one possible criterion of strong equivalence. Functional description will differ in run time if they are different in degree of computational complexity where the number of operations carried out by the two systems differ. From such systems we would withhold the attribution of strong equivalence.

In devising a functional model of the human cognitive system, the intentional realist's model will have to be strongly equivalent to our actual system in order to be deemed correct. The core cases, in which intentional states attributed to agents, will be those in which those states feature in the program instructions or data base of that functional model. Any other attributions of intentional states will be deemed IST states. The moral to be learnt by the intentional realist is clear enough. She gets to distinguish core cases from the non-core cases along Pylyshynian lines, and so can maintain the mixed taxonomy of intentional state attributions.

### 3.3.2 LOT

In his formulation of a non-monadic relational account of propositional attitudes, Fodor makes the claim that central to the RTM is the language of thought hypothesis (LOT). Basically, LOT is the thesis that the mental representations themselves, in addition to the mental state, have structure. However, at times it appears that Fodor takes LOT to be the hypothesis that mental states—having constituent structure. He says: "LOT claims that mental states—and not just their propositional objects—typically have constituent structure" (1987 p. 136). But what makes mental states structured is that one is prepared to be a representationalist about propositional attitudes. Instead of thinking that a mental state is a monadic functional state one might believe that the attitudes have parts, viz. they are composed of a relation and a representation. So, the alternative way to construe LOT would be to grant that the attitudes are structured (i.e. being constituted by a relation and a

---

13The next chapter is devoted to input-output analyses of cognitive systems, and there I will say more about how inputs and outputs are deemed to be relevant.

14For a nice description of the various ways in which we perceive depth and how this relates to the modularity of our cognitive system see Davies (Unpublished pp.10-11).
representation) in virtue of the RTM version of intentional realism, but that it is the representations that are structured according to LOT. This construal is in fact supported by the Fodorian text. He says:

For example, it's compatible with the story I told above [viz. the intentional realist story—JF] that what I put in my intentional box when I intend to raise my left hand is a rock; so long as it's a rock that's semantically evaluable. Whereas according to the LOT story, what I put into the intention box has to be something like a sentence; in the present case, it has to be a formula which contains, inter alia, an expression that denotes me and an expression that denotes my left hand.

... If we wanted to be slightly more precise, we could say that the LOT story amounts to the claims that (1) (some) mental formulas have mental formulas as parts; and (2) the parts are 'transportable': the same parts can appear in lots of mental formulas. (1987 p. 137)

What putting a rock in one's intention box amounts to is being a representationalist about propositional attitudes. What the LOT claim will amount to is a further commitment to the representation, which gets put in the box, being structured. Actually, I think Fodor has great trouble being consistent about these issues in the course of his prolific writings on the subject. Sometimes he takes the RTM and LOT stories to be the same; other times he implies that they are distinct.15

3.4 Dividing Functionalism's Functions

Viewed in the light of the previous subsections, intentional realism is deeply in debt to functionalism. Functionalism provides both (i) a functional level of description (more about this in chapter 3) and (ii) criteria of individuation of mental states-types. In providing the intentional realist with (ii), however, functionalism can be construed as either providing an account of the class of folk psychological states, or if it turns out that some other functional level states are required in order to explain cognitive capacities, providing an account of those other kinds. In other words, functional specifications of mental states can provide a reduction of the traditional class of mental states or processes, or provide a class of some other states or processes. Why should we keep these two functions distinct? Because it provides a further way of

15For some further discussion of these issues see Braddon-Mitchell and Fitzpatrick (Forthcoming).
taxonomising the positions regarding the status of intentional psychology. Let us look more closely at these two functions.

3.4.1 Reductive and Explanatory Functions

Why we need to keep these two functions of functionalism separate becomes evident when we cast our minds back to the theorists featuring in the above taxonomy. Obviously Fodor et al. are functionalists in the traditionally accepted reductive mould. I call them *reductive functionalists* since they believe that functionalism generates both a level of explanation and criteria for the individuation of folk psychological mental state-types. However, it turns out that Steve Stich believes that the mental states over which The-One-True-Cognitive-Psychology quantifies are going to be functionally individuated. He remains skeptical most of the time, and agnostic at the best of times, regarding the inclusion of intentional states in The-One-True-Cognitive-Psychology. But nevertheless he acknowledges that The-One-True-Cognitive-Psychology will be developed at a functional level of description. We may call functionalism which accepts only function (i) *explanatory functionalism* since what is important is the level of explanation that functionalism provides and not the likelihood of folk intentional states being defined functionally.

A cursory glance at Stich's (1983) will show that he is an explanatory functionalist. In outlining his *Syntactic Theory of the Mind* (STM), he postulates entities he calls "B-states" and "D-states", which are mapped to a class of abstract syntactic objects. These states are individuated according to their causal roles, and are thus functionally individuated. Their causal roles might be very much like that of beliefs and desires but they cannot be beliefs and desires since B-states and D-states do not have any *semantic* properties—they have no *content*. Hence, they are not beliefs and desires since, presumably, having some content or other is part of what being an intentional state is.

It should be noted that the Churchlands are not normally taken to be functionalists in either sense. Fodor in his (1985) interprets the Churchlands as denying that there is a functional level of explanation, since on their view behavioural science is, or at least ought to be, neuroscience (Fodor 1985 p. 83). These comments will require revision in light of chapters 2 and 3. It is of course true that in his (1981) Paul Churchland decries functionalism for being a theoretical means for making do with bad psychology. However, given the distinction between reductive and explanatory functionalism and the considerations to be adduced in chapters 2 and 3, the Churchlands can
maintain a modified position. There may well be a functional level of explanation which falls under the rubrik of the neurosciences. Talk of a neurofunctional level (Churchland 1986 Ch.9) of description commits one to this very position. If neuroscience is a functional level of description, then the Churchlands can be described as explanatory functionalists since neurofunctional states may well be mental states (whether they are folk psychological states is another matter).

Dennett can also be described as an explanatory functionalist. Remember that he does not think that functionalism generates reductions of the intentional to the functional, but he does admit to the design stance being relevant to a sub-personal cognitive psychology such as The-One-True-Cognitive-Psychology. On the design stance we are interested not in the way in which a system gets realised, but in the functional components (homunculi) which interact in cognitive processes. That's a functional level of description, but a level which says nothing about the probability of there being functionally defined intentional states.

3.4.2 Functionalism and Semantics

There is a third function to which functionalism is often put. In section 3.1 above, I gave an all too brief rendition as to how (b) might be accounted for, i.e. how intentional state types were to be individuated. The intentional state types to be individuated were to be the class of propositional attitudes—states such as beliefs, desires, intentions, rememberings, etc.. That turns out to be a fairly coarse grained taxonomy of mental state types. Ultimately, a finer grain of individuation is going to be necessary, one where the belief that \( p \) gets to be distinguished from the belief that \( q \) and the desire that \( r \) gets distinguished from the desire that \( s \). In other words, the intentional realist will have to provide an account of (c), of intentional state content. It will have to give an account of belief and desire content in addition to an account of believing and desiring.

Our intentional realist who buys into functionalism relies upon functionalism to give an account of believing. In addition, the intentional realist might also invoke functionalism in order to determine the semantic features of intentional states—the content of beliefs and desires. This invocation is usually called functional role semantics. On this view, the content of a belief will be determined, at least in part, by the relational properties of the belief in question.\(^\text{16}\) The relations of a belief to other beliefs

\(^{16}\)Functional role will be only a part determinant of content if you think that causal connections to the world also play a role in the determination of content.
usually called its *epistemic liaisons*. It's the position of a belief in the causal network constituted by the intentional states that determines the content of that belief. While there is a fair amount agreement regarding this general description of functional role semantics, there is less agreement on the actual details of the idea. For Fodor, for instance, because there is the isomorphism between the semantic-cum-inferential relations between propositions and the causal relations among mental states, when some proposition implies another, then the isomorphic mental states will cause each other. In other words, because \((P \& Q) \rightarrow P\), if an agent believes that \(P \& Q\) then that belief will tend to cause the agent to believe that \(P\) (Fodor 1985 p. 88). On a view such as Stalnaker's (1984), there is no causal relation between one's believing that \(P \& Q\) and \(P\): believing the former does not cause the agent to believe the latter. Instead there exists a *logical* relationship between the beliefs, such that when the agent believes that \(P \& Q\), she also believes \(P\) just because the former implies the latter. For present purposes one may feel free to choose either of these options.

The point about functional role semantics is that the intentional realist who is a functionalist may opt that semantics or she may not. The intentional realist may want function to determine believing or beliefhood independent of the determination of the content of any of those intentional states. Fodor, who makes the points of the previous paragraph, encourages this separation of the functions of functionalism, since he believes that functional role semantics leaves the door open to meaning holism (1987 ch. 3), and it's meaning holism which will cause the intentional realist problems when eventually confronting the problem of the determination of content.

Why might causal role not be relevant to the determination of content? One reason is suggested by Fodor. We really want an account of the content of intentional states that will *survive* variations in functional-causal roles. For consider Paul's hope that Reagan nukes Iran (he thinks Hollywood copyrights are sacred) and David's dread that Reagan nukes Iran (he thinks attacking the Iranians will spark World War Three). Given that they both come to believe that Reagan *will* nuke Iran, that belief will cause elation in Paul and panic in David. Presumably the contents of both Paul's hope and David's dread are the same. But, Fodor claims, the *causal roles* for these intentional states are different. So, it's hard to see what causal role has to do with content (1987 pp. 70-1).

If this is right then the functionalist would be crazy to opt for functional-role semantics. Any account of the content of propositional

---

*Accounts of content which take seriously such causal connections in the determination of content are often called “two factor” theories.*
attitudes which claims that differences in attitude set implies difference in content will have to be wrong, given that the explanations in terms of intentional states are supposed to generalise across differences in attitude set. However, I think Fodor's dread of meaning holism has led him to a somewhat hasty conclusion. The reason why content and causal roles seem to come apart in the Fodorian example is that it is only the actual causal consequences of Paul's and David's belief that Reagan will Nuke Iran which differ. However, it is not clear that in giving a functional-role semantical story one needs to be restricted to actual causal roles in this way. For consider the causal roles which are constitutive of mental state-typehood. It is central to the functionalist's manifesto that one counts actual and potential causal roles in the individuation of mental state-types. So, the same should hold for functional-role semantics. It is the actual and potential causal roles which should determine the content of intentional states. This being so, we see that both Paul's and David's belief will come out with the same content, even though they have different actual causal roles. If there had different other anterior mental states regarding the prospect of Reagan's leaning on the button, then Paul would panic as well as David. That is because the counterfactual causal roles would have been the same. The trouble with thinking about meaning holism in the psychological as opposed to language case, is that one wants to restrict the set of entities which are the partial determinants of meaning to individual psychological histories. It's not clear, however, that in the language case, the set linguistic items which are normally thought to contribute the meaning of an individual item must be restricted to each language user. If that set is taken to range over speakers in a linguistic community, or perhaps, even over the entire species, then the differences in belief sets of an individual should pose no problems.

There is another reason why Fodor should not be happy with his argument. Fodor tells us that he does not believe in intentional causation. It is not the intentional properties of mental states which do the causing, but certain causal properties they possess. However, claims Fodor, the causal relations amongst intentional states "typically contrive to respect their relations of content" (1987 p. 12). Now if Fodor's argument goes through, it's hard to see how the claimed parallelism between the causal and intentional relations between the attitudes can be maintained, as Fodor requires.

The question of the semantics of intentional states is an important one, and it is certain semantical problems which are thought to be the greatest hurdles for intentional realism to overcome. We shall return to these issues in chapter 8.
4 Intentional States

Intentional realism has been characterised thus far as the view that intention states such as beliefs and desires will find home in The-One-True-Cognitive-Psychology. One is an intentional realist if one adopts all of the consistency, reducibility and singularity of explanation theses. If one accepts these theses, then one is able to help oneself to either guise of functionalism presented above. We may now turn to the explication of the doctrine of intentional realism given by Jerry Fodor. On his view, intentional realism “postulates states (entities, events, whatever) satisfying the following conditions:

(i) They are semantically evaluable.

(ii) They have causal powers.

(iii) The implicit generalisations of commonsense belief/desire psychology are largely true of them.” (Fodor 1987, p. 10)

Let’s take a look at each of these conditions in turn.

4.1 Semantic Evaluability

By “semantically evaluable”, is meant that propositional attitudes are true or false (in the case of beliefs), frustrated or fulfilled (in the case of desires), or justified or silly (in the case of fears). The attitudes are semantically evaluable because they have content. Roughly, the content of a propositional attitude is that which makes an attitude, say a belief, the belief that it is. So, the content of the radar operator’s belief about the blip on the screen is that the blip is that of an Iranian F-14. The content of a propositional attitude gets expressed by the declarative sentence following the “that-clause” in the sentential complement of the propositional attitude verb. If you like, the radar operator’s belief “expresses the proposition” that the blip on the screen is that of an Iranian F-14. The operator’s belief turns out true because it expresses a true proposition. That is, because it corresponds to a certain actual state of affairs. The satisfaction of this condition is crucial and we turn to it in detail in chapter 8.
4.2 Causal Powers

In explaining why Reagan ordered the destruction of the Iranian patrol boats I used 'because'. The intentional realist takes this to be a causal 'because'. Propositional attitudes are mental states that (a) can be caused by the environment, (b) can cause other mental states and (c) cause behaviour. These causal powers are thus attributed to the very same mental entities that intentional realism takes to be semantically evaluable. Remember that the causal relations amongst the attitudes parallel their content relations. It was the operator's belief that the blip was that of an F-14 that caused him to believe that the approaching plane might be hostile. The causal powers parallel content relations because the inferential relations between the beliefs are respected by the causal relations.

Fodor thinks that contents parallel causal powers only because he does not believe in intentional causation. That is, “I don’t believe that contents per se determine causal roles ... Technical reason: If thoughts have their causal roles in virtue of their contents per se, then two thoughts with identical contents ought to be identical in their causal roles. And we know that this is wrong; we know that causal roles slice things thinner than contents do” (Fodor 1987 pp. 139-40). We saw in the previous section why this is wrong. There must be causal roles that slice things at the same thickness as contents, otherwise it is difficult to see how there could be the claimed parallelism of causal powers and contents.17

I take it that this condition does not require that all causal relations between mental episodes should have corresponding content relations. There may be many other causal relations deriving from, say association, which get to feature in our mental lives and which are not content relations at all.18 Consider a case in which Smith's thinking about sex always follows from his thinking about gum trees. So, the mental states which possess semantic properties in parallel to certain causal properties are presumably a subclass of all the mental states. This condition is only meant to apply in the cases cognitively caused behaviour.

17Fodor himself says: “The thought that ~p, for example, has the same content as the thought that p on any notion of content that I can imagine defending; the effects of entertaining these thoughts are nevertheless not guaranteed to be the same. Take a mental life in which the thought that p & (p → q) immediately and spontaneously gives rise to the thought that q; there is no guarantee that the thought that p & (p → q) immediately and spontaneously gives rise to the thought that Q in that mental life” (Fodor 1987 p. 140). If Fodor believes this (and I certainly don’t since it seems to me that the content of p and p should not be typed as the same—but that's another story) then by his own example the causal roles/content parallelism is going to break down.

18I owe this point to Kim Sterelny.
4.3 Folk Generalisations

Because the intentional realist is committed to mental causation, she must also postulate generalisations (counterfactual-supporting) in order to get any explanatory bite for intentional psychology. Explanatory bite for intentional psychology would consist of, in part, positing generalisations across (a) an individual agent’s psychological history and (b) different agents. Fodor tells us that he doesn’t have a shopping list of generalisations that intentional psychology must have if it is to be an intentionally realist theory. But something like the following are going to have to be included:

if A sees a red apple in front her, then she will come to believe that there is a red apple in front of her, provided the lighting conditions are adequate, etc.

if A believes that \( p \) and desires that \( q \) and does \( x \), then, \textit{ceteris paribus}, \( p \) and \( q \) are causally sufficient for the production of \( x \).

A would not have done \( x \) if A either failed to believe that \( p \) or failed to desire that \( q \) (Fodor 1985 p. 77).

A’s uttering “\( p \)” is normally caused by her believing that \( p \).

Presumably generalisations similar to those mentioned in section 1 will also feature on the shopping list.

A \textit{caveat} regarding this list of generalisations. In saying that he does not have a shopping list of generalisations required to be honoured for intentional realism to be true, Fodor is very coy. One reason for his coyness is that he thinks a lot of what commonsense tells us about the attitudes must surely be false (1987 p. 15). There are undoubtedly many other things going on in the mind other than what common sense tells us about. So, the generalisations The-One-True-Cognitive-Psychology come up with will have to reflect this. Still, the generalisations mentioned, or some variations thereof—in order to allow for the hitherto undiscovered properties of the mind—will have to feature, if intentional realism is the case.

\[\text{footnote}{He even says that “a lot of what common sense believes about anything must surely be false” (1987 p. 15). If that’s the case, then I fail to understand Fodor’s very own commitment to the ontological proclamations of commonsense, that if there were no beliefs and desires then that would be catastrophic. Maybe commonsense’s ontology is radically mistaken.}\]
4.4 Vindication

Fodor takes intentional realism to be true, and hence intentional psychology to be vindicated just in case that The-One-True-Cognitive-Psychology quantifies over states and processes something like propositional attitudes. States or processes will be something like the propositional attitudes when they satisfy conditions (i) through (iii). So, it might well be the case that the only attitudes which get to feature in The-One-True-Cognitive-Psychology will be beliefs and desires, since a belief set containing only them will obviously satisfy (i) and (ii), and the generalisations required by (iii) would be seem to be stateable in terms of just one state-type representing the world and another state-type generating motivation.

However, if it turns out that the anti-Humean belief-as-desire thesis is correct, then maybe The-One-True-Cognitive-Psychology can make do with quantifying over only beliefs and no other attitudes. In this case, the putative desires-as-beliefs would have to possess the requisite causal roles (according to (ii)) and the generalisations required by (iii) would have to be stateable solely in terms of beliefs.

If either of these possibilities eventuates do we want to claim that intentional psychology is vindicated? I don’t know. But the realist had better have some story as to how many of the attitudes need to be scientifically respectable, in order for the realist to be able to claim that common sense psychology gets vindicated. This is not to say that she must come up with a list now, but rather that the intentional realist must have sorted out the principles by which vindication can be decided, otherwise there could be no decision made as to whether intentional realism is true.

For the purposes of the current work, I am willing to grant that in order for there to be a scientific psychology which vindicates the common-sense framework, only one propositional attitude type such as belief (augmented by desires-as-beliefs) need be quantified over. The first reason why I grant this is that the one attitude type will satisfy conditions (i) through (iii) if the belief-as-desire thesis is tenable. The second reason is that the arguments I adduce below will be unaffected by the number of attitude types. I want to claim that there is not going to be a theory of cognitive implementation that possesses the requisite properties for there being even a single attitude type quantified over.

\[20\text{The problem this poses for the intentional realist is discussed in Huw Price (1988). For more general information on the belief-as-desire thesis see Lewis (1988) and Smith (1987).}\]
Signpost

Where have we arrived at in this chapter? I hope we have demarcated various positions with respect to the relationship between intentional psychology and The-One-True-Cognitive-Psychology. We have also identified the major properties of intentional realism. Whether intentional realism is true is an empirical question. I think there are certain other properties of intentional realism, which as well as being important in themselves to identify, might lead to a means by which confirmation or disconfirmation of intentional realism might come about. It is to these properties we now turn, in Parts I and II. In Part III we examine some objections to the intentional realist programme.
Herbert Simon once defined a complex system simply as something which is made up of a large number of parts that interact, such that the whole is more than the mere sum of the parts (1981 p. 195). According to Simon, a crucial property of complex systems is their hierarchical nature; a hierarchic system is one which is composed of interrelated subsystems, with those subsystems in turn being composed of subsystems, with some elementary subsystem featuring as the basic constituent (1981 p. 196). This account allows for the class of complex, and indeed, hierarchic systems to be vast: atoms, molecules, rocks, amoebas, human beings and the Australian economy all count as complex hierarchic systems.

Now while we can for the most part agree with Simon's approach to complexity, in which it is the decomposition of a complex system by which we attempt to explain the interesting properties of the system, there would seem to be a nondecompositional alternative to attempting to explain the properties such a system. On such an alternative account, a complex system may be treated as a kind of black box. What states and processes we wish to attribute to the system we can determine from the behavioural or functional properties (or lack thereof) of that system. Even if we don't perform hierarchical decomposition upon the system, we know that it must be complex—the fact that it is macroscopic, say, will ensure that it is. So, we have two ways we can go in explaining the properties of a complex system, by attributing states and processes of the system as a result of decomposition, and as a result of input-output analysis. Examples of this alternative analysis of the capacities of complex systems might be the classical dispositional properties attributed to such systems. A glass exhibits certain behavioural regularities (it breaks when dropped) and we attribute a property to it on the basis of that input-output regularity (it has the property of being fragile).
In this chapter I want to take a closer look at these alternative approaches to explaining certain properties generated by systemic complexity. More particularly, I want to draw out some of the differences between these two approaches that seem relevant to cognitive theorising—cognitive systems being complex systems, and as I hope to show later hierarchical systems. So, in section 1 we compare and contrast the two approaches. In section 2 we take a quick look at a certain confusion that can arise, especially for cognitive theory, if one does not follow the distinctions made in section 1. We may then explore some of the intricacies of the decompositional approach—especially with respect to functional analysis vs what we may later call microanalysis—with an eye to seeing how they can relate to cognitive theory later on. Finally in section 4 we can take a quick look at why there is complexity at all.

1 The Two Levels

In this section I want to outline the two approaches to analysing complex systems. First we can look at the basis of the distinction and then see how that can apply to cognitive theory.

1.1 Level One and Level Two

I am going to call an analysis of a complex system a Level One or I-O analysis just in case it attempts to explain the capacities of a system by postulating states and processes defined with respect to the inputs and outputs of the system. In the most general terms possible, what count as inputs and outputs for a complex system are going to be whatever conditions produce state transitions in the system, in the case of inputs, with those state transitions themselves, being the outputs. Inputs and outputs might be anything from the lowering of temperature and solidification, respectively, (in the case of a sample of water) to a question and a verbal response, respectively, (in the case of an adult human being). Many different input-outputs combinations can be applied to a complex system depending upon which of the properties of that system are of interest. One might burn a human corpse and note the difference in weight loss, but that won’t tell one much about how its limbs moved, whereas an electrical stimulation to that corpse might well prove useful.

Although one might have some preconceived ideas as to which outputs should go with which inputs, there would seem to be no specific input-output
pairs. Suppose one thought that the important output from the combustion of some complex system was the smoke given off. From such a Level One analysis one might infer that there is some substance (phlogiston, say) given off in all instances of combustion, that might explain the process of combustion. In fact, though, it turns out that there is some other more relevant output, viz. the increased weight of the system after combustion. So the initial inference would seem to be wrong.

The combustion example shows us that Level One analyses are often used to make judgements about the status of the system construed not as a black box. An analysis which decomposes the black box into subsystems, say, by opening the system or putting it under a microscope, we may call a Level Two or decompositional analysis. Although Level One analyses often provide evidence for certain facts about a Level Two analysis, the inference regarding there being some substance given off during combustion is actually part of a Level Two analysis. That's because that substance involved (phlogiston or oxygen) is a subsystem, or one of the components of, the complex system. We can generalise this idea to claim that the attribution of states to a complex system is relative to the Level of analysis. There are many states of a complex system we may attribute to that system which make reference to the subsystems and components of that system. States may be construed as properties of complex systems enduring through time.

Although Level One analyses cannot attribute states to a complex system which refer to subsystems or components of that system, there may still be Level One state attributions. Indeed, there must be such Level One state attributions if there are to be the state transitions constitutive of the input-output relations mentioned above. Such Level One state attributions might well be centres of gravity or dispositions generally. The point about states attributed in Level One analyses is that the properties corresponding to these states are in a sense holistic or global, or maybe even emergent, in the non-pejorative sense of the term. If one postulates a Level One state of a system, then it would simply be a mistake to try to identify that state at Level Two; looking for a body's centre of gravity by cutting it open would be such a mistake. I call such mistakes Level mistakes, and they are the subject of section 2.

There are some important points to note in closing this section. Level One state attributions might well describe the states as being “inner” states.

---

1. A Level One analysis will rule out some Level Two analyses but not others. A specification of the inputs and outputs of a system will exclude those Level Two theories which generate the wrong outputs, for instance. However, Level One analyses will not differentiate between input-output isomorphic Level Two analyses. To that extent Level One evidence will always underdetermine the correct Level Two analysis.
Chapter 2

Complex Systems 32

of the system. This is not to say that there is some state which one could find by opening up the black box, as it were. The state is inner because it resides within the confines of the system marked off by the boundaries of the black box. Turing machine descriptions might well appeal to inner states in their descriptions, but that is consistent with that description being at Level One.

In decomposing a system according to a Level Two analysis, in all likelihood, the system is going to decompose into kinds featuring in some respectable science. For example, Level Two analyses might postulate the molecules of chemistry, microchips from electronics or cognitive modules from cognitive psychology. This is a significant point which we will return to in section 3.

Level One and Level Two analyses may also iterate. That is, once a complex system is decomposed into subsystems, we are free to perform either a Level One or Level Two analysis on those subsystems. If those subsystems then further divide into subsystems under a Level Two analysis then we are again free to perform either mode of analysis on those further subsystems. Furthermore, the complex system with which we started may also result from some prior Level Two analysis of some larger complex system.

1.2 Levels and Cognition

The previous section pertains to complex systems generally. Since cognitive systems will be a species of complex system, the issues raised apply equally to cognitive systems. Most often, cognitive systems are embedded in some more complex system—computer hardware or an organism, say, (brains-in-a-vat might count as a cognitive system unembedded in this way, but I'm not sure)—where the isolation of the cognitive system is obtained via a Level Two analysis. As we saw in the previous section, there are a variety of possible inputs and outputs of a system that count as the raw data for a Level One analysis. Some possible inputs might be the presentation of some task demand to the system which would yield the performance of that task as output, or perhaps the inputs might be the stimulation of the system's transducers resulting in some motor response. The question as to which inputs and outputs are relevant to the analysis of the system qua cognitive system is crucial. For the moment I will have not much to say on that score; more later in chapters 4 and 7.

---

2 I suppose it's up for grabs whether or not transducers, or indeed any perceptual mechanisms, should count as part of a cognitive system. To exclude them, though, one would be required to be able to distinguish between perception and cognition. I certainly don't have any account to offer in this regard, nor, I'm sure, does anybody else. So, I'm going to include perceptual mechanisms as part of a cognitive system.
Given the range of possible inputs and outputs constitutive of a cognitive system, what kind of analysis of the properties of that system will be possible at Level One? If we construe the outputs of the system in terms of behaviour, then a Level One analysis is going to yield some idea of the capacities of that system. For instance, we know that because of the cognitive system's capacities, the bearer of that system can play chess, recognise faces, use language and negotiate its environment, etc. Moreover, the Level One analysis can give us some limited idea of how these capacities are carried out (see footnote 1), and the relation between these various carryings out: cognitive processes take time to perform, which allow for analysis of reaction time, and the performance of some capacities interfere with the performance of some others. This suggests that Level One analysis can provide at least two benefits: it can enable us to decide which capacities the systems possess, where the capacities represent functions from inputs to outputs. Level one analyses give us the functions of the system that we wish to explain. Secondly, Level One can give us some evidence that may enable the cognitive enquirer to decide between various explanations—where those explanations involve Level Two analyses, as per the previous section. Strictly speaking, these considerations in the case of cognitive systems should also be common to all complex systems.

In presenting the points of the previous paragraph, I should point out that there is a sense in which Level One analyses do not really explain the capacities of the system. Rather, it is Level Two which provides the decomposition of the system into components which we take to be explanatory. The functions individuated at Level One are a complex built out of certain input-output pairs. So, when talking about facial recognition, say, the function is not merely the behavioural act of recognition, but the recognition given that there is a face as input. Since the Level One capacity is a function from inputs to outputs, one cannot explain the capacity by adverting to the input to the system. What one can do though, is predict the potential outputs of the system from the given inputs. If we have postulated some Level One state which represents Bruce's face, then we can predict that a system will perform the act of recognition given Bruce's face as input. To the extent that we might want to explain a given output, a Level One analysis can be explanatory. However, such an explanation is not an explanation of the capacity to recognise faces. Dan Dennett recognises this aspect of Level One analysis all too well. His intentional stance (or intentional systems theory—IST) (Dennett 1978 & 1987) is a species of Level One analysis, but only gives predictive ability rather than explanatory potency. According to Dennett, it is the design or physical stances which yield explanations of the capacities of, at least,
those complex systems which count as cognitive systems. The design and
physical stances are Level Two analyses.

I take it that the sense in which Level One analyses are not more than
merely minimally explanatory is analogous to Moliere’s physician attempting
to explain the tendency of opium to send people to sleep by appeal to its
soporific virtue. As the reader interested in explanation will probably realise,
one of the benefits associated with the current emphasis on the pragmatic
component of explanation (van Fraassen 1980) is that there are some why-
questions to which the Moliere approach can be explanatory. It might have
been thought that the taking of opium merely provides the signal for the
spirits to put one to sleep, rather than it being a low level intrinsic property of
the opium itself that in fact puts the taker to sleep. Or again, the proffered
explanation points to the fact that it is not a peculiarity of a particular
sample of opium or some particular person’s metabolism by which that person
fell asleep; the explanation hints at some covering law or generalisation
regarding sleep and opium (Boden 1981 p. 119).

There are, however, other closely related why-questions for which we,
as sophisticated scientific enquirers, are interested in obtaining answers. We
want an explanation of the soporific tendency of the opium. We get that,
presumably, by examining the microstructural features of the opium, and the
resultant effects of those features upon human physiology. This suggests a
better way of understanding what is going on in terms of the explanatory
potency of Level One and Two analyses: we need to get the explanation of the
Level One functions from the Level Two analysis of the system. The reason
for this is twofold. The first is that the states which we end up with after the
Level Two analysis might be more basic or better understood than those at
Level One. Very often the states and processes postulated by the Level Two
analysis will coincide with kinds quantified over by some respectable science.
Secondly, the Level Two analysis will often postulate simpler capacities that
go to make up the capacities of the system identified at Level One. So, further
Level One analyses are going to be possible of these capacities—this being
what we should expect given that analyses iterate as per the previous section.
This view of Level Two analysis will be taken up in greater detail in section 3.

Now it might well turn out that there is some correspondence between
the states postulated by Level One and Level Two analyses: operating at
Level One we might postulate a face memory or representation, and when we
perform a Level two analysis we find that such a state also features at that
Level; presumably there might be a memory subsystem at Level Two
containing the representation, or perhaps there even is a facial recognition
subsystem. If this turned out to be the case in some instance of explaining the
properties of a complex system, then we would say that some Level One function also features at Level Two. This, however, is no requirement. It might well be the case that the Level One function has no correlate at Level Two; there might well be no subsystem which either recognises faces or stores representations of faces. The Level One capacity of the system might be just a property of the system evident only when the system is construed as a black box—it is a global property of the system.

A couple of points about this representational example before we proceed. Firstly, the question of which Level is appropriate for representational talk is crucial for the current work. That topic is taken up in detail in chapters 4 and 5. Secondly, the question as to when some Level One property features at Level Two is also crucial. I now want to introduce some more terminology. In the cognitive case, I want to say that when we perform a Level Two analysis upon a cognitive system, then the decompositional states we individuate shall be called cognitive modules. There are many senses of 'module' and 'modularity' in the cognitive literature. Although I have not spelled out this sense in any detail yet—that's the task in Part II—I intend this sense in the most general way possible in the context of cognitive theory.

1.3 Levels and Causation

I have claimed that Level One analyses have two functions. Firstly, and most importantly, they specify the capacities of a system; they provide performance specifications. Secondly, they may attribute to that system states and processes which generate limited explanatory power in terms of predicting the behaviour of the system. In addition to being a pragmatic enterprise, explanation also seems inexorably causal: even if an explanation does not explicitly cite a cause, there is some cause to be found somewhere within the domain of the phenomena being explained. One might think, then, that given the limited explanatory potency of Level One analyses, then those analyses will not advert to causally salient properties of the system under analysis. On this line of thought, it will only be Level Two properties of the system which are causally salient. Level One properties are at best analogous to dispositional properties, and their causal credentials are far from clear. For example, Jackson and Pettit might claim that programme explanations feature at Level One in the case of systems to which we attribute beliefs and desires. Such a Level One analysis will merely “programme” for a more

\[^3\text{Dispositions are generally not thought to cause their manifestations. For a discussion see Prior (1985).}\]
detailed explanation in terms of causally efficacious properties—presumably at Level Two.

There are two responses to this line of thought. The first is that in specifying the capacities of a system, a Level One analysis will, depending upon the system involved, specify causally relevant properties of the system. This has to be the case if Level One analyses are to apply in the specification of the capacities of the subsystems of previous Level Two analyses of more complex systems. That is, the iteration of Level One and Level Two analyses ensures that at some point the capacities identified at Level One of some systems will be causally relevant. So, following Jackson and Pettit (1988), if one introduces the distinction between causally relevant and causally efficacious properties, Level One states may not be causally efficacious, but might well be causally relevant.

The second point has to do with the Level One states postulated in order to aid prediction. If they are like dispositional states then they are only going to be as causally efficacious as the best metaphysics tells us that dispositional properties are. If it's true that glasses do not break because they fragile or that propositional attitude ascriptions merely programme for some causal explanation rather than citing causally efficacious properties, then Level One states will not be the causes of the system's outputs.

Okay. What the considerations of this overall section suggest is that unless cognitive theory wants to restrict its explanatory potential, its aim should be to determine which capacities at Level Two generate the Level One functions. Seemingly, only then will cognitive theory get the explanations which it wants. It might well, of course, turn out that no Level One functions get to feature at Level Two. Two possible interpretations of this result seem possible. The first is that cognitive theory turns out to quantify over states and processes far different from those which we might have expected from any Level One analysis; I assume here that pretheoretic psychological generalisations are Level One analyses, since what else have we had to go on. other than behaviour and stimuli? As mentioned in the previous section, Level Two analyses tend to postulate kinds which feature in the hard sciences. The hard sciences (especially neuroscience, say) are relatively recent developments. So, any long standing theory about complex-cum-cognitive systems will presumably be at Level One, simply because there was no Level Two theory to utilise. Secondly, one might insist that such a foreign Level Two analysis, if correct, must not be a cognitive analysis. This dispute is one of the central issues for chapter 4.
Chapter 2  Complex Systems  37

2 Level Mistakes

We saw in 1.1 that when we attribute states to some complex system we must be careful about the Level at which we are attributing the states. We saw that a Level mistake was the postulation of states at Level Two, which are, in fact Level One states. It's important to avoid such mistakes especially given the aim of cognitive theory as introduced in 1.2: if it is cognitive theory's task to determine which functions get performed according to a Level Two analysis, then a Level mistake will trivialise that task. In this section we take a closer look at Level mistakes, with the aid of some philosophy of science literature.

Remember that Level One analyses describe the operation of a complex system in terms of inputs and outputs, and very often that analysis will postulate states defined at that Level. An instance of such analysis evident in the history of science is that of Liebig's attempt to explain animal nutrition. Liebig developed what he thought were the intermediate chemical processes responsible for the nutrition, work, and heat generation of an animal, based solely upon the chemical composition of animal tissue, the chemical analysis of dietary inputs and waste outputs. Without any study of the actual chemical reactions going on within an organism, he tried to "determine the chemical reactions occurring in the body by attempting to calculate what compositional changes must occur in the food to produce an animal's waste products and liberate energy for its work and heat" (Bechtel 1982 p. 561). It turns out that Liebig's calculations did in fact tally quite well with the chemical inputs and outputs of the organismic system; but despite this, his model failed to capture the actual processes of metabolism, with actual processes being far more complex than Liebig's model allowed. It was only after analysing metabolic process by "opening up" the black box, as it were, that the real model of metabolic processes were discovered, and that is a Level Two analysis.

I have claimed that Liebig's analysis is at Level One. On the contrary, one might think that Liebig's analysis was, in fact, a Level Two analysis, but a Level Two analysis that went horribly wrong. I'm sure that there must be many Level Two analyses of complex systems that do go horribly wrong. But in the Liebig case, his analysis was based upon the inputs and outputs of the system. It is for this reason that we must classify it as a Level One analysis. To be sure, Liebig might have mistakenly thought that he was analysing the actual Level Two processes, but that is a mistake about the kind of analysis

---

4 Liebig's attempt is recounted in Bechtel (1982).
5 I'm not going to recount the reasons why the model was ultimately rejected here. I refer the reader to Bechtel (1982).
he was performing, and not a mistaken Level Two analysis. There is, though, an important point about evidence lurking here. Even though, Liebig's analysis is at Level One, the data collected at that Level may well have some evidential or confirmational role to play with respect to some other Level Two analysis. If the input-output correlations are of the wrong kind, then a proffered Level Two analysis will fail. Plausibly, the case of phlogiston is an instance of this: the change in weight of a substance (a Level One property of the substance) led to a Level Two theory (the phlogiston theory of combustion) being rejected. Similarly, a Level One analysis might postulate states, which require some particular type of Level Two state. In the absence of these Level Two states, the Level One analysis will have to be revised. For example, suppose (in fact I argue this below in chapter 7) that folk psychological states are Level One states of us. If it turns out that under a Level Two neurophysiological analysis of us, our brains are mere radio transmitters connected to Martians who control our actions, then we might want to claim that we do not truly possess beliefs and desires.\(^6\)

In his discussion, Bechtel (1982) calls the kind of mistake made by Liebig a *vacuous functional analysis* and that is just what I take a Level mistake to be. A vacuous functional analysis is one which "does account for the inputs and outputs of a system but does not employ the same set of functions to produce this output as does the natural system" (1982 p. 549).

Having pointed out Liebig's error, Bechtel then goes to diagnose the potential for making such a mistake on the part of functionalist inspired cognitive psychologists. He claims that methodologically, functionalists differ little from the behaviourists in that they attempt to account for how sensory stimuli lead to behaviour. As Bechtel himself admits though, the functionalist allows for intermediate informational processing variables to play a role in the production of behaviour. "But," he says "like the behaviourists, many cognitive psychologists think one can develop a model of cognitive processes without studying the neurophysiological processes involved" (1982 p. 564).

Bechtel, correctly, assumes that analyses of neurophysiological processes of systems such as us are at Level Two. It is for that reason that cognitive psychologists who fail to study neurophysiological processes might well make a Level mistake, in an analogous fashion to Liebig. Failing to do this might mean that even though cognitive psychologists take themselves to be offering a Level Two analysis, as did Liebig, the states and processes they attribute might really be mere Level One postulations. Indeed, the intermediate informational processing variables alluded to by functionalist cognitive

---

\(^6\) I think there is an important principle underlying the attribution of intentional states evident in this example. I introduce what I call the *principle of agency* in chapter 7.
psychologists might be attributed at Level One (I will have more to say about functionalism and Level One in chapter 7).

Bechtel introduces the issue of the relation between the neurosciences and cognitive psychology because he wants to present arguments against what I call below in Chapter 3 the *autonomy of psychology thesis*. Now this move I applaud; I argue against various versions of that thesis in Chapter 3. If avoidance of Level mistakes can be used as an argument against that thesis, then all the better for the arguments to come in that chapter.

However, that thesis is relevant to the issue of Level mistakes only to the extent that cognitive psychology is a committed Level Two enterprise. In 1.2 I suggested that The-One-True-Cognitive-Psychology should be construed as a Level Two enterprise. However, Bechtel gives the impression that he believes that cognitive psychology is a Level One enterprise. He says that the research programme of the cognitive psychologist does not fundamentally differ from that of the behaviourist (1982 pp. 563-4). If any psychological programme is going to count as a Level One programme, I think Behaviourism should. If that is so, then it seems that The-One-True-Cognitive-Psychology might well be a Level One enterprise.

The trouble with this line of thought is that it is not obvious that cognitive psychology is a Level One analysis. Bechtel is right about Fodor (1968 & 1975) and Putnam's (1975c & 1975e) arguing that neurophysiology has little interest for psychology. However, the information processing states their models employ need not be Level One analyses (although they may well be despite the use of 'inner' in describing these states—see section 1). Fodor (with Pylyshyn 1988), in fact, is quite explicit in claiming that cognitive states feature in an organism's internal language of thought, and are realised by, rather than being identical to, neurophysiological states. They say that the mental representations which constitute cognitive states:

> are assumed to correspond to real physical structures in the brain...
> For example, the relation 'part of,' which holds between a relatively simple symbol and a more complex one, is assumed to correspond to some physical relation among brain states. (Fodor and Pylyshyn 1988 p. 13)

Realisation is, in this case, a relation that holds among states postulated by Level Two analyses: it is psychological states which are realised by neuroscientific states, but both types of states are individuated by a Level Two analysis. I will say more about realisation in section 3. But for now, there seems to be ample evidence that those at the theoretical end of
cognitive psychology take their enterprise to be at Level Two since neuroscience is blatantly a Level Two enterprise.

The question as to which Level cognitive psychology operates is extremely important. As will become evident below, especially in chapter 4, the situation regarding the level at which cognitive theorising operates is slightly more complex than the description just given. Although many cognitive theorists commit themselves to Level Two analyses, it is not quite clear that their accounts can be Level Two analyses given some of their claims and in-house disputes. My suspicion is that a lot of contemporary cognitive theorising does in fact make a Level mistake in postulating what they take to be Level Two analyses, unconsciously sliding from Level One analyses. The examples which come most readily to mind are of the cognitive structures postulated by factor analytic and componential approaches to psychological theorising (see Sternberg 1977 & 1982), as well as the cognitive psychologist's scripts. As we shall see in chapter 9, scripts are putative knowledge structures which store information about everyday activities which are supposedly invoked when performing those actions. Of course, if such a slide does occur, intuitions which are plausible from the point of view of Level One often get utilised as support for a cognitive model's Level Two status, with actual data from Level Two being thought irrelevant to such cognitive models. If this does happen then it would be disastrous for cognitive theory. Hence the importance of recognising Level mistakes.

I claimed in the previous section that Level One is less explanatorily potent than Level Two, but yielding of prediction. This can be seen clearly in the Liebig case. He thought that he was offering a Level two analysis, but in fact gave a Level One analysis. His model accurately mapped the relations between input and output, and the formulae used in such mappings could be used for prediction, even though those formulae do not feature in the Level Two analysis which would show how the outputs really are generated from the inputs.

3 Level Two

In the previous sections we’ve looked at what goes to make up a Level One analysis. In this section we are going to look at Level Two analyses in more detail. In the first part we look at two varieties of Level Two analysis. In the second part we look at what happens when a system is continually subdivided into subsystems under analysis.
3.1 Functional vs Microanalysis

Remember that Level Two analyses are those in which the complex system is not analysed as a black box. Such an analysis, then, involves decomposing that black box into parts; in effect, we're opening the box up. How do we go about opening the box? The answer to this question depends very much upon what one wants to explain. I take it that the chief purpose of analyses, generally, is to gain some explanatory bite. And as hinted at above, I think that explanation involves a pragmatic component. So, it turns out that it's possible to perform analyses which fail to satisfy our explanatory interests. For this reason it is important to make the analysis-explanation distinction. If there is a such a distinction to be had, then we had better sort out which why-questions we are asking, and hence what explanation we are demanding before we go ahead with our Level Two analysis.

There are many why-questions that can be asked about any complex system. One might want to know what the system is made of. On the account of complexity offered by Simon, a sample of gold constitutes a complex system; the decomposition of it requires some kind of analysis which identifies its microstructural components. We may call such an analysis a microanalysis. The point about microanalysis is that whatever the microstructural properties and kinds that science deems to be constitutive of the system, they will provide the answer to our why-question. The fact that our sample decomposes into atoms of the same kind with the same atomic number is what is important in this type of analysis. Under Level Two microanalysis, the subsystems into which our complex system decomposes need only be atomic; that is enough for that system to qualify as a bona fide complex system, and to be properly decomposed given our interests.

Now suppose that we have some other sample which under microanalysis turns out to be not gold. However, that fact discovered from microanalysis might not bother us, since it is not what the system is made of...
that is important. Perhaps what we want explained, in this instance, certain phenomenal and macroscopic properties of the sample, such as when it melts, why it won't scratch or bend easily etc. Now the fact that this other sample has the same macroscopic and phenomenal properties as gold need not concern us. Even once we have discovered that it is not gold we might still wonder why it has certain of the properties that it does. Perhaps the properties involved stem from the bonding properties of the constituent molecules or atoms together with the temperature of those molecules or atoms. The fact that it's those particular types of molecule or atom may be of no importance here—although it would have been important if we were doing microanalysis. Although microanalysis can generate interesting results in the case of systems such as lumps of gold, it is not clear that microanalysis is going to provide much interesting information in the case of more complex systems such as organisms. Depending upon which part of an organism one microanalyses the results of the analysis might be different. Systems such as organisms are not uniform in their constitution. Maybe in the case of organisms we would microanalyse the system into cells, say, rather than constituent molecules or atoms. I think something like this is the right strategy to adopt in cases such as organisms. However, I think the difference between it and microanalysis is substantial, and deserves outlining in detail. I am going to call such a Level Two alternative to microanalysis functional analysis.

The reason why I call this alternative Level Two analysis functional analysis is that it is what the subsystems do that is crucial to their individuation rather than what they are made of, as is the case with microanalysis. What happens in functional analysis is that we are asking a different why-question from the microanalysis case; we are not asking why does this sample have the property of falling under some kind or other, but rather, why the sample exhibits some particular range of properties, with these other properties generally being established at Level One. Because of this different interest, the subsystems into which the complex system decomposes according to this Level Two analysis will be different (at least initially, but see below) from the microanalytic case. In this case we want to explain certain capacities of the system from a Level Two perspective. So, the fact that the sample is gold would seem unimportant; we want to know what gold and the other sample have in common that generate the properties commonly exhibited. Functional analysis requires interest in certain subsystems of the complex system according to Simon's account; microanalysis requires interest in some other subsystems.
Even though it is functional analysis in which the decomposer of a complex system is generally interested, not all decompositions will divide a system into subsystems that would coincide with a functional analysis. Consider Simon’s (1981) parable of the two watchmakers:

There once were two watchmakers, named Hora and Tempus, who manufactured very fine watches. Both of them were highly regarded, and the phones in their workshops rang frequently—new customers were constantly calling them. However, Hora prospered, while Tempus became poorer and poorer and finally lost his shop. What was the reason?

The watches the men made consisted of about 1,000 parts each. Tempus has so constructed his that if he had one partly assembled and had to put it down—to answer the phone, say—it immediately fell to pieces and had to be reassembled from the elements. The better the customers liked his watches, the more they phoned him and the more difficult it became for him to find enough uninterrupted time to finish a watch.

The watches that Hora made were no less complex than those of Tempus. But he had designed them so that he could put together subassemblies of about ten pieces each. Ten of these subassemblies, again, could be put together into a larger subassembly; and a system of ten of the latter subassemblies constituted the whole watch. Hence, when Hora had to put down a partly assembled watch to answer the phone, he lost only a small part of his work, and he assembled his watches in only a fraction of the man-hours it took Tempus. (Simon 1981 p. 200)

While it is evident that Hora’s watch decomposes into parts, it is not at all obvious that these parts are functional parts. Functional parts are those out of which the capacities of the system specified by an earlier Level One analysis are constituted. It might well be the case that any number of the subassemblies combine to perform not only one task, but different tasks simultaneously. Or even more extreme, perhaps the subsystems are individuated merely by being at the top, middle and bottom of the box; one could if one wished cut the black box into three pieces as part of a Level two analysis. Unfortunately, there will probably be important functional features of the system that are overlooked by such an analysis. If certain (probably correct) accounts of how the brain works are correct, then dividing it into
subsystems according to neuroanatomical taxonomies will not yield a proper functional analysis of the system.8

What the subsystems postulated by a Level Two functional analysis turn out to be, will vary according to the complex system under scrutiny. In the case of gold all we might need refer to are the inter-atomic bonding characteristics and the heat of the system. In the case of a carburettor, we might need to refer to venturi, plenum chambers and floats. In the cognitive case, we might have to postulate memory, perceptual and motor control systems. As mentioned above, the analyses of complex systems may iterate. So, of any of the subsystems postulated by a Level Two functional or micro-analysis can themselves be subjected to further analyses, either Level One or more Level Two functional or microanalyses. An interesting line of thought to follow concerns what happens when we continue to perform functional analysis upon the various subsystems individuated by prior analyses. This is the subject of the next section.

Before we move to the next section though, a point of clarification about functional analysis. I have been painting a picture of functional analysis in terms of Level Two. Strictly speaking this is not correct; Level One analyses are really also functional analyses for the reason that they are analyses specified in terms of what the system does. The particular functional analysis of a complex system at Level One is a high level functional analysis in terms of inputs and outputs. Nevertheless that can still be a functional analysis. It is for this reason that Bechtel calls Level mistakes vacuous functional analyses; the functional analysis at Level One is mistakenly thought to be a functional analysis at Level Two. The difference between the functional analysis at Levels One and Two is that the former takes only inputs and outputs to be relevant to explaining the properties of the system (it treats the system as a black box), whereas the latter takes states individuated independently of the inputs and outputs to be relevant to the explanation of those capacities (it opens up the black box).

That Level One analyses are functional analyses can be seen in an example from Cummins (1983 Ch. 2). On his view, functional analysis consists in analysing some dispositions or capacities possessed by a system into some less complicated dispositions or capacities (p. 28). According to Cummins, there are two types of functional analysis: interpretive and descriptive analysis. Cummins introduces this distinction with an example of analysing a system of relays in program form. That program might, let us suppose, consist of instructions such as 'CLOSE RELAYS A THROUGH D' or, on the other hand, something like 'BRING DOWN THE NEXT

8Well, not initially anyway. See below for more on this.
SIGNIFICANT DIGIT. Let's further suppose that it is the actual closing of those relays that is the bringing down of the next significant digit in the system under scrutiny. Cummins claims that the former instruction describes what happens in the system, whereas the latter instruction interprets what the system does. Notice that the interpretive analysis attributes a capacity to the system which may be realised by many different descriptions—opening a set of relays might well realise the same capacity. Cummins actually defines interpretive analysis as “explaining a sophisticated capacity whose inputs (precipitating conditions) and outputs (manifestations) are specified via their semantic interpretations” (1983 p. 34). To the extent that interpretive analysis is restricted to inputs, and it specifies a capacity which can be realised by many different descriptive analyses, it is a species of Level One analysis, with the interpretive-descriptive distinction mirroring the Level One-Level Two distinction.

There are a couple of ways this can be exemplified. Perhaps we have some system under scrutiny which performs long division. A Level One analysis of that whole system might postulate a state of the system responsible for the bringing down of the next significant digit. If it turned out that this operation was performed by some relays closing at Level Two then that would be an instance of Level Two states conforming to the states postulated by Level One. Alternatively, one might postulate by a Level Two analysis a state of the system which brings down the next significant digit. Now although that state of the system is identified under a Level Two analysis, it makes the identification by adverting to the capacity of that state. Consequently, it is not the way in which that state performs its function (closing relays A through D, say) that is of interest here; that could only be established by a further Level Two analysis. So, when we individuate states at Level Two it is really their potential Level One status by which we individuate them, since Level One identifies capacities, and those states are individuated with respect to those capacities. This is what we would expect given the account of Levels in section 1.

3.2 Levels of Nature

Because there are complex systems everywhere, it might be useful to group some of those systems together according to our explanatory and descriptive interests. This seems intuitively satisfying since some systems seem to have more in common than others: a memory and a perceptual system and an atom are all complex systems for our current purposes, but given that both token
memory and perceptual systems are composed out of atoms, these two systems would have that point in common, but not in common with an atom.

Generally speaking, subsystems which have a number of properties in common will feature as part of some Level Two analysis; it’s by performing a Level Two functional analysis that we individuate the perceptual and memory systems. So it looks as though by performing a Level Two analysis we are actually grouping subsystems together according to our explanatory interests in performing that analysis. Such a grouping of complex system together we may call an explanatory-descriptive level (with a small \( 'l' \)). Every time we decompose a complex system into further complex subsystems by a Level Two functional analysis we move down to a new explanatory-descriptive level. Consequently, there are going to be many levels of explanation. The reader should keep in mind, though, that although there are many levels, there are only two Levels (with a capital \( 'L' \)). Levels of explanation must be distinguished from Levels of analysis—which is just what we should expect given the distinction between analysis and explanation made in 3.1.

Levels (with a small \( 'l' \)) feature prominently in the recent work of Bill Lycan. Lycan has been preaching the virtues of Level Two analyses of cognitive systems for years, in the guise of homuncular functionalism (Lycan 1981). Where levels enter the story most explicitly is in Lycan’s rejection of the function-structure distinction (1986a & 1987). Acceptance of this distinction commits one to hold that functional analysis does not proceed down through many Level Two analyses. Lycan attempts (and I think successfully) to show that there are in fact many levels of nature, with each of these levels being functionally analysed. Shedding Lycan’s immediate interest in the philosophy of mind, for the moment, one can see a budding account in complexity theory.

Traditionally, functionally analysed complex systems were thought to be of a kind similar to mental states and computational states of computers. Functional analysis was supposed to stop once one had got to states such as these, with these entities being realised by structural or physical states such as neurones, microchips or perhaps even cells. Functionalists in the philosophy of mind certainly seem to have believed in some version of this two-level or absolute function-structure distinction. See the following section for a discussion of this point. But, Lycan objects, these realiser states seem to be functionally individuated entities as well. One of his pet examples is the following: mass of subatomic particles; arrangement of atoms; collection of molecules; piece of very hard stuff; metal strip with articulated flange; mover of tumblers or key (Lycan 1987 p. 43). A key is complex system. Indeed it is a
paradigmatically functionally individuated complex system. Once we decide to functionally analyse that system we arrive at systems which themselves seem to be functional. Collection of molecules, here, is certainly a functional description of the system, since one only requires some relevant collection of molecules, not necessarily that particular collection, in order for there to be the higher level properties of hardness and flangeness.

Once we get down to the level of collections of atoms, however, I'm not so sure that we are still speaking functionally. This level is the same level to which microanalysis takes us. So perhaps that is the level at which function stops. I don't, for present purposes, care which way one goes here. Lycan himself allows for the possibility that a functional description at the atomic level may not be feasible, but functional analysis "persists as far down as could possibly be relevant to psychology (well below neuroanatomy, for example)" (Lycan 1987 p. 45).

We should note two issues which follow from this discussion. The first is that functional analysis is going to provide us with a microanalysis if we perform enough Level Two analyses. That's not to say that we must perform all those Level Two functional analyses in order to perform the microanalysis. If our explanatory interests are such that all the intermediate levels are unimportant to us, then we are free to ignore them and perform a microanalysis.

The second issue should allow us to state more precisely what a level really is. Hitherto, I have described a level as merely a grouping of complex systems according to some pragmatically determined properties they are deemed to have in common. From the Lycan discussion, and the fact that functional analyses will give us the level arrived at by microanalysis, it should be evident that the complex systems which constitute some high level are related to the complex systems at a lower level in some special way. Keys are in some sense constituted by atoms, and cognitive modules would seem to be made up of neurones. For such reasons, keys and modules would seem to depend upon certain facts obtaining at the lower level; some complex systems seem to depend upon some other complex systems. Moreover, it is being dependent upon some other class of systems that is the point in common between the complex systems which we group together to form a level. Following popular usage, we may say that some complex systems supervene on some others; or more particularly, one level supervenes upon another.

Now the relation of dependence is a rather fuzzy one, and only a certain sort of dependence relation is going to be postulated here to introduce the notion of supervenience for complexity theory; Bruce, undoubtedly depends upon me for many things and in many ways but we do not want to
claim that he supervenes on me. Roughly, X supervenes on Y just in case there can be no change in X without change in Y. What this means is that there can be no change in something’s being a key without some change in its microstructure, although microstructural changes are not sufficient for changes to its status as a key. Take the house analogy: taking a brick away won’t make the structure a nonhouse, but take away all the bricks and you will no longer have a house.

One advantage of characterising levels in terms of supervenience is that it will generate a hierarchy of levels. Presumably, there will be some complex systems upon which an awful lot of other complex systems supervene. Such complex systems will be near to or at the bottom of the hierarchy of levels. At the bottom, of course, will be the basic constituents of reality, the base level upon which everything supervenes. The entities appearing at this level cannot be complex systems (I’m also not sure if they can be correctly described as systems either) but will feature as the components of some Level Two analysis nonetheless. Higher level complex systems are going to be economies and organisms, with middle level systems being organs and cells. The reason why they feature at middle levels being that they supervene on atoms, say, but not upon organisms and economies. Recalling Simon’s definition of a complex system offered at the beginning of this chapter, the grouping of levels into a hierarchy in virtue of the supervenience relation is exactly what we should expect.

Before moving to the function-structure distinction in more detail I should mention that in addition to bearing the supervenience relation, levels also bear the relation of realisation. A functional kind, such as a key, is realised by the collection of molecules which make up the key, just as the house is realised by the collection of bricks out of which it is constituted. In the cognitive case, we will also see below, that mental or cognitive states get realised by neurophysiological states, and perhaps, if functionalism is correct, also by states of a computer. When the relation of realisation is invoked to describe the relation between entities of two levels, it is normally because these entities are thought not to be the same type of entity. Realisation is invoked when so-called type identity fails. When some level or the state contained therein is realised by some other level or state, mere token identity is said to hold.

3.2.1 Function and Structure

We saw above that Lycan criticises the function-structure distinction. According to that distinction, camshafts are classified as structural, whereas
valve-lifters and mouse traps get classified as functional. The difference between functional and structural states is that the former are abstract. That is:

> When we identify a certain mousetrap with a certain mechanism, we do not thereby commit ourselves to the possibility of saying in mechanistic terms what all the members of the set of mousetraps have in common. Because it is (roughly) a sufficient condition for being a mousetrap that a mechanism be customarily used in a certain way, there is nothing in principle that requires that a pair of mousetraps have any shared mechanical properties. (Fodor 1968 pp. 115-16)

And again:

> Every mousetrap can be identified with some mechanism, and being a mousetrap can therefore be identical with being a member of some (indefinite) set of possible mechanisms. But enumerating the set is not a way of dispensing with the notion of a mousetrap; that notion is required to say what all the members of the set have in common and, in particular, what credentials would be required to certify a putative new member as belonging to the set. (Fodor 1968 pp. 116-17)

Physically heterogeneous systems qualify as mousetraps because of some abstract property they have in common: they all function so as to kill or catch mice. The relation between functional and structural states is an instance of realisation; each of the physically heterogeneous physical systems realise a certain functional description, hence multiple realisation.

What the likes of Putnam and Fodor mean by the function-structure distinction seems to be that some kinds are functional (valve-lifters) while other kinds are physical or structural (eg. camshafts). Putnam (1975c p. 371), for instance, calls functional states logical states and emphasises that those states get realised by structural states which he takes to be in some sense “physical”. Examples of physical states seem to be anything from vacuum tubes, electronic relays or human clerks sitting at desks. So envisaged, the function-structure distinction seems to be an absolute distinction; some kinds are physical kinds and some other kinds are abstract functional kinds. However, maybe that’s wrong. Maybe there is no single level of description at which something is functional and some other level of description at which something is structural. What gets described as functional or structural might well be relativised to levels of description, so that what Putnam and Fodor take to be structural can, in fact, be described as functional—with
respect to some other level. In other words, perhaps the two-levelism built into the function-structure distinction is mistaken.

What Lycan attempts to do—as does Kalke (1969) in anticipation of him—is show that there are not just two levels of description at work when one performs functional analysis, such as in the case of psychological theorising. Instead, there is a hierarchy of levels in which the possible levels of description form a continuum. The “function-structure” distinction gets relativised: a level of description gets labelled structural only relative to some higher level which we take the lower level to realise. It is this impoverished view about function and levels in psychological theorising that leads Lycan to believe that the function-structure distinction is pernicious.

Some examples seem to cast doubt upon the absolute function-structure distinction. Take Fodor’s mousetrap. There is clearly a level of description at which whatever satisfies the functional specification “mouse input yields dead-or-caught mouse output” counts as a mousetrap. At that level, the classic device consisting of base, bait holder attached to a spring loaded lever to trap the mouse, gets to count as a mousetrap—call it the Acme mousetrap. But at that same level, a device that recognises mice, destroying them with a laser, and Sterelny’s cat both get to count as mousetraps, too. All of these objects preserve the same functional relations in that they perform-the-same-function/get-put-to-the-same-use, viz. satisfy the description “mouse-input-dead-or-caught-mouse-output”. Putnam defines functional isomorphism as the preserving of functional relations (Putnam 1975f). In which case, all three of our candidate mousetraps are functionally isomorphic.

It’s easy to see here why functional descriptions can be multiply instantiated in physical systems, since Sterelny’s cat, the Laser Terminator, and the Acme mousetrap have no physical description in common: one is an organism, another a piece of technology yet to be invented, and the third a handyman’s invention requiring no more than the manipulation of steel or iron.

That two physically heterogeneous objects can be functionally isomorphic raises an interesting question. Consider two examples of the Acme mousetrap, one being made of traditional materials such as a wooden base, and iron spring and lever, the other being made entirely of steel. Both of these mousetraps perform the same function. But there seems to be much more in common between them than our other candidate mousetraps. The question is what that something in common consists in. Both seem to have certain properties in common viz. springs, bait-holder and lever. These properties seem to be states of the mechanisms, which are functionally individuated—what could be more functional than being a bait-holder or
spring? It's because of this internal functional similarity that these mechanisms seem to be performing the same function in the same way. In short, there seems to be a stronger form of isomorphism exhibited here. Pylyshyn calls processes which perform the same function in the same way strongly equivalent, and those which perform the same function specified at some high level such as that of input-output as weakly equivalent (1980 p. 120).

We can describe the sense in which our two Acme mousetraps are strongly isomorphic by distinguishing between descriptions of states of systems or mechanisms in terms of a decompositional Level Two analysis and those described in terms of input-output Level One functional roles, on the other. In other words, strong equivalence amounts to some decompositional Level Two functional isomorphism and weak equivalence amounts to I-O Level One functional isomorphism. This means that in addition to our mousetraps being functionally isomorphic under an I-O Level analysis along the Fodorian lines, there is also a decompositional Level at which some mousetraps count as functionally isomorphic while others do not.

The decompositional Level analysis of a system or mechanism in terms of functional roles would become relevant when explaining our mousetrap's capacities—although their being a mousetrap is determined by an I-O Level analysis—according to Haugeland's “morphological” explanations (Haugeland 1981a). In such explanations we refer to the functionally defined structures such as springs, bases, bait holders, etc., which combine to generate the mousetraps' capacity for catching mice. Only Acme mousetraps, and not Laser Terminator mousetraps, exhibit those structures, though. In the case of our two candidate Acme mousetraps the structures referred to in our explanation are composed of different materials (iron and wood in the one case and steel in the other). But because those structures do the same thing, we have a finer grained level of functional description and isomorphism than in the case of the Level One functional description. Another way to express the fine grainedness of these descriptions is to think of them as being less abstract than Level One functional descriptions. Roughly, a description is more abstract if it allows more objects to fall under that description. Level One functional descriptions, such as “mousetrap” allow more objects to fall under it than the Level Two based descriptions such as “Acme mouse trap”.

So, just as ‘mousetrap’ gets to mark out a kind, indeed a functional kind, ‘Acme mousetrap’ marks out a less abstract functional kind. But remember that it was an Acme mousetrap that realised a mousetrap in the

---

9 Even though it's structures which feature in these descriptions, that doesn't make the description structural since they are functionally individuated structures. For comments on "structural" terminology see Lycan (1987 pp. 47-8).
first place. The supposed structurally described realiser, in fact, is a *functional* realiser. One might want to say that the acme mousetrap is a physical realisation of the functional kind MOUSETRAP; but equally one should say that the steel mousetrap is a realisation of the kind ACME MOUSETRAP. Patently, what we have here are various levels of functional description consistent with the various Levels of analysis being offered in this chapter. If there is such a hierarchy of functional levels, then the kinds that realise any given functional taxonomy are going to be a structural state relative to whatever level of functional description it is realising. I think that Putnam, for instance, still fails to put enough emphasis on the various functional levels of description-explanation, as I hope to show in chapter 7 where I criticise his own recent work in which he attacks functionalism, his very own philosophical offspring. The claim I have been making here is that by failing to appreciate the multiplicity of levels and Levels, one might well be tempted to interpret the function-structure distinction absolutely, rather than relatively.

### 4 Why Complexity?

Our all too brief excursion into complexity theory would not be complete without some mention of why there are complex systems at all. We can cite the following reason.

Simon uses his parable of the two watchmakers to draw attention to the advantages of going complex. By performing some quantitative statistical analysis of the relative difficulty of the tasks facing Tempus and Hora, Simon concludes that it will, on average, take Tempus four thousand times as long to assemble a watch as Hora (1981 pp. 201-2). As Simon says (!): “The time required for the evolution of a complex form from simple elements depends critically on the numbers and distribution of potential intermediate stable forms” (1981 p. 202). We may call this feature of complexity its *production advantage*.

The production advantage of complexity has been noticed by various theorists in diverse fields. Richard Dawkins distinguishes between *single-step* selection and *cumulative selection* in his (1986). In the simplest terms, the probabilities for the emergence of a complex system by single-step selection are extremely low. Similarly, the probabilities are low for a computer program to randomly generate a line from one of Shakespeare's Sonnets. However, if that program is designed in such a way that it attempts to improve upon each successive attempt by, say, comparison to an actual line from a Sonnet, then the probabilities for generating that line are dramatically
improved. It this kind of process which underlies cumulative selection, and which supports the existence of complexity.

The production advantage of complexity has also been noticed in the specific cognitive-perceptual case by David Marr in his principle of modular design. Marr says:

This principle is important because if a process is not designed in this way, a small change in one place has consequences in many other places. As a result, the process as a whole is extremely difficult to debug or to improve, whether by a human designer or in the course of natural evolution, because a small change to improve one part has to be accompanied by many simultaneous, compensatory changes elsewhere. (Marr 1982 p. 102)

Marrtian modules are individuated functionally. So, they are going to be candidates for the subsystems identified by a Level Two analysis of some cognitive-perceptual system. We will return to the question of the modularity of certain complex systems, viz. cognitive systems, below. For now, what the production advantage suggests is that it will be advantageous to examine cognitive systems in terms of their complexity.

**Signpost**

Why begin a work on cognitive architecture with reflections on complexity theory? There are two benefits which, to a certain extent, we have already perceived, and will continue to perceive on our subsequent travels. The first is that because cognitive systems are complex systems, the distinction between Level One and Level Two analyses might give us a handle on some of the issues raised in the cognitive case. The second reason is that the issues from complexity involving the decomposition of complex systems into parts is crucial to an understanding of issues relating to the modularity of cognitive systems, the central issue in Part II.

So, keeping in mind the distinction between Level One and Level Two analyses, and the fact that complex systems divide into subsystems, we may proceed to apply this talk of levels to some theoretical issues to do with cognitive theory. After that, we can apply the resultant theoretical results from Part I to the main concerns of this work.
Chapter 3

Autonomy

In the previous chapter I argued that The-One-True-Cognitive-Psychology ought to be, and is taken to be, by cognitive theorists, a decompositional Level Two enterprise. It is claimed, by theorists such as Fodor (1975), Simon (1981) and Pylyshyn (1984), that it is only at such a level—a cognitive level—of explanation-description that one can capture the generalisations required to explain certain ranges of behaviour, in particular those behaviours mediated by cognitive or thought processes. It is this explanatory requirement that is thought to endow The-One-True-Cognitive-Psychology with an autonomous status. Here are some examples from the literature:

Computer programs describe a sequence of operations performed on input without considering the physical materials that make up the machine ... Psychological descriptions of information processing based upon people's behaviour are analogous to computer programs in that they describe mental operations without considering the anatomy or neural activity of the brain. (Huttenlocher 1973 p. 174)

If it is the organisation of components, and not their physical properties, that largely determines behaviour, and if computers are organised in the image of man, then the computer becomes an obvious device for exploring the consequences of alternative organisational assumptions for human behaviour. Psychology can move forward without awaiting the solutions by neurology of the problems of component design. (Simon 1981 p. 26)

Physical and neurophysiological terms taxonomise the world in ways that do not permit us to express such generalisations. They often distinguish aspects of the world and of behaviour that are equivalent
with respect to their psychological import, and sometimes fail to make distinctions that are psychologically relevant. For this reason, descriptions cast in such terms typically fail to capture important psychological generalisations concerning human behaviour. (Pylyshyn, 1984 p. 17)

This autonomous status is normally described as cognitive psychology’s being *irreducible* to a more basic fundamental level of description such as that of the neurosciences. We may call this idea the *functionalist anti-reduction thesis*, since it has its origin in the functionalist theories of the mind proposed by, classically, Putnam (1975c,d,e,f) and Fodor (1968).

Remember that functionalists thought that the states realising the functional roles which specified mental states are not the same kind of thing as that specified by the functional role. For example, the property of being a valve-lifter is not the same property as being a camshaft; there are a range of possible instantiations of valve-lifters which are not camshafts. Since physical kinds are going to cross cut functional kinds in this way, the functionalist insists that there will be no *type-type* identity between mental states and brain states. Because the kinds at the cognitive level cross cut those at the physical (most often neuroscientific) level, then (i) only states and processes at that level are relevant to cognition—i.e. the details of the physical hardware which realises some functional model are irrelevant to the development of that psychological model; and (ii) the generalisations (or laws) relating the kinds of the psychological level will differ from the generalisations relating physical kinds, again resulting in the physical level’s being irrelevant to the psychological level. It’s this putative irrelevancy which forms the basis of the claim that the cognitive sciences are autonomous from the physical sciences. The sense of ‘relevant’ used here is crucial: it would seem that so-called “bottom-up” constraints upon cognitive theory might not be forthcoming because of this irrelevancy. Since I think there are arguments against intentional realism based upon such constraints, the autonomy issue becomes important for some of the chapters which follow.

In recent times, the functionalist anti-reduction thesis has come under attack, most noticeably by Robert Richardson (1979 & 1982) and Berent Enç (1983). What is characteristic of the Richardson attack, in particular, is the separation of the issues of reduction and explanatory-cum-ontological parsimony. On his view irreducibility amounts to what he calls a *de jure* autonomy. At the same time, Richardson allows that cognitive science and psychology may exhibit what he calls *de facto* autonomy, in the sense that ‘psychology can give a proprietary characterisation of its domain. Of course, anyone committed to its being a science *must* be committed to this much at
least" (Richardson 1979 p. 538). In this way, one can have the neurosciences relevant to The-One-True-Cognitive-Psychology with respect to the issue of reduction, but irrelevant with respect to whether that discipl ought be pursued. That is, one might still be interested in doing cognitive psychology even though it may be reducible to a more basic science such as neurology.

While I think that this splitting of the autonomy and reducibility issues is to a certain extent correct, I want to argue that it does not fractionate the debate along quite the right lines. In fact, contra Richardson, I will argue that reduction, once properly understood, should facilitate explanatory parsimony. As well, the general issue of "independence"—the sense of which will become clear below—would seem to be quite orthogonal to the issue of reduction, a fact which is obscured by both the Richardson and Ené accounts. There are, I think, at least four senses in which a level may be autonomous from another, with reducibility being just one of the four senses. The task of this chapter is to elucidate those forms of autonomy in the hope of determining those senses to which The-One-True-Cognitive-Psychology ought be committed. I will also argue, contra Richardson, that it is under certain conditions only that cognitive psychology would retain its status as a discipline worth pursuing, if it were in principle reducible.

In sections 1 through 4 we look at each of these four senses of autonomy, and determine those to which The-One-True-Cognitive-Psychology should be committed. Section 5 will then be an examination of the relation between The-One-True-Cognitive-Psychology and the neurosciences, given the results of sections 1 through 4.

1 Descriptive Autonomy

The first sense of autonomy is which we may call descriptive autonomy. Put simply, descriptive autonomy consists in a level's providing a level of description-explanation comprising its own taxonomy of kinds, and the particular generalisations relating those kinds. As Richardson puts it, the thesis is "the supposed capacity of [a level such as] psychology to provide in its own terms descriptions of its own phenomena and its domain" (1979 p. 555).1 Because cognitive psychology is conceived of as a functional level of description, these kinds and generalisations will be specified in terms of actual and potential causal roles.

1In actual fact Richardson characterises his de facto sense of autonomy in this way. As we shall see this characterisation would not seem to make a level possessing this form of autonomy explanatorily interesting, and, hence, worth doing.
The way to conceive of descriptive autonomy is this. Pretend that cognitive psychology is complete, and we have specifications of all its kinds and ensuing generalisations. Descriptive autonomy amounts to the claim that the completed science does not have to draw upon the concepts of another science in stating those specifications and generalisations. Now suppose that psychology required reference to be made to some physical kinds from some other descriptive-explanatory level. Does it follow that psychology is not descriptively autonomous? No: just because psychology is complete and these references are required then those kinds must also be psychological kinds, if only because they feature at that psychological level.\(^2\)

By claiming that psychology is descriptively autonomous one is claiming that there is a distinctive descriptive-explanatory level. Just because some of the kinds postulated at that high level turn out to exist at some lower explanatory descriptive level is consistent with there being that higher psychological level. The point is that not all of the kinds featuring therein feature at the lower level.\(^3\) That some kinds feature at different levels of description is a commonplace in science: pressure and volume feature in both thermodynamics and statistical mechanics. The case for psychology and the neurosciences ought not be special in this regard.

In the functionally characterised cognitive case, we can therefore distinguish, in the manner of Richardson and Patricia Kitcher (1980), between a purely functional cognitive science, and one with a mixed taxonomy of kinds, with some being functionally specified, and some being intrinsically specified by reference to some other, more basic level.

There are bound to be many descriptions of kinds which might constitute a level that we fail to find interesting: a descriptive vocabulary might include the Fodorian kinds "is transported to a distance of less than three kilometres from the EifFeI Tower" (Fodor 1975 p. 14) or "is an H-particle [if the coin comes up heads]" (Fodor 1987 p. 33). These, though, may not be very explanatorily interesting or useful. In terms of the previous chapter, we have grouped together some complex systems which don't naturally have very much of interest in common.

This sense of autonomy is, in itself, not very interesting from our current concerns since it is in part constitutive of what being a level is. All

\(^2\)This is a simplified version of an argument offered by Richardson (1979 pp. 538-539).

\(^3\)That there can be kinds featuring at two levels seems unproblematic and consistent with the classical account of reduction offered by Nagel (1961). Bridge laws are intended to relate those kinds of the secondary theory to the kinds of the primary or reducing theory that fail to feature in that secondary theory. Do we describe the kinds which already in the primary theory as being primary or secondary theoretic entities? Presumably, they're both.
explanatory-descriptive levels are going to be autonomous in this sense. Notice, though, that it is this sense of autonomy which Richardson seems to call his *de facto* autonomy. However, contra Richardson, in order for a level to be explanatorily interesting it had better be more than just descriptively autonomous. It is this sense of autonomy which is advocated by the above quotation from Huttenlocher.

2 Methodological Autonomy

Now suppose that we want to restrict our attention to levels of explanation-description which we find *interesting*. This might restrict our attention to a very small number of possible descriptively autonomous levels. What will make these *methodologically* autonomous levels interesting will be explanatory and predictive value. A level will presumably have explanatory and predictive value in virtue of the generalisations which relate the kinds at that level, where we think we require such generalisations in order to generate the explanations of the phenomena in which we are interested.

It is methodologically autonomous levels which get to be called *special sciences* as envisaged by, say, Fodor (1975), Pylyshyn (1981) and Simon (1981). Again, one should assume that there are no epistemic or developmental issues creeping into the picture here; one should pretend that all the levels have been completely characterised and all the information is in. The methodologically autonomous levels are those which we find interesting and useful. If it is this explanatory end which determines something's being a *science*, then being able to give a proprietary characterisation of its domain is not going to be sufficient for a level's being a science, as claimed by Richardson. In order for that to be the case, not only must descriptive autonomy be a component of *de facto* autonomy but also a methodological autonomy must be a component as well: a *de facto* autonomous level would have to be explanatorily interesting as well. However, one can only have *de facto* autonomy in this way by abandoning the separation of the reduction and parsimony issues. The following considerations suggest why.

When one gets down to the basic motivation for interlevel reduction, that motivation seems to be an explanatory one. Assume that we have decided upon which levels are explanatorily interesting, we might find that some levels are related in various ways. More specifically, we might find that one level (what we may call the *primary level*) not only does its explanatory work, but also do the work of some *other* level (what we may call the *secondary level*). In other words that primary level generates all the explanations and predictions originally gained from the secondary level. The
explanations generated by the primary level must range over the domain of
the secondary level and capture the same level of generality as the expla-
ations of the secondary level. It is in such cases, I claim, that we have
discovered that the one level reduces to the other.

The above quotations from Pylyshyn and Simon seem to make a com-
mmtment to the methodological autonomy of psychology; I take it that that's
what Simon means when he says that it is organisation which determines
behaviour. If organisation actually determines behaviour, then we will be
able to explain behaviour only in organisational terms. There is a further
component to the quotation from Simon, however. That will be discussed in
section 3.

Two important issues are raised in our present context. The first has
to do with theories of reduction proper. The preceding remarks are designed
to be constraints which must be met by any detailed theory of reduction. We
should, therefore, take a more detailed look at reduction, particularly given
that we might want to know whether or not The-One-True-Cognitive-Psych-
ology will end up being methodologically autonomous. The second issue has to
do with the dispensibility or otherwise of a level which seems to be reducible
to another level: if a level is reducible, then presumably parsimony will
dictate that we dispense with that secondary level. We look at these issues in
turn.

2.1 Theories of Reduction

You will recall that functionalists claimed that psychology generally will not
be reducible to any branches of the neurosciences because of the failure of
type identities between the potential primary and secondary theories. With
no type identities, it was claimed, there could be no "bridge laws" to link the
primary level to the secondary level. This objection to the methodological
autonomy of psychology rests upon the classical view of reduction given by
Nagel (1961). Nagel placed two conditions upon the secondary level's being
reduced to the primary level. This objection to the methodological
autonomy of psychology rests upon the classical view of reduction given by
Nagel (1961). Nagel placed two conditions upon the secondary level's being
reduced to the primary level. Firstly, the kinds of the secondary level which
are absent from the primary level must be "suitably related" in virtue of the
bridge laws to the kinds in the primary level. Nagel calls this the "condition of
connectability". It's this condition which the functionalists and theorists in
the philosophy of cognitive science exploit in advocating methodological
autonomy.

The second condition Nagel calls the "derivability condition". On this
condition, each law of the secondary level must correspond to a law in the
primary level, such that by using the bridge-laws, the laws of the secondary
level can be derived from those laws of the primary level. When these conditions are fulfilled we may say that the two levels fail to be methodologically autonomous since the higher secondary level reduces to the lower primary level.

The crucial thing to realise about these conditions for methodological autonomy is that whatever they are, they must ensure that the primary level generates all the explanations and predictions originally gained from the secondary level, according to the above explanatory requirements. Nagel goes somewhat further in this explanatory requirement. Not only must the same predictions and explanations be forthcoming from the primary theory, but that primary theory must also explain the secondary theory's ability to generate the explanations and predictions that it does.

In order for the Magellan conception of reduction to apply in the present case, we must assume, firstly, that a level has the properties of a theory which are relevant to reduction, and, secondly, that levels will utilise laws in their explanations and predictions. While the former assumption seems fair enough, since what a level is, is a grouping of complex systems together under some kind of theory about their having certain explanatorily interesting things in common, the latter assumption is not at all obviously the case. It might well turn out that the only real laws are those of the lower level hard sciences, in which case one would require a modified theory of reduction—a modified Nagelian theory, perhaps—which ranged over generalisations rather than laws.

In any event, even assuming that these assumptions are plausible, the functionalist inspired cognitive theoriser's use of the classical Nagelian conception of reduction has been called into question in recent times. Richardson (1979), for instance, challenges the common interpretation of the connectability condition. In rejecting the methodological autonomy of cognitive psychology, they assumed that the bridge laws or, better, inter-theoretic connections had to be bi-conditional in form, such that type identities were required to be made before derivation of the secondary level's laws or generalisation could be made. There is, however, an alternative interpretation of the connectability condition which requires only that the bridge-laws relating the kinds from the two levels be one-way conditionals of the form: if A (a term of the primary level or theory) then P (a term of the secondary level or theory). On such an alternative reading of Nagel, a one-many relation between levels is not sufficient for failure of reduction. Says Nagel: "the linkage between P [a term in the secondary theory] and A [a term in the primary theory] is not necessarily biconditional in form, and may, for example, be only a one-way conditional: If A, then P" (Nagel 1961 p. 355n);
and again: “the state of affairs signified by a certain theoretical expression ‘A’ in the primary science is a sufficient [my emphasis—JF] (or necessary and sufficient) condition for the state of affairs designated by ‘P’” (Nagel 1961 p. 354). So for any generalisation of the level which realises the cognitive scientific level, it may be possible to reduce that higher level to the lower provided that the one-way conditionals are forthcoming, and clearly the functionalist’s proclamations about type-type identity do not preclude that eventuality.

The ideology lying behind this form of reduction is one which takes reduction to be domain relative or specific. The domain relativity of reduction has been advocated by Richardson and Enç, and the Churchlands (Patricia 1986 & Paul 1984). Although I disagree with some of the details, the general idea is that reductions will be possible for a subclass of the class of entities possessing cognitive psychological states. The claim is that whenever members of that class possess some property, usually a realising hardware of a particular kind, then that is sufficient for the attribution of some cognitive psychological property, in accordance with the one-way conditional connections between the hardware level and the cognitive psychological level.

In order for domain specific reduction to be explanatorily viable, one must show that reduction is possible with the presence of mere one-way conditional intertheoretic connections. One must show that the generalisations of some secondary level are derivable using those one way connections. There are arguments in the literature, in particular some offered by Patricia Kitcher (1980), purporting to show that these derivations will not

---

4I have changed Nagel’s terminology here in order to make it consistent throughout the examples in the chapter.

5My description of domain specific reduction is idiosyncratic compared to that featuring in the literature. The standard description allows for bi-conditional bridge laws by advocating type identities of a restricted type such as cognitive-states-for-humans rather than the more general cognitive states. Now that type might be type identical with a neurological type. The restricting, in this way, of the types that one wants to identify is advocated by Jackson, Parfet and Prior (1982). The trouble with this, is that it runs counter to the functionalist’s mistaken—so I claim in chapter 7—hope that a psychology, and in particular a cognitive psychology, will pertain to any creature provided it is organised in the right way. In order to get the requisite generality, this version of domain specific reduction will have to make the claim I am just about to make viz. that the possession of that restricted type cognitive-states-for-humans will be sufficient for the more general type. Doing that, however, loses the type identity between the general type and the neurological type in just the way I am describing. In my description of the domain specificity of reduction I have omitted this type-identifying step.

6I should point here that the potential reduction of The-One-True-Cognitive-Psychology will have to be domain relative, for reasons which will become obvious in chapter 7. The idea will be that The-One-True-Cognitive-Psychology is chauvinist in that it excludes non-neurological state bearing creatures from being functionally equivalent to us. I mention this just to wet the reader’s appetite.
be forthcoming, although I think there are ways to circumvent these arguments. Undoubtedly, the question as to whether The-One-True-Cognitive-Psychology is methodologically autonomous is an important one. However, I think it to be a question too large answer here; it is just too tangential to the concerns of this work. Instead, I want to assume that methodological autonomy fails, in order to determine how that affects the explanatory parsimony and reduction connection, which Richardson claims splits. We move then to the issue of dispensibility.

2.2 Dispensibility

If it turns out that The-One-True-Cognitive-Psychology is methodologically autonomous, then it follows that there is explanatory justification for the discipline—presuming that the explanations and generalisations contained therein are required. We now consider a more pessimistic scenario.

If cognitive psychology (for humans) were to domain specifically reduce to some branch of the neurosciences, that is, the neurosciences can do all the work that cognitive psychology in humans can, then does it follow that we should dispense with that cognitive psychology? The answer to this question has to be "no". One reason might be this: the very domain specificity of the reduction means that the explanations which we know apply in the human case, due to humans' possessing a neurology, will not be applicable to any other possessors of cognitive states which do not possess a neurology. It is this scenario, in terms of domain specific reduction, which allows Richardson to break the connection between parsimony and reduction. We will need cognitive psychology just because we want to capture more creatures in our cognitive net than just humans. One might want to claim that neurology does allow us to capture martians and smart computers in the cognitive net just because the domain specific reduction guarantees that neurological descriptions have functional properties which will apply to creatures without neurones. I think the right response to this is to claim that neurological descriptions endow these properties just because of some psychological rather than neurological story being true. For reasons which will become obvious in chapter 7, I do not like this reason for the negative answer since I think that there will be different One-True-Cognitive-Psychologies across humans, martians and smart computers, contrary to the functionalist inspired cognitive theoriser's hopes. But more on that later.

For a better reason for the negative answer, let us suppose that some level is not methodologically autonomous with respect to some lower level, not in the domain specific sense, but in the full blooded sense. This being the
case, one might be tempted to claim that we may as well dispense with the higher level, since we get, by hypothesis, all the explanatory and predictive power from the more basic level; there is no explanatory need for the higher level. In this case it would seem that parsimony and reduction do not come apart.

However, upon further consideration, this should not guarantee that the higher level may be dispensed with. There may be certain pragmatic grounds for retaining the higher level contrary to the ideal of parsimony. Perhaps the descriptions at that level provide a shorthand method for making predictions or providing explanations. We are, after all, finite things and if Bruce’s cheering at the cricket match can be predicted from his neurobiological make up only after one million years calculation, then the level providing that prediction won’t be methodologically very useful. If this is the case, then we should not dispense with that level. It is this pragmatic component which might do a lot of the work in determining when we have a special science. Even if psychology is reducible to one of the neurosciences, its status might still be that of a special science because of such pragmatic necessity. Whether or not there is such a pragmatic component to the dispensability or otherwise of some level will depend, of course, upon the particular details of the explanatory-descriptive levels in question.

This last point, though, overlooks an important issue regarding pragmatics and explanation. These days in the philosophy of explanation, a lot of noise is often made about explanation’s being a pragmatic notion. That being so, then perhaps we don’t get explanatory equivalence across two levels unless we have pragmatic equivalence. This seems right to a degree; even though it is, in principle, possible to predict or explain Bruce’s behaviour after a million years calculation, we would certainly not claim these as explanations or predictions just because we would never have access to the results. The putative explanations at this level would be of no use.

What about more plausible pragmatic differences across levels which we think are not methodologically autonomous? If predicting Bruce’s behaviour took an hour longer in, say, neurobiological terms than in psychological terms, then perhaps that difference would count as an explanatory difference, hence undermining the possibility of reduction. I mention this difficulty in passing only. To decide whether such pragmatic difference should count as an explanatory difference will take us too far into the philosophy of explanation. I suppose that from the point of view of methodological autonomy, whether such pragmatic differences are going to count against the possibility of reduction will depend, in part, upon how strong is one’s parsimony. The extremely parsimonious theorist might well put up with an
hour difference in time taken to explain Bruce's cheering, whereas the less parsimonious might not.

### 3 Developmental Autonomy

If it is the level of the program at which cognitive science resides, then the psychological program running in us must be *discovered*; even if we assume that the functionalist is right, and that there is no, in principle, problem with the reality of algorithms, we still have to devise the algorithm which makes up that program. Developmental autonomy is the thesis that during the *stages of formulating* the kinds and generalisations featuring at some level, in our case the program for cognition, that formulation can proceed at that level in its own terms, without recourse to the conceptual resources of some other descriptive-explanatory level.

Because developmental autonomy has to do with the postulation of the kinds and generalisations of the cognitive sciences—*i.e.* cognitive science is autonomous with respect to its algorithm formation—we can say that it amounts to a division of *developmental* labour.7 This seems to be the view of many theorists such as Pylyshyn (1980 & 1984). The cognitive psychologist, it is claimed, has the task of formulating the “mental programs” and algorithms which are responsible for generating mentality. The task of the neuroscientist is reduced to the engineer theorising about the hardware on which these programs run. The neurosciences are a structural enterprise not dealing with the functional organisation of the mind.

Developmental autonomy seems to be the second component of the Simon quotation above: the way things turn out in the hardware realising a functional model is irrelevant to psychology.

The kind of claim being made by a developmental autonomy thesis depends upon the modal status of the claim. There is a strong version of developmental autonomy which claims that psychology *must* proceed on its own without reliance upon the neurosciences. The weaker claim is that psychology *can* proceed on its own, and whether it does so, will depend upon the vicissitudes of the levels concerned. It is not likely that the strong claim can be maintained: physiological organisations might have some role to play in the formulation process of psychological theory construction. Patricia Kitcher's example is:

---

7It's important that this division of labour is developmental and not descriptive. Descriptive autonomy can also be construed as a division of labour, but a division which still be in place when all the facts are in. Either form of autonomy can be the case with or without the other forms being in place.
Chapter 3

Suppose that physiologists discover a welter of pathways from the various regions of the brain associated with imagination to neural areas which control motor activity. This would suggest that imagination only appears to be an entirely armchair activity; part of imagining may involve the testing of motor readiness as well as the testing of hypotheses. (1980 p. 135)

Patricia Kitcher (1984), following Philip Kitcher (1984), calls a more basic level’s contribution to a high level explanation “explanation extension”. Explanation is not reduction, but allows there to be explanatory relevance between levels. For example, transmission genetic’s explanation of the inheritance of sickle cell anemia is enhanced by chemical knowledge of blood cells and the protein production of the sickle cell gene (Patricia Kitcher 1984 pp. 102-3).

Conversely, it might be prudent from the point of view of the development of the neurosciences to employ psychological categories in the determination of neuroscientific categories. Hypothesises Kitcher:

if psychology tells us that A-states always produce B-states, and if we know that A-states are sometimes instantiated by neurophysiological states of kind N23 and that all instances of B-states in human beings occur in a certain region of the brain, then we should look for some mechanism connecting N23-states to that cerebral location. (1980 p. 135)

While these examples suggest that the strong version of developmental autonomy is untenable, they also suggest the possibility of two levels being developmentally useful for each other. If this were the case then it would show that the weaker version of developmental autonomy is also misguided. Fodor, for one, thinks that there is “fit and mutual adjustment” (1968 p. 110) between the levels, while Wimsatt speaks of the “dialectics of reduction” in which two levels undergo
coevolution—they are the major factors producing change in each other: A lower-level model is advanced to explain an upper-level phenomenon which it doesn’t fit exactly. This leads to a closer look at the phenomenon, and perhaps results in some change in the way in or detail with which it is described. This will also lead to changes in the lower level model and may suggest new phenomena to look for. (1976 p. 231)
Fodor and Wimsatt’s proclamations hardly constitute an apodeictic case against the weak developmental autonomy of psychology. In order to reject that thesis one must show that such coevolution must occur. Now I’m not sure what would be an acceptable argument for this conclusion. However, the denial of the weak sense of developmental autonomy seems to be implicit in much scientific practice. Workers in most disciplines tend to specialise. Working within a subdomain of a discipline the differences between levels diminish in terms of their developmental impact. Consider Marr’s work on vision (Marr 1982). On his view, the only way that that subdomain can be adequately understood is by examining the domain from each of his, now famous, “three levels”: the computational, algorithmic and implementational levels. When studying vision, Marr has not developmentally limited himself to the more abstract levels of description. When working in a particular subdomain, one wants to know all one can about the that domain in question, and knowing about the implementational descriptions of that domain is going to increase one’s knowledge of the domain.

It is important to note the relationship between developmental and methodological autonomy: the two are really independent. One might be tempted to think that methodological autonomy, given its denial of type-identity, actually implies developmental autonomy. Why? Because if it is the kinds of The-One-True-Cognitive-PsychoIogy which cross-cut the kinds of a lower level—and is, hence, methodologically autonomous—the developments with respect to one level could not have any influence upon the other. On this view, demarcation disputes just don’t happen—in fact, can’t in principle happen. Kitcher’s examples mentioned were actually used by her to bring out the independence of the methodological issue from the developmental issue. The reason why they are independent is that all that seems to be required for developmental dependence between two levels, in either direction, is token identity. This is the reason why Kitcher’s examples are plausible: A-states are token identical with states of kind N23. Presumably, token identities will still be made even if no type-identities can be made, as would be the case if a level turns out to be methodologically autonomous from another. I think the deployment of the functionalist anti-reduction thesis in an argument for developmental autonomy, and vice-versa, is misguided.

---

8I thank Karen Neander for suggesting this line of argument to me.

9Unfortunately, Enç (1980) seems to be disposed towards such arguments.
Assuming that we have devised some putative algorithm of the mind, we still have to determine whether that the program we come up with is the right one—our hypothesised functional model must be confirmed or disconfirmed. This presents yet another way in which a functional level description might be autonomous from some lower level in which the algorithm is realised. If a functional model is immune from confirmation or disconfirmation from a lower level we can say that it is confirmationally autonomous.

Some statements of the putative autonomy of The-One-True-Cognitive-Psychology implicitly suggest that the neurosciences cannot fill the confirming or disconfirming role functionalism requires, in which case psychology would be confirmationally autonomous. This very strong view seems to be attributed to functionalists, perhaps somewhat unfairly, by Paul Churchland (1981): functionalism could have saved four-spirit alchemy and phlogiston if it had been thought of. However, such construals seem rather strong given the functionalist’s reliance upon the notion of realisation. If a given functional specification of a cognitive psychological history gets realised in the physical system within the domain of the enquiry, the functional description is confirmed. If that functional specification fails to be realised, then that description is disconfirmed. In other words, confirmation or disconfirmation results when token identities are (or are not) discovered. If a belief token is identical with a brain state token, a psychological theory is true of the organism with that brain state.

Evidence that functionalists are not committed to confirmational autonomy can be seen in Fodor (1968):

... it is clear that a psychological theory that attributes to an organism a state or process that the organism has no physiological mechanisms capable of realising is ipso facto incorrect. If memory is a matter of forming traces, then there must be subsystems of the nervous system that are capable of remaining in the relevant states for periods that are at least comparable to known retention periods. If no mechanisms exist, then the trace is the wrong model for the functional organisation of memory. (p. 110)

In such cases of confirmation or disconfirmation, the psychological mechanisms are token (non)identical with functionally characterised states or processes of the organism or system in question. On this view there is no way that we will have to make do with “bad psychology” or let psychologists “get away with murder” (Fodor 1985 p. 82).
It should be obvious from the considerations of this section that confirmation autonomy is in a sense domain specific. Cognitive psychology is true of us just in case we possess wetware that exhibits the right sorts of properties. If cognitive psychology is to be general enough in order to capture Martians in its net, then they too will have to possess silicon-ware, or whatever it is that realises their psychological states, with the right sorts of functional properties.

The final point to notice about confirmational autonomy with respect to the generality of the putatively autonomous discipline is that cognitive psychology might be confirmationally autonomous with respect to its laws, even though the level, as such, is not confirmationally autonomous. This can be seen in the domain specific reduction case of section 2. The reason that we could not get explanations and predictions about martians off the ground even with domain specific reductions to neurology, is that we could not foresee that those laws applied in the Martian case. Indeed, because there is probably an indefinite number of distinct realisations of cognitive psychological states—martians, Venusians, alpha centurians, etc.—we could not possibly confirm the generalisations (or laws) with respect to all the domains where cognitive psychological ascriptions are appropriate.

5 Function and the Neurosciences

Much of the impetus for the whole autonomy debate in the psychological case stems from the functionalist's insistence that mental and cognitive states are functionally individuated states whereas the neuroscientific states which realise those functional roles are "physical" or "structural" states (Putnam 1975c p. 371). In the previous chapter I argued against there being an absolute function-structure distinction as assumed by Putnam. One benefit of dismissing the two-level function-structure distinction is that we are no longer obliged to treat the relation between psychology, or The-One-True-Cognitive-Psychology more specifically, and one of the theories which putatively reduce it, the neurosciences, say, as an absolute one of function to structure. What that means is that although we take some heterogeneous collection of neuroscientific states as realising a functional state, we are not

---

10I take it that this is what is shown by Kitcher's argument (1980 p. 137) and not that reductions with one-way bridging connections are impossible. The confirmationally autonomy of The-One-True-Cognitive-Psychology with respect to its laws only hold if The-One-True-Cognitive-Psychology is supposed to apply to species other than us. If it is chauvinist, as I have hinted at in previous footnotes and argue explicitly in chapter 7, then The-One-True-Cognitive-Psychology will not even be confirmationally autonomous with respect to its laws.
obliged to individuate those realising states structurally or physically, so that they are typed as being heterogeneous. We can type these realising states, as Patricia Churchland (1986 pp. 361 & 382) has described it, neurofunctionally. When one thinks about it, it is function which distinguishes anatomy from physiology; anatomy deals with structure whereas physiology deals with function. We may now speculate that it will be some high neurofunctional level or levels which are going to be of importance when considering questions of reducing The-One-True-Cognitive-Psychology. The-One-True-Cognitive-Psychology might well end up being, dare I say “mere”, high level neuroscience! Indeed, it’s considerations such as these that allow the conceptual possibility that there can be a discipline such as neuropsychology: strictly speaking, the standard functionalist conception of psychology makes neuropsychology impossible.

For example, we identify Broca’s and Wemicke’s areas of the brain not because they were interesting anatomically, but because of the role they played in language production and comprehension. That is, what counts as the same brain state is not individuated neuroanatomically at all. Brain states S and T might be different brain states anatomically individuated but might count as the same brain state under some other method of individuation.

I take it that such an alternative method of individuation is going to be functional, and the interesting method if one is interested in the relationship between psychology and the brain. That this alternative non-anatomical method of brain state individuation is going to be the interesting one can be seen in the history of the neurosciences. We’ve known a hell of a lot about the anatomy of the brain for quite a while. The commissures, cortex, cerebellum, neuroglia, etc. have been identified for one hundred years with limited understanding. That’s because we have yet to isolate the functions that various parts of the brain perform—where parts of the brain are functional parts.12

What is at issue here is the way in which we take something to fall under a given type, i.e., how we determine what kind of thing something is. Presented with a fire engine and a tomato we might want to know if they are the same type of thing. With respect to their colour they are of the same type viz., both are red things, but with respect some other types viz., mode of transport, edible substance or shade of red, they are not of the same type. In

---

1. This is the exact description applicable if cognitive psychology is species chauvinist, as I am going to argue in chapter 7.

12. Enç makes a similar claim in his (1983). However, such claims made without allowing for reduction to occur with only one-many relations will not secure reductions since type-identity will still not be forthcoming due to the possibility of there being mental states exhibited by martians and robots.
this case, it does not follow that because the fire engine falls under the type mode of transport that it cannot be the same type of thing as a tomato. All this I take to be obvious, but I've recounted it anyway. I claim that the same goes for brain states. Just because two states of a brain or brains are different in some respect, it does not follow that they are not the same in some other respect.

There is an analogy between the cognitive and neuroscience and the case of thermodynamics and kinetic theory. We know that two gaseous systems with the same mean molecular kinetic energy can have different actual instantiations of molecules combining to generate the same mean kinetic energy. That difference, however, is not one that affects the attribution of the more abstract average property to the two diverse systems. The same is true, I claim, of cognitive descriptions realised in differently structured brains. Just as kinetic theory utilises some more abstract properties of the systems within its domain, The-One-True-Cognitive-Psychology will claim that it is abstract properties of brains which it quantifies over. As we have already seen, these properties might not be the fullblooded cognitive properties attributable to us and the much venerated martians and smart computers, but the properties amenable to reduction on the domain specific conception of reduction. With this story about kinds firmly in mind, combined with the reflections upon methodological autonomy in section 2, we may well decide that cognitive psychological states reduce to some neuroscientific states. That should not be surprising given that cognitive states are functionally individuated, and certain neuroscientific states can also be functionally individuated.

Signpost

One reason why cognitive theorists might have argued for some autonomous status for The-One-True-Cognitive-Psychology is that they felt that the pursuit of their discipline might be questionable if autonomy were not the case. I have tried to show that on some disambiguations of the autonomy thesis such worries are unfounded. In those cases where some form of autonomy does not hold, there is no need to worry that the supposed usefulness of their pursuits is mistaken. Even in the worst possible case, when all the facts are in and we know that The-One-True-Cognitive-Psychology fails to be methodologically autonomous, there is the possibility that we will not be able to dispense with the discipline for whatever pragmatic reasons come to light. To this extent, I hope I have shed some light on the autonomy of psychology debate. The implication for the current work is
this: if some lower science can developmentally affect some higher level, then there may well be some lower level scientific results which can shed some developmental light upon the question of whether the intentional realist conception of The-One-True-Cognitive-Psychology will be vindicated. Before investigating this avenue, though, some more theoretical groundwork needs to be laid.
In chapter 1 we saw that cognitive systems were a species of complex system. In virtue of what does some complex system count as a cognitive system? While this is itself an important theoretical question, an answer to it might aid us in attempting to decide the correct method of analysis of a cognitive system. If two theorists differ as to what is constitutive of a cognitive system, then that might affect the analysis proposed. Presumably, a complex system will count as a cognitive system just because of some class of its Level One or Level Two properties. I want to argue in this chapter that it is Level One properties of a system which form the basis for our describing the system as a cognitive system.

Closely related to this question is that of which level of explanation-description is appropriate for cognitive systems. There is thought to be (by the likes of Fodor and Pylyshyn (1988), for example) a distinct cognitive level of description. Our task, then, is to decide when our analysis of a cognitive system is a cognitive level analysis of that system. The answer to this question will also be stateable in terms of Levels of analysis. I assume that a cognitive level of description is applicable to cognitive systems only.

To start our ball rolling, I am going to take the second of our questions first, and work backwards to the first. Then in section 3 we can see how these considerations apply in a particular case from the cognitive theoretic literature. Section 4 concerns the issue of inexplicit representation in representational systems resulting from the discussion in section 3.
1 Representation and Cognition

In their "Connectionism and Cognitive Architecture: A Critical Analysis" Fočor and Pylyshyn answer our second question in terms of the representational states of systems:

"... any level at which states of the system are taken to encode properties of the world counts as a cognitive level; and no other levels do. ... Correspondingly, it's the architecture of representational states and processes that discussions of cognitive architecture are about. Put differently, the architecture of the cognitive system consists of the set of basic operations, resources, functions, principles, etc. (generally the sorts of properties that would be described in a "user's manual" for that architecture if it were available on a computer), whose domain and range are the representational states of the organism." (1988 pp. 9-10)

So, a cognitive level of description is one in which representational states of the system/organism feature, and representational states are those which encode features of the world. Let's, for the moment, assume that this view is right. What this view does is to shift the topic of our enquiry into the realm of representation theory. We now want to know what it is for a description of a system to advert to its representational states. This raises two further questions: how are those representational states of the system individuated; and secondly, in virtue of what are those states representational? The former question is one which ought be answerable in terms of Levels, while the second requires a theory of representation.

Care is required in dealing with these questions since the answers we come up with might be too general for the cognitive enquirer; the answers might encompass mental representations and non-mental representations such as photographs and linguistic items. The quoted passage requires the representations mentioned to be mental representations, otherwise noncognitively significant representational states of the organism will suffice for a cognitive description.1 Of course, the case of cognitive representation will be parasitic upon the general case, but the cognitive enquirer will have to

---

1 A case of nonmental representational features of an organism so featuring might be something like the following. Suppose that there is a strict correlation between some neuronal firings and body temperature, such that we can tell what the organism's body temperature is from the firings. This is representation in the sense that a thermometer or a fuel gauge represent things, and they would be representational states of the organism, but yet they would fail to count as representations of interest to the cognitive enquirer, and would not make a description in terms of them a cognitive description.
decide upon what constitutes the difference between representation and mental representation. The discovery of that difference must wait until we return to our first question.

For the moment, however, I want to concentrate on the nonmental case in order to see what lessons can be learned about representation from it. We may then proceed with the mental case that is relevant to cognitive theory. Now to our two questions.

1.1 Representational States

How are the states of a putative representational system individuated? I'll offer a tentative answer to this question, and then answer some objections.

1.1.1 Level One vs Level Two

Let's frame the discussion in terms of examples. Assuming that pictorial representation is bona fide representation, suppose we want to purchase a system which pictorially represents certain aspects of departmental life for a graduate recruitment promotion, which includes such scenes as discussion in the tea room, the packed Thursday afternoon seminar room, etc. Suppose we receive two tenders. Both provide us with a machine which, once turned on, flashes the scenes of departmental life on a screen with an accompanying soundtrack. In fact we provided both the soundtrack and the photographs. So, the machines are input-output equivalent; all one needs to do is switch the machines on and we get the same outputs.

Clearly, each of our tendered systems are representational since we see representations of departmental life in the output of the machines (why we will come to shortly). All we have to go on, though, is the output of the machines. Why do they generate the output that they do? Let us, for the moment, assume that the systems generate the outputs they do because they possess a range of representational states. Now in virtue of what do we make that claim? We make that claim by analysing the systems from Level One. The criteria by which we judge that the systems possess representational states are based upon an examination of the inputs and outputs of those system. At that Level the systems are black boxes. We don't need to open up the machines to confirm that they don't represent our departmental life, if the scenes pictured are of the brothel down the road from the University. Suppose that one display depicted the department and the other the brothel. We may conclude from the outputs of the system that the representational states of
the two machines differ. Again, we have made that judgement at Level One. Suppose the machines contained forty scenes from the department. We could then postulate forty representational states through which the machines pass in the course of their presentation.

We may now ask more about these representational states of the systems postulated at Level One. Do the machines (the machines originally tendered with the correct scenes, now) possess the same representational states? Our answer must surely be “yes”; where could any difference lie?

There might well be differences in the states of the tendered systems when analysed from other than Level One. Suppose we perform a Level Two analysis of the machines: we open them up to see why they possess the representational states they do. The first machine, call it model A, turns out to contain a photographic transparency projector which projects these images onto a screen once turned on—we initially wondered why its manufacturers did not return the slides taken by the departmental shutter-bug. The second machine, model B, contains no transparencies—its manufacturers returned our slides. It contains a CRT which produces images as a result of the playing of a video disc onto which the slides had been recorded. On our Level Two inspection, there appear different states of the system at that Level responsible for the Level One representational properties. Furthermore, it seems that there is some representational story to be told at Level Two in the case of model A. There appear to be forty Level Two states of the system which represent the life of the department—viz. the forty slides. There would also appear to be no such representational story to be told at Level Two in the case of model B (although see 1.2 for more on this point).

That there is some correspondence between the states postulated by Levels One and Two in the case of model A is not a necessary component to some system’s being a representational system at Level One, as evidenced by the existence of model B. To have assumed that is to infer some representational states of a system at Level Two just because of some Level One postulation; and that would be to make a Level mistake as outlined in the previous chapter. Again, model B shows the mistake inherent in such an inference.

Keeping these points from the theory of representation in mind, we can conclude that if a cognitive level of description is one which ranges over the representational states of the system, then that description need only be at Level One. There is also the contingent possibility of some cognitive level of description featuring at Level Two.

This summary raises the central problem of this present chapter which confronts cognitive theory. In chapter 1, we noted that cognitive theory
saw itself operating at Level Two, and that its aim was to account for Level One properties in terms of Level Two properties, discovering the correspondences between the properties at the two Levels, if indeed there were any to be found. The results so far in this chapter suggest that, to the extent, cognitive theory is going to be interested in the representational properties of systems, then cognitive theory might not be a Level Two enterprise at all, and that to the extent that it is looking for Level Two cognitive descriptions which follow from Level One analyses, it might well be wasting its time. We shall return to this problem explicitly, in a short while.

1.1.2 Derived Representation

In the previous section I assumed that we could attribute representational states to the systems in question. Perhaps, though, this is mistaken. Perhaps these systems don’t possess representational states, but represent only to the extent that we as their designers possess representational states. If this is correct then our commencing with the non-mental instances of representation in the hope finding out something about the mental instances has the representational cart before the horse.

While metaphysically mental representation might be prior to non-mental representation, it does not follow that it is epistemically prior. Even though the non-mental case depends upon the mental case, we might still increase our knowledge of the latter by looking at the former. Maybe it’s true that the putative representational states of the simple systems described depends upon us, its designers, possessing representational states. At the same time, though, that might tell us that any cases of representation might involve some purposive or teleological component. We might possess representational states just because we had a “designer”—God, or natural selection, maybe. The upshot of this is that if it’s problematic to attribute representational states to our simple non-mental examples, it doesn’t follow that that’s a problem for the simple case. Hence, we can avoid moving in the first place to the complex mental case.

1.2 Representation

In the previous section we looked at how to go about individuating the states of a system that do the representing. States represent by virtue of encoding features of the world. Now in virtue of what are features of the world encoded in states? The answer to that requires a theory of representation. Strictly
speaking I don't need to answer this question in this context; I'm a firm believer in divisions of labour, and I certainly don't have such a theory to offer. However, let's take a quick look at some constraints upon such a theory.

One theory of representation we may call the resemblance theory. According to this theory something encodes features of the world just in case there are certain structural isomorphisms between those features of the world and that entity. So, a photographic transparency might count as representation of a scene on this view. I don't have an account of just what structural isomorphisms amount to. However, it's clear that a video disk which has had the photographic transparency recorded onto it encodes features of the world too—but such structural isomorphisms would not seem evident. Actually, this is not quite right. There obviously are some structural isomorphisms present in the context of the system in which video disk operates, viz., running on a CD player which is connected to a CRT. We can say, then, that there is an alternative causal-functional theory of representation in which it is certain properties within a system that determines an encoding of features of the world. The causal-functional theory captures the photographic example, but in that example the structural isomorphisms are obvious even without recourse to stating the causal-functional context. We may say that the causal-functional theory allows for both systemic representation (with no structural isomorphisms) and explicit representation (with structural isomorphisms), whereas the resemblance theory allows for only explicit representation.

It is important to note that the causal-functional context in which a system features can be varied by adjusting the boundary conditions of the context. As in all cases of functional description such boundary conditions must be fixed in order for the descriptions to capture what they are designed for. In the representational case, by adjusting the causal-functional context, linguistic items such as 'dog' can be made to represent things other than dogs, just as the scenes of the brothel can be made to represent the sober intellectual life of the department. Certain causal-functional contexts depend upon others. The CD within the display only represents in virtue of featuring within the causal-functional context provided by that display system, and that system represents only to the extent that it features in some wider causal-functional context.

---

2This is not quite right, but will do for now. If we show a photograph of a scene to an Azande Indian they might not pick the structural isomorphisms we take for granted, and so might require some causal-functional background in order to pick out these isomorphisms. But I take it that there is, in principle, a difference between the representational structure of the videotape and the photographic negative. This case provides an epistemic objection only, which happens to be able to be overcome by filling in some background information not possessed by the Indians.
I am not going to have anything to say about how these two theory types should pan out in any detail. One relevant point which springs immediately to mind is this: the power of the resemblance theory is limited. ‘Dog’ represents dogs, but not because of any structural isomorphisms required by the resemblance theory. Clearly, we need the causal-functional theory to give an account of linguistic entities, with the system in question being some socio-linguistic context.3

It is possible to have what we may call derived encoding of features of the world. In such cases a representational system uses pre-existing representational states in its encodings of features of the world. An example might be the recording of sentences about some place onto video disk—what’s stored on the video disk counts as representing a scene because the sentences describe that scene. I take it that the classic example of derived representation is that of the computer; a data base may represent various states of affairs, not because the states of the computer system have any representational powers on their own, but because the states represent other representational entities—bits of information, say.

Now, what theory of derived representation should we adopt? Clearly, any resemblance theory will not do, since if it were true our video disk sentences would not count as representations of the world. One would have to do something like photograph sentences on this theory (although, of course, strictly speaking, the sentences would not be representations in the first place on the resemblance theory). So, the causal-functional theory is what we require when dealing with derived representational states.

What about these theories of representation and our display models A and B? From the perspective of the Level One analysis of the two systems, a resemblance theory would seem to be adequate. The inadequacies of this theory would only prove evident if, say, the systems represented aspects of the department derivatively: there might appear certain linguistic entities on the screens which describe the department. In that case the applicability of the causal-functional theory becomes evident.

At Level Two, though, things get a bit trickier. Although a resemblance theory would suffice to account for the forty representational states postulated at Level Two in the case of model A, that theory would be inadequate for model B at Level Two. The CD featuring in that Level Two analysis requires some causal-functional story to be told in order for it to count as representational. Now there might be a tendency on the part of

---

3Perhaps sentential level linguistic items are capturable by a resemblance theory since a sentence representing some state of affairs does contain representations of individual features of that state of affairs—‘John loves Mary’ contains individual representations of John and Mary, after all.
theorists to think that just because a system represents according to some theory of representation, then that theory will also allow us to individuate representational Level Two states, such as that of the CD. However, although this was the case in our example, there is no necessity that this occur in every case. Consider model B, if a resemblance theory were true: there can be no representational Level Two states of that system since the relevant structural isomorphisms are lacking. This might well have occurred in the case of model A even though the resemblance theory were true. There might have been no CD component in the Level Two analysis which we could describe as representational. What I am suggesting is that there is a sort of representational equivalent of a Level mistake which can be made in circumstances such as these. This mistake makes the false assumption that the representational states of a system must feature at Level Two in addition to Level One.

One might claim that although the Level Two decomposition of the display model B fails to show that there are forty representational states of the larger system, there still might be those forty representational states attributed to the CD component of the display. The CD has certain Level One properties which we want to describe as representational just because the rest of the system counts as a causal-functional context. In fact we will want to attribute forty Level One representational states to the CD. Now under a Level Two analysis even the causal-functional theory of representation will have problems individuating forty representational states of the CD at Level Two. Similarly, the same thing might well have happened in the first Level Two decomposition, if the modularity of the Level Two decomposition (the CD is a module at Level Two, according to chapter 1) were not forthcoming. If there is no representational description at Level Two in this way, then there can be no Level Two cognitive description either, by hypothesis.

This point is crucial, it seems to me. We now should turn to the cognitive case to see how it applies there. Before doing that in section 3, however, we need to move from the general representational story of this section to the specific mental story required for the cognitive case. Therefore, it is to answering our first question that we now turn.

2 Cognitive Systems

From the discussion thus far it might be obvious that I think cognitive systems are representational systems of some sort. I am now going to assume that they are representational systems which describe not only as representational, also as mentally representational. Now that does not get us
very far, especially since I think that one can assume that without buying into the account of mental representation famously proposed by the likes of Fodor (1975 & 1987): on such an account a system might need be equipped with sentences in a language of thought (Fodor 1975), or it might not. For the moment, all we need do is think of mental representation in psychological terms: if we think there is a particular psychological story to be told in the explanation of a system's behaviour then we can attribute mental representation to that system. So, we may now rephrase our question as to when do complex systems count as cognitive systems as: when do complex systems count as systems attributable of mental representation?

Sorting out the cognitive from the non-cognitive complex systems is something that Fodor has attempted to do (Fodor 1986). I think it is instructive to briefly survey the Fodorian attempt, not because I think that account is correct (far from it), but because it raises the general issues we want to tackle in the context of Levels (and levels) of analysis.

2.1 Fodor on Paramecia

Fodor attempts to mark out the difference between cognitive and non-cognitive systems by identifying a capacity of cognitive systems which the noncognitive systems lack—a property which is the mark of mental representation over and above mere representation. That property, Fodor claims, is the ability of a system to respond selectively to nonnomic stimulus properties. What kind of difference does that amount to? Consider what Fodor calls a primal scene:

Anything counts as a primal scene in which:

\[(3a) \text{An organism (or system—JF) A sees something S.}\]

\[(3b) S \text{ has some property O, such that:}\]

\[(3c) A's \text{ behaviour comes to exhibit some property C in consequence of (3a) and (3b). (1986 p. 6)}\]

Roughly, this means that it is A's seeing S and S's possession of O that enter into the explanation of why A's behaviour came to be C. Now what kind of property might O be? Fodor notes that not every property S possesses is likely to feature in the explanation of A's behaviour resulting from perceptual encounters with S. Some properties of S are going to enter into lawful
relations amongst the properties involved such that we can get, what Fodor calls, a primal scene of the first type, viz. a primal scene in which there is "a lawful connection between a certain property of the 'stimulus' (viz. S's property of being O) and a certain property of the ensuing behavioural response (viz. A's behaviour coming to be C)" (Fodor 1986 p. 9).

Fodor claims that it is this kind of primal scene in which ordinary complex systems and non-mental representational systems partake. Thermostats, fuel gauges and paramecia all partake in primal scenes of the first type. The reason for this is that they can respond only to nomic properties of the stimulus S, where nomic properties of a stimulus are (all and only) those which enter into lawful relations.

Now there also appear to be primal scenes in which the behaviourally relevant stimulus property is non-nomic. Fodor's example is of a primal scene in which he sees a crumpled shirt and as a consequence remarks that it is crumpled. The property of crumpled shirtedness is non-nomic because that property does not enter into lawful relations. He calls these primal scenes of the second type. It's Fodor's claim that it is only systems attributable of mental representation that partake in primal scenes of the second type. So, the capacity which cognitive systems possess but which non-cognitive systems lack is the partaking in "primal scenes in which the behaviourally efficacious stimulus property (the one that goes for 'O' in the primal scene formula) is non-nomic" (Fodor 1986 pp. 10-11); or as already stated, the difference between one of our gang, and one of theirs is that we respond selectively to non-nomic stimulus properties whereas they cannot.

Fodor gives three conditions in order for primal scenes of the second type to occur:

(1) There must be a psychophysical relation between A and S, in virtue of which some of the nomic properties of S affect (non-behavioural) states of A.

(2) The detection of these psychophysical properties must eventuate in A's coming to represent S as being O.

(3) There must be a lawful connection between A's representing S as O and the appropriate behavioural consequence C.

What these conditions mean is that certain psychophysical properties of S cause A to go into states which carry the information that S has those psychophysical properties. From those states A infers that S has O, or in other words, comes to represent S as having O. A's consequent behaviour
coming to be $C$ results from $A$'s representing $S$ as $O$ (rather than $S$'s actually being $O$), with this relation between this representing and the consequent behaviour being lawlike (given certain other of $A$'s psychological states). One way to put Fodor's proposal is that mental representation allows for a system to perceive and respond to properties that are not transducer detectable. We can see this point more clearly by looking at Fodor's diagram (1986 p. 15):

$S$ has properties:

- $P_1$
- $P_2$
- $\ldots$
- $s_1$
- $s_2$
- $p_n \rightarrow s_\ldots \rightarrow \text{represents } S \text{ as } O \rightarrow \text{behaviour becomes } C$
- $O$
- $s_m$
- $\ldots$

Fodor says that the dashed lines represent nonnomological necessity. $s_1$-$s_m$ carry the information about the light reflectance of the stimulus; they're psychological states, although not intentional states.

While Fodor's way of looking at our current problem is intuitively appealing, I don't think that the specific Fodorian working out of it is going to work. The reason goes something like this: Fodor thinks, quite rightly, that there is some difference between paramecia and us. The responsiveness to nonnomic stimulus properties is supposed to provide a property constitutive of such a difference. That property, however, depends upon there being a principled distinction between nomic and non-nomic properties. To be sure, Fodor chooses, as always, prima facie plausible examples to exemplify the distinction he wants to make; but I'm not sure that the apparatus he uses in introducing the distinction actually yields that distinction. Let's see why.

According to the proffered story, nomic properties are those which enter into lawful relations. Now as Fodor himself continually points out (1975 ch. 1 & 1986 p. 9), laws hold of things in virtue of their properties at different levels of description: their physical, chemical, hydrodynamical properties, etc.. There is also a psychological/cognitive level of description which will, presumably, have its own postulated properties with lawful—or at least
lawlike—relations between them. This means that there are going to be, and have to be on this story and Fodor's, nomic properties at the cognitive level. What might some of these nomic properties be? I'm willing to bet that some will have to be the properties which Fodor thinks are non-nomic. Take the property of being a crumpled shirt for instance. There are a couple of reasons why this might be a nomic property at the psychological level. The first is that psychological laws are going to be stated in terms of mental or cognitive states. As we have seen, such states will be specifiable in terms of inputs, outputs and relations to each other. If that's the case, then if there are mental states to do with crumpled shirts, then crumpled-shirtedness will be just the sort of property that enters into the lawful picture, and hence ought be called nomic properties on the story given.

What kinds of psychological laws might such properties feature in? Well, some mental states, viz. intentional states, are endowed with content. Fodor, in coming up with his own theory of content in terms of what he calls asymmetric dependence, claims that 'X' means Y because (i) there is a nomic relation between the property of being a Y and the property of being a cause of 'X' tokens; and (ii) if there are nomic relations between other properties and the property of being a cause of 'X' tokens, then the latter nomic relations depend upon the former (Fodor ATOC p. 31). Now presumably, if complex systems which are also cognitive systems, assuming the appropriate stimulation, come to posses mental states about crumpled shirts, and hence those mental states' content involves crumpled-shirtedness, then there must be some nomic relation between those systems and the property of being a crumpled shirt. We can state this problem in terms of Fodor's own diagram: in order for A to represent S as having O, that is, to possess some mental states with the content that S has O, there must be some other arrows mentioned in the diagram consistent with Fodor's overall programme. Such arrows would pick out nomological necessity, but would link O with A's representing S as O.

There is a second way to establish where Fodor goes wrong. It has to do with the fact that the psychophysical nomic properties which lead to A's representing some non-nomic property are going to be supervened upon by that non-nomic property. It is the physical surface properties of the shirt.

---

4One problem with following the lead from Fodor here is that, strictly speaking, it might be the case that there are only strict laws at the bottom level of description, whatever that bottom level turns out to be. Now as I have said in chapter 3, I'm not sure that higher levels are going to make reference to laws. Fodor seems to think that there are high level laws. If he's wrong, what I have to say should equally apply to the weaker generalisations, say, of the levels concerned. My objection to be outlined against Fodor, however, will require that the psychophysical "laws" he mentions in the diagram are also not, strictly speaking, laws.
which cause the light to reflect in the way it does and so affect A's transducers in such a way that A eventually comes to represent the shirt as being crumpled via the inference from the states which carry the information that the shirt has such and such pattern of reflectance. But why should we accept as Fodor does that it is the physical properties that have brought about the states which carry that information? Why can't we simply claim that it was the crumpled shirtedness which played the role in the formation of those states carrying information about light reflectance? On this alternative scenario, being a crumpled shirt would be a nomic property. Of course, because crumpled shirtedness supervenes on the physical properties of the shirt, vastly different sets of physical properties can realise that putative nonnomic property. Somewhat similarly, the states from which A infers that there is a crumpled shirt in view can vary according to whatever the physical properties of the shirt are, with the same inference being made despite there being such a difference. So, whatever the actual states are which are involved in A's inference, we can consistently describe them as carrying information about physical properties, while at the same time claim that: it is the supervenient property which eventually brings about A's representing there being a crumpled shirt in view. In terms of Fodor's diagram, there is no requirement to advert to $p_1...p_n$ instead of $O$ (being a crumpled shirt) in the etiology of C.

In addition to this story about nomic properties, the aim of the Fodorian approach is unsatisfying from the point of view of our current task—he, after all, just wanted to distinguish us from paramecia, whereas I want to decide what is characteristic of cognitive as opposed to complex systems. Even if Fodor's view is correct, it's not going to help us decide whether gorillas, cats and cockroaches are species of cognitive systems or not. If “only nomic properties of stimuli can elicit behavioural responses” (Fodor 1986 p. 10) from non-cognitive systems, we must also know if it is mere selective response to at least one nonnomic property that guarantees cognitive membership, or whether there is a range of nonnomic properties to which a response is required. Clearly, cockroaches respond to stimuli (it ran away last time, but flew this time) vastly differently from my car's fuel gauge (every time I refill my tank it responds in exactly the same way), but cockroaches don't, I presume, respond selectively to something's being a crumpled shirt. In short, any criteria for distinguishing between cognitive and noncognitive complex systems has to decide where, say, infraverbal creatures fit in. Until we get such decisions, we ought not be happy with the criteria.

Built into Fodor's account is the belief that there is some line to be drawn between the cognitive and noncognitive, in which case this problem for
his account is one of placing infraverbals, say, on either side—or perhaps on—that line. There may, however, be no such line to be drawn. Perhaps being a cognitive system is as graded concept ranging from paramecia to us and even, perhaps, beyond. If that's the case then we should expect to find some property which the more easily describable cognitive systems possess to a greater degree than those which we have trouble describing in this way. Responding to a greater number of nomic properties might be such a property; paramecia apparently respond to none, while we respond to many, and perhaps infraverbals respond to some, but not as many as us. This is a more promising approach, if only Fodor's nomic-non-nomic distinction could get off the ground. By employing this strategy of using a graded property to distinguish degrees of cognitive systemhood, I now offer an alternative to Fodor's account.

2.2 Abstraction

What kind of property might complex systems exhibit that noncognitive systems don’t? There is a kernel of truth in Fodor's account. Fodor’s posits the property of some system’s being able to respond to non-nomic properties of stimuli as the one which is the sign of the cognitive. What seems right with Fodor’s account seems to be that the property cognitive systems possess is one framed in terms of responses to stimuli—or Level One properties. Fodor mistakenly thinks that the relevant difference in responses to stimuli across the cognitive-noncognitive gap was the responding to non-nomic stimuli. What I propose is that the difference would seem to be a difference in ways of responding to a possible range of stimuli. Consider the fuel gauge. It is not only restricted in possible inputs—telling it that the tank is full won’t get it to go to the full mark—but also restricted in the possible responses to those inputs—all it can do is go to the full mark. Actually, I think there are few distinct properties lurking here, which we may examine in turn. I call these properties abstraction properties, and it is these properties which cognitive systems possess to a greater degree than the noncognitive systems.

2.2.1 Medium Abstraction

Since they are complex systems, cognitive systems have inputs and outputs—ie. stimuli and behaviour. It is those inputs and outputs which will determine certain states of the system. Let us, for the moment, call these states cognitive states. A paradigmatic example of a cognitive state might be
a belief. Now my belief, say, that Miles Davis recorded *Kind of Blue* in 1959 is typically generated by my reading the cover notes of that album, other things being equal. It is equally likely, though, that I might have come to form that belief as a result of my musicologist friend telling me that Miles recorded that work in 1959. Or alternatively, perhaps I was (contrary to fact) present at the recording sessions in March and April 1959 and thought that this year was to be a milestone in the history of jazz. These examples suggest that the perceptual inputs which lead to the generation of a cognitive state are not restricted to any medium or modality. I call this property of cognitive systems modality or medium abstraction. That a belief might be medium abstract is not crucial here. A dog’s coming to recognise a bone might eventuate due to both visual and olfactory stimuli.

That cognitive systems tend to abstract away from modalities can also be seen in the case of behavioural outputs. A cognitive state can produce behaviour operating in any medium or modality. In the case of my belief about the recording date of *Kind of Blue*, it can play a role in the causing of my writing or uttering the sentence “Miles Davis’ *Kind of Blue* was recorded in 1959”. Obviously, the degree to which a system is domain abstract will depend upon the range of transducers possessed by the system and how complex that system turns out to be.

### 2.2.2 Stimulus and Response Abstraction

In addition to cognitive systems being medium abstract, it would seem that it’s not just that they can possess states generated by stimuli across modalities and mediums, but also that almost *any stimuli* can generate a cognitive state. If this is right then medium abstraction with respect to stimuli and behaviour is merely a species of what we may call *stimulus abstraction* and *response abstraction*. In order for cognitive systems to be stimulus abstract we would need to find them possessing states generated by, and playing a behaviour-causing role in, varying stimuli within a modality or medium. This is just what we do find. The cover notes of *Kind of Blue* might have had the sentence ‘Recording date: March 1959’ or ‘*Kind of Blue* was...”

---

5For the time being, we can assume that the minimal requirements for a system’s possessing beliefs is that it exhibits inputs and outputs of the right sorts. That is, perhaps Dennett is correct about the propositional attitudes. There may be more that is required in order to attribute beliefs and desires, but that will be an issue for chapter 6.

6I use the term ‘stimulus abstraction’ instead of ‘stimulus independence’ as used by Pylyshyn (1984) and Devitt and Sterelny (1987). They take stimulus independence to be the lack of any necessary connection between particular stimuli and particular behaviour. I call this idea S-R abstraction below.
recorded during March and April 1959' or even (as part of a historical
introduction 'This milestone was recorded fours years before Kennedy was
assassinated'.

The same situation applies to behaviour as well as stimuli—that is,
cognitive systems might be called response abstract. As a result of my
believing that *Kind of Blue* was recorded in 1959 I might utter the words
"*Kind of Blue* was recorded in 1959" or alternatively "Four years before
Kennedy was assassinated was the year in which *Kind of Blue* was recorded".

2.2.3 S-R Abstraction

Accepting that cognitive systems are stimulus and response abstract, what
does that really tell us about the production of behavioural output from the
system? We know that by being stimulus and response abstract a cognitive
system's states may have been generated by some other stimuli or given rise
to some other behaviour. But given this, it is still possible that there is some
necessary connection between some stimulus and response pairs. An example
of this might be that my reading the question 'Recording date of *Kind of
Blue* == ?' might cause me to utter "*Kind of Blue* was recorded in 1959" and
nothing else. Upon receiving the stimulus 'What year did Miles record *Kind of
Blue*?' my belief might lead me to utter only "Four years before Kennedy was
assassinated". So, even though a cognitive state can be generated by an
number of different stimuli-types, the relation between stimulus and
response might be quite fixed: for any stimulus $S_i$ and any response $R_n$ there
is a necessary connection between $S_i$ and $R_n$ such that $S_1$ results in $R_1$, $S_2$
results in $R_2$...

One of the lessons learned from our flirtation with Behaviourism is
that, at least, the human cognitive system would seem to exhibit no such
necessary connection between stimuli and response. We may call this lack of
connection *S-R abstraction*. S-R abstraction, or something like it, has been
argued for Pylyshyn (1984 pp. 12-15). The prospects for this being the case
seem fairly good. In order for cognitive systems not to be S-R abstract, it
would have to be the case that we could predict behaviour from the presence
of a stimulus alone, without recourse to any states of the organism. Even

---

7Pylyshyn calls S-R abstraction stimulus independence and Fodor speaks about
non-nomic properties of stimuli. I cite these arguments for S-R abstraction for the
reader's interest, and not because I agree with the arguments contained therein.
Devitt and Sterelny (1987 pp. 4-5) claim S-R abstraction to be a property of human
language. This claim sits nicely with Fodor's contention that it's no accident that
verbal organisms are the ones which detect non-nomic properties of stimuli, and that
it's these organisms whose behaviour we describe cognitively (1986 pp. 17-18).
granting that I both believe that *Kind of Blue* was recorded in 1959 and desire to convey this fact to friends at every opportunity, the exact form of my behaviour, whether I utter a sentence, point to an old calendar or even nod my head, would not seem determinable from an initiating stimuli.

### 2.2.4 Domain Abstraction

Domain abstraction is the feature of cognitive states by which they can be generated by incoming stimuli of any source. The idea here is that in the case of beliefs, for example, what we can form beliefs about is pretty well unrestricted. We form beliefs about things that don't exist (unicorns have four legs), that existed before the rise of our species (dinosaurs were dumb), and that are abstract (nine is larger than seven). Whatever is a potential source of stimuli—potential because even though we might never form beliefs about something, we probably *could* given idealisation with respect to such things as length of life span and the development of better instruments, *etc.*—may be called a *domain*. Our cognitive states would seem not to be restricted to any particular domain or domains.

There are many sources of potential stimuli one might think could not generate cognitive states, since we are restricted in what counts as stimuli due to the make-up of our transducers. We have no transducer capability by which to directly detect neutrinos, infra-red radiation or abstract objects. When we form cognitive states we need not detect the presence of some object by its directly affecting our transducers in the way that we see an apple just by light reflecting off the apple. We don't have those sorts of transactions with neutrinos and numbers. But we do have more indirect theory laden transactions via stimuli which are somehow correlated with neutrinos—specks on a photographic plate.

We can see in these abstraction properties where the kernel of truth in the Fodorian account lies. What Fodor takes to be non-nomic properties of stimuli are really abstract properties of stimuli, and it is these properties which typically feature in the states of a system which possess these abstraction properties. Crumpled-shirtedness is a classic abstract property multiply realisable by psychophysical stimuli. S-R abstraction is in fact quite close to idea behind Fodor's non-nomic property story: it is those systems that enter into primal scenes of the second type which are S-R abstract. The distinctive feature with respect to their being responded to by cognitive systems is not their nomicism, since abstract properties will *have* to partake in lawful regularities if psychology is at all possible, but their high level nature. These abstraction properties are not only of importance when
distinguishing cognitive from run-of-the-mill complex systems, but also when examining various models of the architecture of cognitive systems. We will, therefore, return to them in later chapters. I should point out before continuing that while I envisage that possession of these abstraction properties is necessary in order to attribute a cognitive status to a system, they by no means are jointly sufficient for the presence of cognition. I think there are some additional properties which must be possessed in order for a complex system to count as a cognitive system. I will return to these in chapter 7, where they are most relevant to a discussion of classical objections to functionalism.

2.3 Paramecia and Levels

Having outlined what I think is crucial in distinguishing the cognitive from the noncognitive, we can now examine some of the general characteristics of that account. It is these characteristics which are an essential component of any account which attempts to draw the cognitive-noncognitive line.

In section 1 it was argued that it is a Level One analysis by which a system counts as representational. We saw that even when a Level Two representational analysis of a putatively representational system was not forthcoming we could still count that system as representational due to a Level One analysis of it. An appropriate Level One analysis is sufficient for the attribution of representational status. The message in Fodor's account is that it is the specific properties, viz. non-nomic properties, that a system represents that is important when we want to attribute mental representations to that system. Therefore, the correct Level of analysis we require in the determination of a system's having mental representations will also be Level One. That this is so can be easily seen from the Fodorian analysis itself. Being selectively responsive to non-nomic properties is specified in terms of the inputs and outputs of a system: the three conditions of when something counts as a primal scene appear to be Level One requirements. Initially one might think (as Fodor himself probably does) that Fodor's account is at Level Two because there is mention made of states and representations, but we know from the previous chapter and the previous section that there can be Level One states of a system, and that what counts

---

8I stress again that Fodor's view is not that it is the responding to non-nomic properties which somehow is definitive of representation, rather than mental representation. If that were his view then it would exclude the mundane examples of representation that featured in the previous section, such as the fuel gauge. Responsiveness to non-nomic properties is putatively definitive of mental representation only.
in something's being representational are its Level One properties. Similarly with my proffered account. The abstraction properties just mentioned make no Level Two commitments. Cognitive systems possess those properties at Level One.

Actually, I think that whatever account one comes up with it's going to have to be a Level One analysis. This point, however, is in dispute in some of the literature. The brunt of Fodor and Pylyshyn's (1988) paper is the claim that there are Level Two constraints upon when a system counts as cognitive. There are certain capacities which are characteristically exhibited by cognitive systems, such as the productivity and systematicity of thought. Now, as I have argued elsewhere (Braddon-Mitchell and Fitzpatrick Forthcoming), these capacities are *behavioural*. To that extent they are relevant to Level One analyses. Fodor and Pylyshyn think that only a particular Level Two account will generate such Level One properties. We argue (Braddon-Mitchell and Fitzpatrick Forthcoming) that provided that there is some other story (such as a teleological or evolutionary one) in place that constrains the Level One properties of a system, then there is no requirement for the Level Two properties to act as constraints upon the presence of the behavioural capacities in question.

For our present purposes, though, the Fodor and Pylyshyn point is somewhat beside the point. Of course, some system's repertoire of behavioural responses is going to be determined by its Level Two states and processes. Rocks fail to count as cognitive systems because of their internal states and processes (or lack thereof). What is crucial for present purposes is that the capacities which are characteristic of cognitive systems are behavioural and specifiable at Level One. The Level Two states which may give rise to the Level One capacities cannot be *constitutive* of cognitive applicability just because many *different* Level Two architectural stories can generate the same Level One capacities. Although those Level Two properties determine the capacities, it is those Level One capacities that are criterial of cognitivehood. There wouldn't seem to be any argument for moving to Level Two to decide upon the cognitive status of a system here.

Finally, I think we are compelled to remain at Level One for the determination of cognitive attribution because of general complexity theoretic reasons of chapter 2. All cognitive systems are going to be high level complex systems because they depend upon many *other* complex systems. Remember, though, that *we* are organisms (where organisms are a species of complex system). To that extent, our cognitive systems are going to have to feature as subsystems of the overall organismic system. That overall organismic complex system would decompose into various subsystems, such as cognitive
systems, digestive systems etc. What is important for such a decomposition is going to be the inputs and outputs of those subsystems. It is the functions of systems such as the cognitive system which go to make up the larger system by which we make the decomposition that we do. That is, it is the Level One properties of the subsystems in terms of which they feature in the Level Two analysis of the overall organismic system. In making that decomposition, the actual way in which those functions are performed, that is, the further Level Two analysis which we might perform upon those subsystems in order to properly explain those functions, is irrelevant to the individuation of those subsystems at that specific Level Two analysis of the organismic system. We postulated that there is some subsystem just because of the capacities of that system, and how those capacities contribute to the overall capacities of the original complex system.

If what I have said in this section is correct, then it gives credence to a long standing criterion for mentality: the famous “Turing Test”. Turing (1950) considered the question ‘Can machines think?’. He thought that if a suitably powerful and programmed computer attached to a teleprinter could “imitate” a human interlocutor then that might be grounds for describing the machine as a thinker; and if such a description of the computer were meaningless in current usage (1950) then the use of words would certainly change by the end of the century so that we could describe the computer in this way (1950 p. 442).

That a computer fooling a human interlocutor in what Turing called ‘the imitation game’ could possibly provide a criterion of mentality or cognition has been often rejected in the following sort of way. Design a computer that obviously does not think—by having it operate according to some “dumb” strategy such as list searching—so that it can pass Turing’s test. In this way certain Level Two constraints are thought to come to bear upon something’s counting as a cognitive system. However, presuming that such a machine can pass the test, not enough information has been given to decide the issue. The problem is that passing the test would be sufficient provided that it is evident that the machine exhibits the abstraction properties I have cited. A machine that works according to a dumb list processing or complex hook-up procedures is not going to be S-R abstract, for instance. Only if the computer can be designed in a generally dumb way that exhibits these properties am I going to have to bite the bullet here. However, this ploy to deflate the importance of Level One analyses of cognitive systems such as the Turing Test needs to show that their dumb machines can really imitate a bona fide cognitive system such as a mature...
intact human being. A machine is an imitation of us only if it possess the abstraction properties introduced in this chapter.

What I think the Turing machine example shows us is that Level One descriptions of systems do not just require a collection of inputs-and-outputs when that Level is providing criteria for some property such as being a cognitive system. Level One analyses may often include certain relations and regularities between those inputs and outputs in the criteria. The abstraction properties I have mentioned are designed to be such relations and regularities in the cognitive system case.

That's all I have to say about why we should answer our second question in terms of Level One. What this and the previous section are supposed to do is convince one that it is Level One analyses which are important to the question of determining how to go about individuating cognitive systems.

3 Connectionism

Now that we've done a little theorising about (mental) representation in the abstract, let's try to apply our insights in a particular case from the literature on cognitive architecture. One of the more new wave approaches to cognition is the so called connectionist or parallel distributed processing (PDP) approach. I am not going to give a detailed description of these approaches here. This approach encompasses many variations in dealing with the modelling of cognitive systems. I refer the reader to works such as Rumelhart et al. (1986), and for summaries Smolenski (1988), Ramsey, Stich and Garron (CEFFP) and Sterelny (Forthcoming). Very basically, on one version, connectionist networks consist of a series of interconnected units or nodes consisting of input and output units, and sometimes hidden units, each of which carries an activation value. Each of the connections are weighted, and sometimes the hidden and output units are assigned a "bias" which affects the unit's activation level. When the network receives some input, activation spreads throughout the network according to the activation levels, and weights and biases, with output finally being produced.

How does such a network encode knowledge? It is normally considered that the knowledge of such a network is encoded in the connection strengths and biases. For this reason, connectionist networks are often taken to store information in a widely distributed manner; there is no node or connection which corresponds to any particular feature of whatever the network is supposed to represent. This is often the mode of representation of those networks which possess learning procedures (such as back propagation)
which enable them to literally program themselves with the help of a trainer in some learning period. However, it is possible to construct networks in which nodes can be interpreted as corresponding to explicit features of whatever the network is designed to encode knowledge about, but these won't be considered below. We'll stick to the most extreme case possible with no localised representation.

Now let's suppose that we want to construct a connectionist model of a small memory where knowledge is encoded in a distributed manner—the kind of network that Ramsey, Stich and Garon (CCFFP) develop. What the network stores, for example, might be a set of propositions. What the system can do is judge the truth or falsity of certain propositions given as input to the system. We may now ask a number of questions about our connectionist model. Is the system our model represents a representational system? From the previous discussion it would seem that we should answer this question in the affirmative: the role of Level One analyses combined with the causal-functional theory allows the system to represent the scene without there existing any Level two structural isomorphisms between the states of the system and that which the system represents. To the extent that the model gets the right answers, i.e. gives the appropriate output for a given input given the causal-functional context, it stores information and is representational. Our system is also derivative in its representational powers since the relations between elements in the scene are represented propositionally, and our network in turn represents certain propositions as holding. Indeed, Ramsey, Stich and Garon concede that it is the Level One properties of their network which determine that it is representational. On their view, if the network correctly judges the truth or falsity of a proposition, then the network can be said to store information about the truth or falsity of those propositions; it is the capacity of a system to perform some task that leads us to attribute a representational status to the system.

Now, granted that our system is representational, how do we describe its workings? The answer to this question is crucial when it comes to deciding about levels of description-explanation. Two descriptions come readily to mind, although there are undoubtedly many more. The first describes the operation of the system in more or less representational terms, when given an input we claim that it is because of the system's encoding (distributedly) of some state of affairs—call it proposition $p$—that it produces the right output. This is a description in terms of the system's Level One properties. There is another description at Level Two. It makes no reference to representations of the system. Instead, it lists the spread of activation through the network in terms of the weights and biases, and claims that that output is the correct
(for the relevant causal-functional context) output for that input. In other words, there is no propositional modularity, as Ramsey, Stich and Garron call it (CEFFP p. 12), evident at Level Two.

The question now is: which description actually describes the system's workings? The answer to this question will, of course, depend upon our interests and which level of explanation-description we think to be appropriate. Such a question is often interpreted as a request for a Level Two analysis. But given the legitimacy of both Level One and Level Two analyses, we can really use either description. It all depends upon what one is after in advancing the description: if one wants an explanation of how the system produces some output for a given input in terms of inner workings of the black box, then one goes for the latter description. If we only want to analyse the system in such a way that we can merely predict which output will result from a given input we can opt for the former description.

Now, if we want to describe our system at the cognitive level which description do we use? Since to describe a system cognitively is to appeal to representational states of the system—and the second description would not seem to do that—the first representational description is the one we must use, given the results of the previous sections. If this is right, then the second description cannot be a cognitive level description. It is for this reason that Fodor and Pylyshyn claim that descriptions of putative cognitive systems in connectionist vocabulary must be descriptions at a noncognitive implementational level of description (1988 pp. 64-66).

This has consequences. Ramsey, Stich and Garron claim that the model described in PDP terms is supposed to be a cognitive level description. But why should we take their word for it, given the discussion so far? Their preferred description of the model is in terms of hidden units, weights and biases; but as we have just seen, such a description is not going to yield a cognitive level of description. Of course, the connectionist wants to claim that the model is a representational system, and that there is some cognitive level description of its operations. That is going to be a possible position for the connectionist since, as we have seen, it is the Level One properties of the system that allow them to make that claim. The problem is that the Level Two description they offer fails to be the description they think it to be.

Does this give Fodor and Pylyshyn their point against our version of the connectionist programme? Does the failure of there being a Level Two cognitive level description mean that such connectionist models lie at the level of implementation, being a mere account of the level which realises the cognitive level description? I don't think so: there is the Level One description of the system which is enough to guarantee the cognitive status of their
model. If this is so, then what's all the fuss about? The fuss generated by Fodor and Pylyshyn is generated by two of their claims, viz. (i) that a cognitive description of a system is one that adverts to representations, and (ii) that cognitive descriptions are at Level Two. What the above considerations from complexity and representational theory suggest is that (i) and (ii) don't have to go together; one can maintain (i) without adhering to (ii). To insist that (i) and (ii) do go together means that one comes terribly close to committing a Level mistake. That (i) and (ii) can be factored apart seems to be a view implicit in some of the connectionist views, such as the extreme version we have briefly considered here. They show how one can individuate a system as being cognitive without postulating Level Two representational states, that is, without there being a Level Two cognitive level of description. However, one can still cognitively describe that system by adverts to a Level One analysis.

What the considerations of this section show is roughly this. The connectionist example I have employed shows that there is a worst possible case for cognitive theory, in which there is no representational level of description between a Level One analysis and an implementational Level Two analysis. I am not saying that such a scenario is the case, only that it might be the case. If it turns out to be the case then The-One-True-Cognitive-Psychology as traditionally conceived will have to rely upon Level One analyses in order to frame its explanations and predictions couched in representational terms.

4 Inexplicit Representation

The question of how much inexplicit information a representational system can have seems crucial for the current problem. Since Dennett's "Get your queen out early" example, striking the correct balance between explicit and inexplicit representation for various representational systems, and determining what are the criteria for differentiating explicit from inexplicit representation, as we saw in chapter 1, has been problematic. The considerations in this chapter are at one with the moral Cummins (1983 & 1986) has been long preaching. Cummins (1983) criticises certain approaches to the explanation of cognitive capacities because they postulate internal manuals in order to explain those capacities. Central to the internal manual model is the view that programs or instructions are represented in the systems that execute them. He then goes on to adduce reasons why this view is mistaken. Now put this way, the internal manual model is just what Fodor and Pylyshyn seem to be advocating, not just as a methodology for cognitive
Chapter 4

Cognitive Systems

analysis, but as a constitutive feature of cognitive description. The quotation at the beginning of this chapter is explicit on this point, since they want the operations of the system to be described in terms of some manual for the operation of the functioning program. As Cummins says:

... there is a temptation, often not resisted, to suppose that systems do what they do because they are programmed to do it, hence to treat programs as causes of behaviour rather than as interpretive analysis of the capacity to produce it. (1983 p. 44)

As we saw in chapter 2, interpretive analysis can be construed as a Level One analysis. So the mistake of the internal manual model is that they fail to recognise the distinction between Level One and Level Two analyses, and to the extent that they do recognise that distinction they leave themselves open to making a Level mistake.

Signpost

This at-worst scenario I have been describing in this chapter has been designed to show that it is Level One properties of a system that constitute its being a representational and cognitive system. To that extent, all one requires in order to generate cognitive level descriptions is a Level One analysis. This at worst case also shows that if cognitive systems are analysed at Level Two, then it is not guaranteed that an analysis will yield a cognitive level of explanation-description, as is exemplified by the connectionist model in section 3. As yet, we have only tackled the necessity claim of views committed either explicitly or implicitly to something like an internal manual model; it might turn out that a Level Two analysis generates representational states consistent with some Level One analysis, but that does not have to be the case.

We, of course, should take seriously the possibility that Level Two analyses do do the work that The-One-True-Cognitive-Psychology envisages, that the worst possible case is not the case. The remaining chapters are going to assume this. In so doing we can examine some ways in which this Level

---

9It's interesting to note that Cummins seems to think that the level of description at which interpretive analysis takes place is not a level at which the description will advert to the causal properties of the system. This is consistent with our discussion in chapter 1. For more on this with respect to the requirement for there being a language of thought see Braddon-Mitchell & Fitzpatrick (Forthcoming) and Braddon Mitchell (1988).
Two decomposition of cognitive systems might (a) be constrained, and (b) turn out. In doing that we will identify a view which underlies intentional realism.
Modularity

Part II
Chapter 5

Function and Domain Specificity

Is this a dagger which I see before me,
The handle toward my hand? ... Or art though but
A dagger of the mind, a false creation?
(Macbeth Act II Scene I)

How might a Level Two decompositional analysis of a cognitive system go? Let's assume for the moment that the subsystems individuated are cognitive modules according to the definition of chapter 1. There are various ways to taxonomise cognitive mechanisms. The standard way is to individuate them according to what they typically do—i.e. the function they perform. Take memory for example. Perhaps there are two memory systems, long-term (LTM) and short-term memory (STM). They are individuated according to what they do; they are simply information storage devices. Whatever factors constrain where a given memory ends up (length of period stored, degree of repetition, etc.), whether one is remembering events, propositions, faces or numbers would seem irrelevant to where the memory resides.

There is claimed to be an alternative individuation of cognitive mechanisms (originally employed by Franz Gall of phrenology [in]fame) in which it is not the function of the mechanism which is the sole taxonomic criterion. Instead, it is that upon which the mechanism performs the function that is criterial. That is, cognitive mechanisms are taxonomised not according to what they do, but rather, by what they do it to (Marshall 1984 p. 213). On such a taxonomy, a memory system might be constrained by what the memory is about: it is the content of a memory that determines the nature of its storage. Such mechanisms are called domain specific. If memory turned
out to be domain specific then there would be, say, facial, number, and event memory systems, etc.

The concept of domain specificity has been made to do work in some recent cognitive theorising. In Fodor's *The Modularity of Mind*, for instance, domain specificity is offered as one of the defining characteristics of modular input systems. Fodor also uses it in distinguishing horizontal faculties such as LTM from putative vertical faculties (of which modules are species) such as separate visual and spatial stores (Marshall 1984 p. 226). Fodor actually employs two taxonomic methods in distinguishing cognitive structures. The first is a functional taxonomy which purportedly differentiates the input analysis of the modular input systems from the fixation of belief of the central cognitive processes. The second is taxonomy by subject matter which differentiated the domain specific from nondenominational cognitive structures. Whether these are independent taxonomic methods will become crucial in what follows.

Similarly, Arbib's schemas are identified, in part, by the domains upon which they operate. Schemas are fine grained functional modules which, in the case of high level image interpretation, have “world knowledge” about aspects of some scene under interpretation. For example, in the recognition of suburban scenes, there may operate a sky schema, a roof schema, a house schema, and a foliage schema (Arbib 1987 pp. 352-9).

Armed with the concept of domain specificity, we might also gain an effective edge in the philosophy of psychology. In order for intentional realism to be true there must be a cognitive mechanism capable of supporting the properties we attribute to intentional states. Since we can believe and desire just about anything—i.e. our intentional states can take as their contents just about any domains, there must be some domain inspecific cognitive mechanisms in which the states postulated by intentional realism are implemented.

In short, what cognitive theory construction wants is to carve the mind at its joints, and domain specificity is claimed to provide an edge for our taxonomic blade.

This chapter is about trying to get some account of this concept of domain specificity. In order to do that we need to determine three things: (a) what a domain is; (b) how a cognitive mechanism can be constrained by or operate upon that domain; and (c) why there are domain specific mechanisms at all.

Does domain specificity provide some sort of blade with which to slice the cognitive pie? It's thought, it seems, that there is a principled difference between cognitive mechanisms individuated according to what they do and
those according to what they do it to. The claim of this paper is that this difference is, in principle, an illusion—just like Macbeth's dagger. My claim is that domain specificity is rife amongst cognitive mechanisms to the extent that we lose taxonomic sharpness, and that this becomes obvious once we understand that taxonomy by domain specificity is closely related to taxonomy by function. The reason for this illusion is that the only account—well, collection of intuitions, really—of domain specificity on offer is not really an account of domain specificity at all, but rather domain specificity qua property of modules. Extraction and critique of this proto account from Fodor's writings on modularity is the content of sections 1 through 3.

In section 4 I outline what I take to be the real issues regarding domain specificity evident when the effects of modularity have been factored out. In section 5 I attempt to extract as much of a taxonomic distinction as is possible from the account given in section 2, although the edge of our taxonomic blade is far blunter than first envisaged.

1 Domains

There are no definitions on offer as to what constitutes either a domain or domain specificity; there are plenty of metaphors though. Fodor says that “domain specificity has to do with the range of questions for which a device provides answers (the range of inputs for which it computes analyses)” (1983 p. 103). This leaves the idea of domain specificity hardly pellucid, however. More specifically, the proto account tells us that input systems are special purpose computational systems which are constrained by (a) the range of distal sources of proximal stimulations they operate upon, and (b) the range of information they can access in the course of their operation (Fodor 1983 p. 47). So, our facial recognition system might be domain specific because it allows us to recognise faces only, and not, say, numbers and middle sized objects; and it operates only upon information about facial features—beliefs about the likelihood of seeing someone or the sound of their voice do not feature in the recognition process.¹

This characterisation of domain specificity is offered in the context of listing the defining properties of modular input systems. My main claim in this section is that the proto account really is not concerned with domain specificity itself at all. It is, instead, an account of the modularity of cognitive

¹What one takes actual recognition to be is important here. It seems that contextual factors do affect our recognition of individuals; we all hesitate upon seeing a familiar face in a strange environment. The act of facial recognition must be restricted to the output “That's a familiar face” rather than “That's Bruce”.

mechanisms. Hence, one's interests in domain specificity remain unsatisfied.
There are a range of proto account red herrings; roughly they have to do with:
(i) the information processed by domain specific mechanisms; (ii) the stimuli
these mechanisms can take as their inputs; and (iii) the processing strategies
of the mechanisms.

1.1 Information

A somewhat hair splitting issue to start off: conditions (a) and (b) seem too
strong. Only (a) would seem to be involved with domain specificity. If domain
specificity has to do with the limited range of questions that the mechanism
is designed to answer, say the question as to whether that face is Norma Jean
Baker's, the sound of Norma Jean's voice might well be used in the process of
recognising her face. (b) has to do with what Fodor calls informational encaps-
ulation—the use of restricted information, visual properties of a face, say, in
the operation of a cognitive or perceptual mechanism. Because the proto
account pertains specifically to input systems which are both domain specific
and informationally encapsulated, I suspect that the need to distinguish
domain specificity from informational encapsulation has been ignored.

In fact, the proto account claims to conceptually factor out domain
specificity from informational encapsulation when talking about domain
specific mechanisms generally (and not qua input systems). Fodor says:

It is perfectly possible, in point of logic, that a system which is not
domain specific might nevertheless be encapsulated. Roughly, domain
specificity has to do with the range of questions for which a device
provides answers ... whereas encapsulation has to do with the range of
information the device consults in deciding what answers to provide. A
system could thus be domain specific but unencapsulated (it answers a
relatively narrow range of questions but in doing so it uses whatever it
knows); and a system could be nondominational but encapsulated (it
will give some answer to any question; but it gives its answers off the
top of its head—i.e., by reference to less than all the relevant
information). (Fodor 1983 pp. 103-4)

Given (a) alone, will this factoring out obtain? I think not. If our hypothesised
face recogniser were domain specific in virtue of operating solely upon the
visual properties of faces—that is, domain specificity has to do with taking
inputs from only one domain in accordance with (a)—then that mechanism
will be encapsulated with respect to all the other domains that it did not take
as its input. For example, the vocal characteristics of Norma Jean constitute a domain; one can gain information about the people in one's proximity from voices. But the facial recognition module takes as its inputs only visual properties of faces. If such a mechanism were domain specific but unencapsulated, then it may perfectly well take the auditory properties of someone's voice as input in order to aid the recognition of that person's face; one is used to hearing *that* voice with *that* face. So, the definition will imply informational encapsulation with respect to any domain other than that upon which the mechanism is supposed to operate.2

We have to be careful here. When Fodor introduces the idea of informational encapsulation he takes it to involve *top-down* information flow. The idea here is that our perceiving—the processes performed by the inputs systems—is unaffected by what we believe. On this view even though we are in a strange context and we believe that we would not be likely to meet Bruce, that should not affect the mechanisms mediating the process of recognising some face as familiar. For convenience, let's call this form of encapsulation *vertical encapsulation*. Now if the sound of someone's voice is involved with a domain specific mechanism's operation, and that domain specific mechanism is vertically encapsulated—which the input systems are, if Fodor is correct—then it might still be the case that that mechanism is unencapsulated with respect to information not involved with the agent's beliefs, but with information present in some other input system. We may call informational encapsulation with respect to information residing in other input systems *horizontal* encapsulation. Now although the formulation of domain specificity given by (a) does not imply vertical encapsulation, it does imply horizontal encapsulation since if a mechanism is operating upon a limited range of inputs, then it cannot take inputs from any other domains—where taking inputs from other domains constitutes information flow.3

We will confront the issue of information again during the discussion of proprietary processing below.

---

2 I thank David Braddon-Mitchell for this point.
3 These days Fodor distinguishes between *synchronic* and *diachronic* encapsulation (Fodor 1984 pp. 39-40). That's in order to counter various claims about background information used by the input systems—see Seidenberg and Tanenhaus (1984). But that distinction won't help Fodor to get around putative unencapsulation of input systems under his definition since even if there is only *top down* information flow between central processes and modules the cross domain information flow must be *synchronic* and not *diachronic*. 
1.2 Inputs

Domain specific mechanisms are those which operate on specific domains. In the case of modular input systems, (a) tells us that a domain is some feature of the distal environment that causes characteristic proximal stimulations at an organism’s transducer surfaces. So, cows (and probably their Twins on Twin Earth, since cows and Twin-cows are transducer property identical), tigers, faces, middle sized objects, and linguistic utterances all qualify as domains.

The term ‘domain’ has quite a large usage in AI and cognitive science. Domains are often identified with “micro-worlds” such as Winograd’s (1972) SHRDLU block world, or various fields of enquiry such as problem solving and memory. The proto account has quite a restricted view as to what constitutes a domain, in terms of the possible contentful inputs to the mechanisms. For example, what don’t count as domains are the objects of the “traditional sensory/perceptual ‘modes’ (hearing, sight, touch, taste, smell) and one more for language” (Fodor 1983 p. 47)—that is sound waves, light waves and molecules in the air. The reason is that mechanisms which take them as inputs are not constrained by the content of what is perceived. We can see and hear tigers, cows, words and sentences; and smell roses and conspecifics. Since each of the perceptual mechanisms cut across content domains in this way they don’t get to be domain specific.

Fodor has been criticised by both Arbib (1987) and Jackendoff (1987) for making modules too big. They interpret him as claiming that the visual, auditory, olfactory, etc. systems are domain specific because they are restricted to operating upon one of the sensory modalities. Says Arbib (and the confusion evident here is repeated at various points in the quoted work, eg. p. 337):

“Vision” and “language” are domains for him,... If we insist on Fodor’s usage of domain, we find that it is not defined but is given by a list (“vision, language, ...”) whose continuation is unclear. Is “reading” a domain; or is it to be regarded as “sort of domain specific,” involving language and vision but not certain other domains; or is it “domain neutral”...? (1987 p. 361-2)

But as we have just seen, Fodor explicitly rejects this interpretation of what a domain is, and is not claiming that the visual or auditory systems are domain specific.4

4This point has been seriously misunderstood. Higginbotham, for instance, says: “Fodor suggests that the fundamental ‘input systems’ for human beings are ‘the
The claims of Arbib and Jackendoff are right, of course, in implying that the traditional modalities are specific in some sense. The visual system takes as its input light waves, the auditory system takes as its input sound waves, and the olfactory system takes as its input molecules in the air. Each of these perceptual systems not only takes these stimuli as inputs but can take only these stimuli as inputs. Our question is whether or not these mechanisms are domain specific. On the proto account of domain specificity given, it seems that they are. Here's why.

Examples of mechanisms specific in the current sense are most transducers. Transducers provide the interface between an organism and its environment. Distal environmental objects cause the impingement of energy at various surfaces of the transducer. These patterns of proximal stimulations are then converted in some lawful way into, say, neural code ready for use by the rest of the organism's cognitive machinery.

These transducers are in some sense specific because they take as their inputs a restricted class of possible stimuli. If we used the same transducer for vision and hearing then we would say that it was not in this sense specific. There is no logical requirement that transducers be specific in this sense. Just as transducers can be fine grained as in the case of taste receptors (where there are four types of transducer, one for detecting each of bitterness, sweetness, sourness and salt) so they may be coarse grained as well. The mechanism might be strange one, even impractical, but still it may be able to take as its input both frequencies of light (within and beyond the range of the retina) and sound waves.

It would seem that transducers satisfy (a) (that a domain is a feature of the distal environment that causes characteristic proximal stimulations at an organism’s transducer surfaces) easily. These transducers take as their inputs a particular narrow range of stimulations. The further question remains as to whether or not the traditional sensory modalities are content domains. Let's backtrack for a moment. The intuitive reason behind the proto
account as to why the traditional sense modalities are not domain specific seems to be that the stimuli upon which transducers operate (lightwaves, molecules in the air, or whatever) carry information across subject matters—we can see both kangaroos and Holden cars. Now what seems to be the distinguishing feature of Holden cars and kangaroos is that the usual inputs to the transducers (lightwaves, etc.) carry with them content regarding kangaroos and Holden cars. Some feature of the world only gets to be a content domain if it can feature as part of the content of some organism-environment transaction.

It is the case, however, that light waves can be the content of such transactions. After all, we have beliefs about light waves. So, not only will light waves satisfy (a) they would seem to be content domains as well. Although, intuitively, Fodor doesn’t want our transducers to be domain specific, they turn out to be on the definition.

We may summarise the problem thus: Fodor’s account of a domain requires that the domain be part of the content transmitted in environment-organism transactions, rather than being part of process by which that content is relayed. A mechanism is domain specific just in case it operates upon a particular range of relayed content. But lightwaves, also constitute content relayed in such transactions. So why don’t they too count as content domains?

What seems to have gone wrong for the proto account here is that a content domain’s featuring in the process of organism-environment interaction allows for that content domain to be solely “operated” upon by a transducer. However, the domain features in such transactions, not in its capacity as the content of the transaction, but rather in its capacity as being part of the process of the transaction. In trendy 60’s terminology, Fodor’s account requires a domain to be the message and not the medium in organism-environment transactions; unfortunately that account makes no provision for making that distinction. Consequently, transducers get to be called domain specific because lightwaves feign being the message over and above the medium.

The obvious suggestion for providing for this distinction is to restrict one’s definition of domain specificity to content domains which are not playing the role of the process in organism-environment transactions. I take it that information theory must have criteria by which to make such a distinction, and that we need to make that distinction is, I take it, uncontroversial. In the case of perception, for example, we distinguish sharply between elements of the process of perception such as lightwaves,
and what is delivered by that process, viz. what we see.\textsuperscript{5} I don’t intend to have anything to say about the principles according to which we make this distinction.

I want to point out that this problem regarding domain specificity is not the problem raised and dismissed within the proto account’s manifesto. One might think that the current problem is just the non-psychologically interesting, trivial sense of domain specificity recognised by Fodor himself (1983 p. 48). But I don’t think that the specificity of transducers is of this trivial kind. On Fodor’s trivial sense, if I read him correctly, if you have some general purpose mechanism, say, which on some occasions mediated the perception of cows, then \textit{qua mechanism of cow perception} that mechanism will be domain specific. That is, to the extent that the mechanism which mediates the perception of cows \textit{performs that very task} on a particular occasion, it gets to be called domain specific. But that very same mechanism might also be domain specific in this sense with respect to tigers, tables, chairs and any middle sized object. Says Fodor: “It is, for example, entirely compatible with the cow specificity of cow perception that the recognition of cows should be mediated by precisely the same mechanisms that effect the perception of language, or earthquakes, or of three-masted brigantines” (Fodor 1983 p. 48). Since the mechanism in question can take as its input more than mere cows, then on the definition, that mechanism cannot be domain specific; the intuition behind Fodor’s trivial sense is not that cows and three-masted brigantines are carried on lightwaves, but rather that in a particular instance the trivially domain specific mechanism can take them both as inputs. Clearly, although our transducers process stimuli caused by, \textit{inter alia}, kangaroos and Holden cars, they still are restricted in the stimuli they may process. Hence, they are not trivially specific with respect to those domains.

2 The Grain Problem

Now that we are equipped with the medium/message distinction—whatever that is, exactly—there remains the question as to how to slice up the domains which feature as inputs to cognitive mechanisms. While the sensory modalities are not domains that Fodorian modules operate upon, there seems to be some sense in which there is a visual or auditory domain—albeit a very

\textsuperscript{5}I take it that no-one would want to claim that although this distinction is made by us, it might not be relevant to transducers. Maybe light waves are the message for our perceptual homunculi and not us. Although such an idea hardly deserves a reply it would run along the lines of personal-subpersonal forms of domain specificity and that Fodor and the rest of us are talking about personal level content domains.
coarse grain of domain. I think there is a problem nestling here, which we may call the grain problem.

In a nutshell the grain problem has to do with how finely or coarsely we slice domains. What the grain problem amounts to can be seen in the case of trivial domain specificity just cited. In this case we had the example of a mechanism which recognises three-masted brigantines, meat pies, kangaroos and Holden cars. According to the standard view, such a mechanism fails to be domain specific; it flouts condition (a). But consider this very mechanism with respect to the domain middle sized objects. It may be that the mechanism will mediate the perception of middle sized objects only. In doing that, the mechanism would seem to be performing some specific perceptual task in the same way that a mechanism that would only recognise kangaroos would. If middle-sized-objects could count as a domain, then there seems to be something pretty specific, indeed content specific, about the mechanism. Of course the domain of middle sized objects is a far coarser grained domain than kangaroos; but nothing has been offered in the proto account which restricts content domains to fine grained domains.

Another example may help make my point here. Take the phonetic analysis of speech. One might want to suppose that there is a mechanism which commences operation only when the acoustic signals it takes as its inputs are taken to be utterances. My claim is that this mechanism should be domain specific. Again, a possible reply to my claim might be that utterances cannot be a content domain since they carry information about various content domains—utterances about sheep and grandmother—and therefore are media rather than messages. To count as domain specific the mechanism would have to process utterances restricted to sheep or grandmothers. Notice here that utterances must also constitute a content domain; one can hear utterances about utterances. In addition, it would seem that utterances are not really the medium at work in this case; it is really sound waves which carry particular content. So even with the additions to Fodor's account required by the considerations of the previous sections we are led to domain specificity.

In point of fact a mechanism such as this hypothesised phoneme analyser is claimed by Fodor to be a candidate domain specific mechanism (1983 pp. 48-9). My point is that if utterances count as a coarse domain and granting that middle-sized-objects also count as a coarse grained domain, then our middle-sized object detector should also count as domain specific. This highlights what I take to be a crucial tension in the intuitions lying behind the proto account. It makes mention of content domains when it really means fine grained domains; but when one comes to its list of any candidate
domain specific mechanisms such as "mechanisms for colour perception, for the analysis of shape, ... systems that assign grammatical descriptions to token utterances" (1983 p. 47), the level of grain is quite coarse.

Why is the grain problem a real problem? For the purposes of his monograph *The Modularity of Mind* Fodor was interested in mapping cognitive architecture. Input systems are described as perceptual systems processing transducer output into a form ready for what he calls central processing where belief fixation occurs. Input systems are deemed to be domain specific because they operate either upon a fairly coarse domain such as colour or shape, or upon some finer grained domain such as my favourite sheep. Given the domain specificity of these mechanisms, Fodor then goes on to contrast them with central processes. Central processes on this account are domain inspecific. Says Fodor:

> Even if input systems are domain specific, there must be some cognitive mechanisms that are not. The general form of the argument goes back at least to Aristotle: the representations that input systems deliver have to interface somewhere, and the computational mechanisms that effect the interface must *ipso facto* have access to information from more than one cognitive domain. (Fodor 1983 pp. 101-2)

From our discussion so far, there is going to be some coarse grained domain with respect to which central processes are domain specific. It's going to be a very coarse domain: THE WORLD or whatever our central processor takes as its object. But if what I have said in this section is right, we may as well call central processors domain specific. If that's the case then the proto account of domain specificity will fail to provide a criterion for taxonomising cognitive mechanisms. If this sounds counterintuitive now, it shouldn't after the section (4.2) on proper functions.

That's the grain problem. Patently, what is required is some way of distinguishing between those cases of attribution of domain specificity which we find interesting, such as the case of a face recogniser, and those which we don't. The criterion used by Fodor is the subject of the next section.

---

6If it looks to the reader as though Fodor is confusing the issue of domain specificity with the issue of informational encapsulation, you're right. The result of this confusion is the argument for why there have to be domain inspecific cognitive mechanisms. I reject that argument on the grounds of this confusion.
We have just seen that the proto account requires some criterion for distinguishing between and within levels of grain of domains in order to generate the plausible list of domain specific mechanisms. The proto account finds such a criterion by answering the question (c): how can there be domain specific mechanisms at all? The answer given by Fodor is that it is eccentric domains which tend to have corresponding domain specific mechanisms.

On this view we get domain specific mechanisms because there is some eccentric feature of the environment that cannot be operated upon by some general cognitive processing strategy. It's the idiosyncratic computational processing of the mechanism—the proprietary code with which it operates—that allows it to operate effectively upon that domain. Says Fodor:

The Haskins experiments demonstrate the domain specificity of an input analyser by showing that only a relatively restricted class of stimulations can throw the switch that turns it on. There are, however, other kinds of empirical constraints that can lead to the same sort of conclusions. One that has done quite a lot of work for cognitive scientists goes like this: If you have an eccentric stimulus domain—one in which perceptual analysis requires a body of information whose character and content is specific to that domain—then it is plausible that psychological processes defined over that domain may be carried out by relatively special purpose computational systems. All things being equal, the plausibility of this speculation is about proportional to the eccentricity of the domain. (Fodor 1983 pp. 48-9)

As Fodor correctly points out, the inference from the eccentricity of a stimulus domain to domain specificity of psychological mechanisms should be made with extreme caution. He says:

Suffice it, for the present to suggest that it is probably characteristic of many modular systems that they operate in eccentric domains, since a likely motive for modularising a system is that the computations it performs are idiosyncratic. But the converse inference—from the eccentricity of the domain to the modularity of the system—is warranted by nothing stronger than the maxim: specialised systems for specialised tasks. (Fodor 1983 p. 52)

The example used by Fodor to bring out this point is that of chess playing. Chess playing is an eccentric stimulus domain. It employs eccentric and idio-
syncratic information, but no modularity theorist is going to want to postulate a chess playing module.\footnote{In point of fact, Fodor's view seems to have changed here. In a footnote in his (1987a) he wants to know why chess playing is modular. By asking this he is not implying that there is a dedicated chess playing mechanism. What he means is that eccentric domains often require certain sorts cognitive processing; Fodor calls this "brute force" modularity. That is, in order to become proficient at chess you've got to employ certain cognitive strategies such memorising as many board positions and resultant moves as possible, in the absence of a dedicated chess processing mechanism. Whatever strategies are employed, the expert, s/he might be good enough in the adoption of that strategy so that it appears that there is a separate mechanism at work. Eccentric domains might be such that there is a mechanism whose operations are idiosyncratic because they take as their inputs idiosyncratic stimuli; such processes Fodor calls "modular in the nature of things" (Fodor 1987a p. 36).}

It is because of the eccentricity of the stimulus that we hypothesise that a domain specific mechanism is operating if an organism is to interact behaviourally with some distal object. But it's only an hypothesis, not an inference. Fodor is surely right that eccentricity is not sufficient for domain specificity given that we don't have a domain specific chess playing mechanism.

What does the proto account take eccentricity to be? Eccentricity seems to have something to do with the processes required to operate upon domains. Says Fodor:

If you have an eccentric stimulus domain—one in which perceptual analysis requires a body of information whose character and content is specific to that domain—then it is plausible that psychological processes defined over that domain may be carried out by relatively special purpose computational systems. (1983 p. 49)

The point here seems to be this. A domain counts as eccentric if the processing strategies by which a mechanism operates upon that domain are not general purpose strategies which can be employed across different domains. So: suppose there is a set of general recognition principles along the lines of a prototype-plus-similarity-metric according to which a perceptual mechanism might operate. Just so long as a domain is not processable in these terms it gets to be described as eccentric. Fodor contrasts the perception of cows with the perceptual recognition of sentences in these terms in order to show that language is eccentric, since the processing principles associated with language comprehension is specialised—if the data is to be believed.

If this is constitutive of what eccentricity is, it seems to require more than is required for mere domain specificity. The reason is that all that is required for a mechanism to be domain specific would seem to be that it
operate on a particular domain. There might well be domain specific mechanisms which all use general processing strategies. What makes these mechanisms domain specific is that they are mechanisms operating on different domains. There will need to be some story told as to why there are these fine grained mechanisms rather than one big mechanism performing all their functions. It is for this kind of reason that I opt for a different explanation as to why there are domain specific mechanisms below.

Again the reason for this account of eccentricity is that the class domain specific mechanisms qua input system can be at least partially explained by reference to idiosyncratic processing. If this is the case though, it's hard to see why it should be included in an account of domain specific mechanisms qua domain specific mechanism. Fodor gets his class of input systems only after he shows that there are mechanisms with the requisite list of properties (such as domain specificity, informational encapsulation, mandatory operation, etc.), and not by inter-defining these requisite properties. Despite Fodor's belief that the class of domain specific mechanisms is more or less coextensive with the class of input systems, we want to know why there are domain specific mechanisms at all, not why there might be modular input systems.

There are two worries with this line of thought. The first concerns information again. On the proto account's view of eccentricity, a domain specific mechanism requires "information whose character and content is specific to that domain". Assuming that this information is internal to the mechanism, so as to avoid issues surrounding encapsulation, this condition would seem to be easily fulfilled—in fact too easily from the proto account’s point of view. Even in the case of a mechanism’s operating according to the principles of a prototype-plus-similarity metric there is "information whose character and content is specific" to whatever domain the mechanism is operating upon. In order for a general purpose mechanism to recognise cows, it must possess at least a cow prototype with which to perfect a match which will lead to recognition.

My second worry is more substantive. Consider the current interpretation of eccentricity in terms of another putative domain specific mechanism, viz. face recognisers (Fodor 1983 p. 47). If the current view is correct then there is such a mechanism just because faces constitute an eccentric domain. But what reason do we have for believing that? It seems that a face is something that should be quite easily amenable to recognition utilising the general purpose prototype-plus-similarity-metric principles just mentioned; faces would require the formation of a prototype just in the same way as my-favourite-sheep would. So why should faces be eccentric domains?
Any claim that faces are something eccentric, and thus requiring idiosyncratic processing needs to be argued for independently. In the case of language and sentence recognition, there might be a *prima facie* case made for the eccentricity of those domains. The account of eccentricity seems to make not being processed by some general purpose strategy *criterial* of eccentricity. So by that account faces are eccentric. However, that's got to be cheating.

What this shows is that as well as eccentricity's not being a sufficient condition of domain specificity—as the standard view allows—further argument needs to be provided before we even accept that eccentricity construed in this way is even a necessary condition.

On the proto account, it is important that there are not mechanisms domain specific with respect to sheep—they don't count as an eccentric domain. However, although it might be obvious that we don't have any domain specific mechanisms with respect to *them*, *they* (sheep) probably do. If the eccentricity story is right then we would be at a loss to explain the existence of domain specific conspecific detector in sheep; moreover, if the story is right it would seem that there *couldn't* be such a mechanism.\(^8\)

The emphasis placed upon eccentricity by the proto account arises from the belief that specialised computational strategies generate certain processing advantages, *viz.* speed and efficiency. If certain cognitive-perceptual tasks are to be performed at all effectively then the operations of such mechanisms should be so as to achieve these advantages. If this line is right, then it would not seem that eccentricity, as such, is the real reason why we have domain specific mechanisms. We shall return to the issues of speed and efficiency in the next section.

### 4 Domains and Function

We have spent the previous section going the through the false starts of the proto account of domain specificity. With an eye to making sure that those aspects of cognitive mechanisms pertaining to modularity are factored out, we can proceed to look at domain specificity afresh.

\(^8\)There's a reply to this to the effect that what counts as eccentric is relativised to a species. So, although sheep aren't eccentric for us, they are for sheep. Intuitively, this has things round the wrong way. Surely we would expect conspecifics to count as noneccentric from the point of view of other conspecifics. This assumes, though, that eccentricity has to do with, say, frequency of occurrence in an environment, which is not the proto account's view.
4.1 Functional Specificity

We saw in section 1.2 that many more cognitive mechanisms were going to count as domain specific than on the proto account's view of domain specificity. I offer the following analysis as to why. The sense in which coarser grained domain specific mechanisms are specific seems to be related to the function we typically attribute to that mechanism. Our middle sized object detector, for instance, is described as having the function of detecting middle-sized objects. When the function of a mechanism is described in terms of some domain—either coarse or fine grained—we may call that mechanism functionally specific. Any mechanism to which we attribute a particular function is going to be functionally specific.

We also saw in section 1.2 that transducers were in some sense specific: the retina takes as its input only lightwaves, and the olfactory senses take as their inputs only air molecules. That specificity also seems to be specificity with respect to function since it's the function of the retina to take proximal visual stimulations in the form of frequencies of light and transform those stimulations into a media appropriate for our visual perceptual systems.

What is the relation between functional specificity and domain specificity? It seems that the reason why we wanted to call a mechanism content domain specific was that the function that mechanism performed was restricted to a content domain; it's the function of the my-favourite-sheep recogniser to turn on in the presence of a particular sheep, the function of the linguistic input analyser to process utterances, just as it's the function of our face recogniser to recognise faces. In each of these cases we attribute a particular function to a mechanism where that function is described in terms of some content domain. This suggests that a mechanism ought be described as domain specific just in case it operates on the domain featuring in the description of its function. On such a construal of domain specificity, the mistake in Fodor's account is that it requires a mechanism to operate upon fairly fine grained domains in order to be deemed domain specific, whereas in fact any level of grain will do. If this is right, then it follows that there is no principled difference between a functional taxonomy of cognitive mechanisms and a taxonomy based upon operations upon a domain. The latter taxonomy turns out to postulate fine-grained functions (event memory, say) rather than coarse-grained functions (LTM, say).

If specificity of function is all there is to being domain specific then what about transducers and the medium-message distinction? If we restrict the class of domain specific mechanisms to those which operate upon domains
which feature as the message in organism-environment transactions, then
domain specificity can be had very cheaply—many more things turn out to be
domain specific than we first thought. If we wanted to be difficult we could
further expand the class of domain specific mechanisms by including those
mechanisms, such as transducers, which are functionally specific with respect
to domains featuring as the medium in organism-environment transactions. I
don't want to be difficult; and so I think we should adhere to the proto
account's intuition that transducers are not domain specific. We are, after all,
concerned here with content domains, and content is restricted to messages in
this context. Assuming this, there is an obvious distinction to be had between
functionally specific non-denominational mechanisms and functionally
specific domain specific mechanisms. We'll return to this distinction in section
5.

Meanwhile, there is an obvious objection to my construal of domain
specificity: it's the claim that this talk of functional specificity seems trivial.
Of course, if the function of a mechanism is to recognise faces then that
mechanism operates upon faces. One reason why one can have such a
response without detriment to the account on offer is this. Functional
specificity seems trivial because we antecedently know the function of the
mechanism. If we know in advance that there is a mechanism which takes as
its input transduced patterns of retinal excitation, then claiming that the
mechanism is domain specific, in that it can only take as its input transduced
patterns of retinal stimulations, seems uninteresting; but the fact that there
is a mechanism which is dedicated to taking these inputs only is going to be of
theoretical importance if you’re trying to map the architecture of a cognitive
system by deciding which functions have dedicated mechanisms.9 So, the fact
that a mechanism takes as its inputs sheep and three-masted brigantines,
can be of interest qua middle-sized-object detector if the presence of such a
detector makes a difference to the architectural model being developed.

4.2 Proper Functions

The above account of domain specificity requires some attention. The general
question as to what function we wish to attribute to a cognitive mechanism
or, indeed, anything is a problematic one. But it is one that must be answered

\footnote{Such a mapping is precisely the task that awaited vertical faculty theorists such as Gall. The domain specific mechanisms would have been the fundamental faculties. Gall, had, however, no conceptual or sound theoretical apparatus by which to come up with such a mapping.}
if the account is to get off the ground. I want now to take a brief look at attributing what Millikan has called *proper* functions (Millikan 1984).

One way to attribute functions to a mechanism is based upon the powers of that mechanism. A kidney has the power to filter blood. But what about kidneys which are diseased or malfunctioning so that they no longer possess that power? A mechanism has a proper function not according to the powers which it actually has, but according to what it was *designed* to do—it's *history* rather than *performance* that counts. With this in mind let's now consider an example of attributing a proper function to a cognitive mechanism.

Consider Braitenberg's (1984) example of some organism's possessing special-purpose hardware that is sensitive to visual patterns exhibiting symmetry around a vertical axis. When this mechanism switches on there is an orientation reaction (in the case of fish, preparation for flight). Why is there such a mechanism? If the proper function we ascribe to the mechanism is: detector of symmetry around vertical axis on the retina, then we might be at a loss to explain why we have such a mechanism—what advantage would *that* afford the possessor of the mechanism? It seems that in nature the vertical symmetry intended to be identified by this mechanism is the face of some other organism. It appears, then, that our vertical symmetry detector really delivers information of the type “there's a face looking at you” where the face is of a predator or mate. An alternative reason why there is such a mechanism is that there is a selectional advantage to be had in possessing a dumb but relatively quick predator detector. Maybe, then, the mechanism delivers the information that “There's a predator”. Or again, perhaps the possession of this mechanism is advantageous for detecting conspecifics or mates, in which it gives the information “There's a potential mate”. Or, maybe we should describe the mechanism as a predator-or-mate detector.

Which functional description of our mechanism is the one we want? If we want to describe its function as a vertical symmetry detector then when it goes off in the presence of the vertical symmetry of Holden cars, the mechanism is functioning correctly. But if we describe the function of this mechanism as being a face detector, then the case of the Holden car's setting off the mechanism is a false positive.\(^1\) Under this latter description the mechanism has a very specific function, certainly more specific than the

---

\(^1\) Even if our detector fires in the presence of Holden cars we may still claim that the its *proper* function is to yield the information “someone is looking at you” since *proper* functions are independent of the actual uses to which something is put and actual breakdowns which cause the mechanism to never fulfil that function. See Millikan (1984 Ch. 1).
former description. That's why the Holden's setting off the mechanism constitutes a false positive in the one case but not in the other.

Now which function we ascribe to such a mechanism is going to determine how domain specific the mechanism is, relative to finer grained domains. If we call our mechanism a vertical symmetry detector then it is domain inspecific with respect to Holden cars and faces. If it is called a detector of when someone is looking at you, then it is domain specific with respect to that function, but domain inspecific with respect to detecting when your kin are looking at you.

If we are to decide upon the degree of grain of domain upon which a cognitive mechanism operates, we are going to have to come up with a theory of how to attribute functions. That is, we must determine the correct way to describe what a mechanism is designed to do. There seem to be a couple of competing views regarding what proper function we should ascribe to Braitenberg's mechanism. The first tells us to describe it as a face detector. Why? Because it is the selectional advantages which the swift recognition of faces endows on the possessors of the mechanism which explain the presence of that mechanism. According to this view, cognitive mechanisms wear their etiologies on their sleeves, as it were: the functions attributed to the mechanisms should itself tell us why the mechanism enhanced the survival of its possessors.

The second view takes more notice of the false positives generated by the mechanism when it is described in the manner of the first view. It advocates a coarse grained function to be attributed to our mechanism, such as vertical symmetry detector. On this description, the mechanism does not wear its fitness enhancing properties on its sleeve: why would selecting vertical symmetries be survival enhancing? This does not mean that there is no good explanation as to why such a mechanism exists. There is a full evolutionary story to be told in terms of the environment in which these mechanisms develop, where the vertical symmetries in question just happen to be faces, and it's their being faces which leads to the survival of the possessors of those mechanisms.

One reason offered as to why the first view should be adopted is that its preferred attribution of function gives one an explanation of why that mechanism bestows the survival advantage that it does: it's the presence of faces and their instant detection which explains that mechanism's presence. But what does this buy us over the previous explanation of why there is a vertical symmetry detector? It seems not much; both seem adequate responses to the question as to why there is such a mechanism. However, that question, viz. why such a mechanism exists, is a different question from the
one which asks in virtue of what does some mechanism have the function that it has. The crucial issue here, and the one that goes unargued by the defender of the first view, is that this latter question ought do the work of the former question. I see no reason this should be so. We can see the reason for this below, albeit somewhat circuitously.

The general issue of the close relation between function attribution and content is correctly recognised by Fodor (ATOC). It is often hoped by some of those working within the field of information theoretic semantics that the misrepresentation-disjunction problem will be solved in teleological-functional terms. Fodor, correctly I think, claims that such a strategy cannot break out of the disjunction problem for the reason that when there is ambiguity about the intentional content of a state of a cognitive mechanism, there is also an ambiguity regarding the function we wish to attribute to the mechanism-cum-state (ATOC pp. 16-7). The problem is that we are dealing with the same problem in the case of the disjunction problem and the attribution of function.

Now while it seems that Fodor is right about this connection between content and function in the cognitive case, for just the reasons cited in section 2.1, the moral he draws seems just wrong. He claims that "Darwin doesn't care" about which way describe Braitenberg's mechanism: describing the mechanism as a face detector means that we easily get false positives when the environment changes, whereas describing it as a vertical symmetry detector one can forsake the false positives. As Fodor puts it, on the former description, when error occurs it is the fault of the mechanism, whereas on the latter reading the world has gone wrong since it has been the detection of faces which led to the mechanism's enhancing fitness, and it's just an accident that there are Holden cars in the world.

As we have seen, Fodor is right that there can be a plausible etiological story to be told regarding the mechanism's being selected under either description. He's wrong, however, in claiming that there is concomitant indeterminacy of function here. There are obvious cases of function attribution to Braitenberg's mechanism which our two views do not support: it hardly counts as a grandmother detector since it's hard to fathom how that could enhance fitness, and there would be too many false positives elicited. Thus, our problem is to determine which of the plausible attributions supported by our views we should adopt.

We already have seen one reason for going for the second view over the first: on the first there is the confusion of the question of why there is such

\[^{11}\text{For more on the disjunction-misrepresentation problem see Dretske (1981 & 1986) and Fodor (1987 & ATOC).}\]
and such a mechanism, and the question of what function some mechanism has. Another reason lies behind the point of the previous paragraph. Darwin does care about how many false positives a mechanism generates, if those false positives affect fitness. Consider the case of our mechanism operating in an environment where there are many more vertical symmetries than mere faces. The question is: do the false positives generated with respect to its being a face recogniser constitute a selectional disadvantage. We can easily conceive of environments where such a mechanism would prove disastrous, with its dying out due constant interruptions to behaviour. Equally, we can conceive of environments in which there are only faces to detect, in which case there would be no false positive under either description of its function. It's the intermediate case that seems problematic: it might well be that it's the recognition of faces that leads to the mechanism's fitness enhancing properties, but the false positives with respect to that function are not disadvantageous. In such a case, the first view tells us that the mechanism is a face detector. The second view tells us that it is a vertical symmetry detector and that the environment sometimes sets off the mechanism in the absence of the feature of that environment which led to its endowing fitness enhancing capacities upon its possessor.

In short, what the second view adheres to is something like the following simplicity principle: one should attribute the coarsest grain of function that will allow some feature of the historical environment to explain the presence of that mechanism. Notice that if one adopts this principle, one will need to, independently of the function attributed, specify that feature of the environment which led to the rise of the mechanism; but that's okay, since the answer to the “why is there such a mechanism?” question will have to do that anyway. By following this principle, it seems that we get the best of both views: it is some particular aspect of the environment that brings about the existence of the mechanism, and at the same time we allow for our not being surprised at the thing's going off in the presence of other aspects of the environment.

Given the reliance of our concept of domain specificity upon the theory of functions, the question of false positives might lead one to be skeptical about the actual extent to which a mechanism really is domain specific. For example, if our hypothesised detector's proper function is to yield the information that someone is looking at you, then when it goes off in the presence of a Holden car, it operates upon domains other than the faces of conspecifics or predators. So: why should we claim that the mechanism is domain specific? The first reply to this question is that in the case of proper functions, it's consistent with the skeptic's story that a mechanism never
perform the function specified as its proper function, or even perform some other function (See footnote 10).

The second reply has to do with, again, the problem of error in the theory of representation. Just because we have a domain specific mechanism before us, we should not think that its domain specificity restricts it from making mistakes—from generating wrong information—where wrong information might be information about some other domain. Macbeth's domain specific phoneme analyser might well switch on when the wind whistles through the floorboards of the castle so that actually hears "Sleep no more." We would surely count this instance of the mechanism's performance as an error, without questioning the domain specificity of that mechanism.

4.3 Modularity and Domain Specificity

Fodor treats the eccentricity criterion with reservation; I think he should have totally avoided it. So, then why do we have domain specific cognitive mechanisms? Let's take faces as an example. If, as I have been claiming, faces are not an eccentric domain, then why might there be a domain specific mechanism for recognising them? Well, perhaps there might be some selectional advantage in having a separate mechanism for recognising faces. A story needs to be told as to how a domain specific facial recognition mechanism might maximise fitness. One might think that perhaps the supposed speed of operation and computational efficiency the input systems seem to possess have provided some fitness advantage to the populations exhibiting these perceptual systems.

Although issues of speed and efficiency—whatever kind of performance parameters these are supposed to involve—provide better reasons for why there are domain specific mechanisms, I should treat them with caution. There are two reasons. The first is that one of the ways envisaged to get speed and efficiency out of a cognitive mechanism is to have that mechanism undergo proprietary processing. As I've already made clear above, it's not evident that proprietary processing has terribly much to do with domain specificity. So, it doesn't follow that they have to be the reasons why there are domain specific mechanisms. This is exactly what I must say given the account of domain specificity I have offered. I'm claiming that any cognitive mechanism is in fact domain specific; and given that there might be slow cognitive mechanisms, we should not expect those mechanisms to be domain specific because of speed and efficiency requirements.

The second reason is more substantive. The chief reason why Fodor's input systems are fast, does not seem to be related directly to their domain
specificity. As Fodor himself suggests, it is informational encapsulation which generates the speed of modules. Now while speed and efficiency might well be good reasons for there being domain specific input systems, I don’t quite see why they should be reasons for there, generally, being domain specific mechanisms. If it turns out that there are non-modular, informationally unencapsulated domain specific mechanisms running on nonproprietary codes, and hence are not fast and efficient, then why would we have them? Again, the point is that we want to factor out the empirical case from the conceptual case: as a point of logic, we want to know what the reasons are for there being domain specific mechanisms per se.

The most plausible reason why there are domain specific mechanisms per se seems to be that it may only possible to have a complex cognitive system such as ours develop by there being a proliferation of domain specific mechanisms. What does that mean? My claim is in the spirit of David Marr’s principle of modular design (Marr 1982) (see chapter 1); it’s not that vision, say, deals with eccentric domains, but rather that the only way to encompass processing of all those domains vision requires—in a evolutionary plausible way—is by the building up of a large system by conjoining separate domain specific mechanisms. I should mention that Marr never mentions domain specificity in stating his principle. I think that my preferred account of domain specificity commits me—happily—to Marrtian modules (not Fodorian modules) being domain specific. The reason is that, as we saw in chapter 1, Marrtian modules are individuated functionally, and, trivially, these modules will be functionally specific. Assuming the medium-message distinction is in place, then that combined with Marr’s principle will deliver us the class of domain specific mechanisms. This has the consequence that central processes turn out to be modular in Marr’s sense; he probably wouldn’t have thought that, but it’s in the spirit of his proposal since it means that developmental changes in either input systems or central processes, say, won’t of necessity require developmental changes in the other. That fits perfectly with the claim of the revised account of domain specificity to the effect that central processes are domain specific.

5 How Much of a Distinction?

Now that we have, I hope, sorted out domain specificity and why there are such mechanisms, we may ask the question: How much of a distinction between domain specific and domain inspecific mechanisms are we left with? One of the signs of a robust distinction is the ease with which it can be wielded—how easy it is to decide between cases. It’s clear that we can decide
between some cases: those mechanisms which are functionally specific but whose specificity falls on the medium side of the medium-message distinction get automatically classified as domain inspecific. Be that as it may, that's not much of a distinction to make since it will mean that LTM and STM will turn out to be domain specific, which is contrary to the intuitions behind the examples of the introduction. So what does the concept of domain specificity provide in the way of an edge for our taxonomic blade?

As a last resort distinction, it seems that the proposed account of domain specificity still allows for a relativised distinction between domain specific and inspecific mechanisms. For any mechanism specific to the domain featuring in the description of its function, that mechanism will be domain inspecific to some other domains usually more fine grained than that which features in its functional description. So, our middle-sized object detector being domain specific with respect to middle-sized objects, is inspecific with respect to three-masted brigantines, kangaroos and Holden cars. Correspondingly, there are many other hypothesised mechanisms which would be classed as domain inspecific with respect to the domain of middle-sized objects—our central cognitive process would seem a candidate here. With this relativised account of domain specificity in terms of functional specificity, we get enough of a distinction to taxonomise cognitive mechanisms relative to one another; but no absolute distinction.

We are left, therefore, with cognitive mechanisms featuring somewhere along a gradient of ever increasing coarseness of function. At the present, that might not seem much, by way of providing a taxonomy of cognitive mechanisms, but the mapping of the architecture of the mind has to start somewhere.

Signpost

In this chapter we have been assuming that the worst possible scenario outlined in the previous chapter is not the case. Given, then, that The-One-True-Cognitive-Psychology can analyse cognitive systems at Level Two and at a cognitive level of description. That Level Two analysis will therefore be able to individuate the class of cognitive mechanisms or modules at that Level or level in terms their representational properties. We have seen that such an individuation will be based, in part, upon the cognitive functions the mechanisms perform and the degree of grain of domain upon which they operate. If the arguments in this chapter are sound, then there would appear to be no single point at which we can differentiate fine and course grained
modules. The class of cognitive mechanisms would not seem to divide into the domain and nondomain specific. It's all matter of degree, or so it seems.

Having done this, we now move on to utilise some of these results in identifying a particular view about Level Two cognitive descriptions. It is a view of cognitive architecture which, I will claim, underlies intentional realism. Having done that, we will, at last be able to return to the question of the tenability of intentional realism.
In chapter 4 it was argued that it is the degree to which complex systems exhibit certain *abstraction* properties which lies behind our describing those systems as representational systems, and, hence, cognitive systems. We also saw that these abstraction properties were Level One properties of those systems. Although the criteria for discriminating cognitive systems are at Level One, The-One-True-Cognitive-Psychology seems to be a Level Two (or decompositional level) enterprise (from chapter 1). We also saw in the previous chapter that at Level Two, it might be possible to individuate cognitive mechanisms whose operations range over coarse grained representational states according to the function we attribute to those mechanisms. By combining the results of these two chapters, I think there is an important theoretical position to be identified—it's importance emanating from the fact that it is a view widely held within cognitive science/psychology. The combination of these results yields a view I will call *cognitive state realism* (CSR). It is the view that representational states of a cognitive system are individuated at Level Two, and that these states are coarse grained representational states. The current chapter is devoted to a detailed exposition of this view.

As I argued in chapter 4, in order for a system to be described as a cognitive system, it needs to possess, to a certain degree, the proffered abstraction properties. To that extent, both infra-verbals and would seem to fit somewhere on the non-cognitive-to-cognitive scale. In so far as we would seem to exhibit these properties to the greatest degree of the cognitive systems of which we know, I am going to formulate cognitive state realism in terms of *us*, since it is clear that we are cognitive systems. Consequently, some of the examples by which I introduce the cognitive state realist thesis are framed in terms of *language* production. That, however, should not affect the substance of the cognitive state realist view I am putting forward.
Chapter 6

Cognitive State Realism 125

1 The Nature of Cognitive States

Cognitive state realism is the view that a Level Two description of cognitive architecture will postulate mechanisms which range of over states that can be described as representations that are coarse grained. These representations are called cognitive states. Cognitive states are deemed to have various properties, and as we shall see, we have encountered some previously, in chapter 4.

1.1 The Grossness Thesis

Because they are postulated as Level Two states of a cognitive system, cognitive states mediate the input-output of the system. Given some stimulus, the cognitive state will play a role in the production of behavioural output. In other words, cognitive states exhibit certain causal roles. For example, the type of causal role required will be one in which the cognitive state, call it $C$, is identified, in part by its actual and potential role in the production of an utterance in the agent's language. When an agent $A$ utters "$p$" at $t_1$, that same state $C$ would also be part of the cause of $A$'s uttering "$p$" at $t_2$—assuming sincerity, language competency, $A$'s desire to both communicate and make known one's beliefs, etc. In addition, the actual and potential causal role by which we specify $C$ will also involve the production of nonverbal behaviour. Suppose $p$ represented that tennis balls were green. Assuming that $A$ also possessed some motivating cognitive state, such as the desire to photograph something green, then $C$ would be typically cause $A$ to photograph tennis balls. We may call this component of the causal role defining a cognitive state the behavioural output condition.

The specification of our cognitive state will also include perceptual input conditions.1 When confronted with a tennis ball and $A$ sees that they are green then, ceteris paribus, $C$ results. The generation of $C$ might also be brought about by verbal transactions. B's uttering of "$p$" to $A$ might well lead $A$ to form the cognitive state $C$, provided we can tell some story about $A$'s trusting $B$ etc. As with the behavioural output condition, the perceptual input condition will be specified in counterfactual supporting terms.

I want to call the conjunction of both of these conditions cognitive state realism's grossness thesis. What the grossness thesis requires is that

---

1Stephen Schiffer gives a perceptual input condition $[P]$ in (Schiffer 1987 p. 31). Schiffer maintains that this condition cannot be met from the point of view of common sense functionalism. Schiffer seems to think that such conditions must be stateable in the theory which employs them. The role I envisage for such conditions makes no such claim. This question is tackled in chapter 7.
whatever one wants to call a cognitive state at Level Two, it must have, in addition to certain causal roles in the production of other such states, certain gross causal roles specified in terms of the perceptual input and behavioural output conditions. I use the term 'gross' because some single state actually gets caused by certain external stimuli and causes certain behavioural output. The grossness thesis ensures that the causal-functional role specifying a cognitive state is not restricted to some local, internal relations. The specification of a cognitive state must be specified in accordance with the grossness thesis because it is gross perceptual stimuli and behavioural output (both verbal and nonverbal) by which we make cognitive attributions.

At this point I should mention that the grossness thesis divides into a strong and weak version. The strong version claims that the very state which underlies the production of nonverbal behaviour, also underlies the production of our verbal reports. This is a particularly important feature of the cognitive state realist view, since our verbal reports are based upon introspected information. According to the cognitive state realist when we introspect the causes of our actions we take those causes to be the causes. That's why the same cognitive state that causes our behaviour, also gets to be the cause of our utterances—at least those utterances concerning our first person explanations.

The importance of this point has been emphasised by Stich in his *From Folk Psychology to Cognitive Science* (1983). Although he does not explicitly state the grossness thesis, in the final chapter, he takes folk psychology to be committed to a particular story as to the “gross architecture' of our cognitive system” (1983 p. 237). That architecture, in so far as it is committed to the grossness thesis, requires that our cognitive system have some overarching behaviour generating mechanism which works according to the interaction of intentional states—beliefs and desires paradigmatically. When we make our first person verbal reports as to why we did something, then according to the grossness thesis that verbal behaviour is also produced by that overarching behaviour generating mechanism. From this it follows that it is the same states we introspect in our verbal reports that are responsible for the behaviour to be explained.

Stich offers some reasons why he does not think this strong version of the grossness thesis can be met. His reasons are based upon some psychological evidence drawn from cognitive dissonance studies and attribution theory. Without going into any detail here, the evidence as presented by Stich suggests that there may be two relatively independent cognitive systems, one (largely unconscious) for mediating behaviour and another (mainly conscious) verbal explanatory system which makes *inferences*
(or often confabulations and post hoc rationalisations) from (a) what it takes to be the causes of behaviour generated by the former unconscious system) and (b) the subject's situation (Stich 1983 p. 236).2

If this line of thought turns out to be correct, then the strong version of the grossness thesis is going to be in trouble. If our cognitive system keeps two sets of books as Stich suggests, then it will not be the one cognitive state which underlies both behaviour and our verbal reports of the causes of that behaviour. The possible fact that we do keep two sets of cognitive books in itself does not cause problems for the grossness thesis. It might well be that the verbal explanatory system has reliable access to what is going on in the actual behaviour generating mechanism. The data suggests, though, that such access is not available to the verbal explanatory system, thus reducing its output to the status of rationalisation and confabulation.

Even if Stich's suggestion is vindicated, however, all is not lost for the cognitive state realist who believes the grossness thesis. There is a weaker version of that thesis which, instead of claiming that there is one state underlying verbal and nonverbal behaviour, claim that there is some Level Two state which features in the production of nonverbal behaviour. Now Stich's line of attack confronts the strong version of thesis only, leaving the weak version for use by the cognitive state realist. In the face of the evidence from the dissonance studies and attribution theory, our cognitive state realist may decide to keep the weaker grossness thesis and so get behaviour caused by gross cognitive states, while admitting that we might not have introspective access to the causes of behaviour (the cognitive state in question did not directly cause one's utterance). On this weaker version, the grossness thesis consists in there being some Level Two state satisfying the behavioural output and perceptual input conditions, but which plays no role in the first person introspective reports of the causes of an agent's actions.

1.2 The Modularity Thesis

There is a further condition, closely related to the grossness thesis, to be met in order for cognitive state realism to be true. This is the modularity thesis identified by Steve Stich: a cognitive system "is modular to the extent that there is some more or less isolatable part of the system which plays (or would play) the central role in a typical causal history leading to" either non-verbal

\footnote{For more detail than Stich gives see Nisbett and Ross (1980) and Nisbett and Wilson (1977).}
behaviour or assertions of belief (Stich 1983 pp. 237-8). How we construe the modularity thesis will depend upon how we construe the grossness thesis. The modularity thesis tell us that if it is the same state which underlies both verbal and nonverbal behaviour (the strong grossness thesis) then that state must be an isolatable part of the cognitive system producing the behaviours. If Stich’s two sets of cognitive account keeping books turns out to be true, then the modularity thesis will certainly apply at least to the states which underlie non-verbal behaviour (the weak version of the grossness thesis).

In order to understand what this thesis amounts to we need to clarify just what “some more or less isolatable part of the system” means. There are multifarious ways in which to take this, depending upon what one takes the system in question to be. Let us suppose that we have a functional description of the human cognitive system. Ignoring for the moment, whether that theory is verified as true, the states and processes postulated within that functional theory, to the extent that they are individuated within that theory at all, can be described as isolatable parts of the system. Now, what Level of functional theory do we have? If the functional theory is at Level One, then it is still possible to have a form of modularity occurring, where the modular states are not isolatable at Level Two, but only at Level One. Consider, again, Ramsey, Stich and Garron’s (from chapter 4) connectionist memory model. That model represents certain propositions as holding at Level One, and each proposition which we isolate can be thought of as being modular, just to the extent that we can distinguish between the propositions represented in the theory at that Level.

What cognitive state realism requires, though, is a sense of modularity stronger than this. On the required reading, what the modularity thesis commits the cognitive state realist to will go something like this. Suppose that a class of cognitive states are tokens in a central memory store. The sense of modularity required is that those states feature in the Level Two analysis in which that central memory store features. On that Level Two analysis the states may feature as states of the memory store featuring in a Level One analysis of that store. Alternatively, those postulated states might feature at a Level Two analysis of the memory store. On either of these scenarios, the kind of modularity involved requires that the states are individuated at Level Two. This is the sense of modularity introduced in chapter 2 and the sense which lies behind Stich’s propositional modularity of chapter 4.

---

3As Stich points out, this use of ‘modularity’ must be distinguished from the Fodorian use. Stich’s use is more closely aligned to the sense introduced in chapter 2.
The modularity thesis combined with the grossness thesis commits the cognitive state realist to the view that it is a state of the central memory store, say, that features in the production of behaviour. If that state is modular in the sense that it features in a Level Two analysis of the central memory store, then that very memory address which is the cognitive state-token must be accessed in the course of that cognitive system's producing behaviour. If the state is modular in the sense that it features in a Level One analysis of the memory store then that memory store, rather than any particular memory address of the store, must be accessed in the course of behaviour production.

1.3 The Physical Realisation Thesis

Now let's assume that our hypothesised Level Two functional theory is true: it's The-One-True-Cognitive-Psychology. That will mean that our functional theory is realised in human neural wetware. Yet another reading of the modularity thesis takes the isolatable part of the realised cognitive system to be some state of the realising wetware. This is a far stronger and more interesting reading of the modularity thesis. It takes the modularity thesis to be interpreted within a physicalistic framework. To that extent, it assumes another thesis to which the cognitive state realist is committed, which we may call the physical realisation thesis. According to this thesis, cognitive states are taken to be realised in states of the brain along the lines of the account of realisation given in chapter 2. When a cognitive state is said to be modular and gross according to the previously introduced theses, it is going to be an isolatable state of the brain which must exhibit these properties; there will be a state of the brain individuated which plays an overarching etiological role in the production of behaviour, and that state must play a role whenever the intentional or cognitive state it realises is postulated in the etiology of behaviour.\(^4\)

Fodor, the most famed cognitive state realist, seems to have this "brain state" sense of modularity in mind since he seems to argue for the physical realisation thesis. He says (in a paper jointly authored with Pylyshyn):

\[\text{the symbol structures in a Classical model [language of thought model—JF] are assumed to correspond to real physical structures in}\]

\[^{4}\text{I hope that at this point the reader does not think that my formulation commits the cognitive state realist to the view that there is a grandmother neurone. I take it that the brain states in question are not anatomical states but functional states of the brain.}\]
the brain and the combinatorial structure of a representation is supposed to have a counterpart in structural relations among physical properties of the brain. For example, the relation 'part of', which holds between a relatively simple symbol and a more complex one, is assumed to correspond to some physical relation among brain states. (Fodor and Pylyshyn 1987 p. 13)

Moreover:

the physical properties onto which the structure of the symbols is mapped are the very properties which cause the system to behave as it does. In other words the physical counterparts of the symbols, and their structural properties, cause the system's behaviour. (Fodor and Pylyshyn 1987 p. 14)

Now these passages suggest that it is not mere isolatability of a hypothesised functional state that is required in the sense introduced initially, but the isolatability of brain states is what the modularity condition is going to require. The suggestion is not, I take it, claiming that those physical states of the brain which cause the system's behaviour are anatomically isolated. Rather, in accordance with chapter 3, physical brain states will get individuated functionally—but it is still in some sense physical brain states which we are individuating. From the point of view of establishing The-One-True-Cognitive-Psychology, this stronger interpretation of the modularity thesis would seem to be required. Two reasons come immediately to mind. The first is that the states postulated by cognitive state realism are thought to be etiologically salient states. Mere hypothesised states in a functional description of a cognitive system don’t cause much behaviour in any of the cognitive state realist’s subjects. The second reason is that only on the second interpretation would we be able to verify that our cognitive psychological theory was in fact instantiated in us, by correlating the tokens from the higher order neurosciences with those of functional psychology.

Functionally individuated brain states might have as their components anatomically individuated brain states which feature in many different functionally individuated states and processes. For example, consider Martin Davies’ “brain state” notion of modularity: “For it might be that the matter involved in two tasks is the same, while the actual states of that matter may be used in the performance of two tasks are distinct” (Davies Unpublished p. 5). We are asked to imagine a field of cells that vibrate in two planes, with the vibrations in one plane featuring in one task while the vibrations in the other plane feature in another task. Or more plausibly, a neuron’s synaptic
connection is important for some functions while it's position in a field of
graded slow electrical potentials might be important for some other. What is
important in this sense of modularity is not which structural or anatomical
component does the work, but that some structural component does the work.
In this way the cognitive state realist avoids both (a) the weaker Level One
construal of the modularity thesis, by having the functionally individuated
states physically realised (according to the, necessarily Level Two, physical
realisation thesis), and at the same time (b) being committed to cognitive
states being physiologically modular states.

1.4 Abstraction Properties (Again)

In chapter 4 I argued that abstraction properties were Level One properties
of a system by which we judge whether or not that system counts as a
cognitive system. In introducing the abstraction properties there, I referred to
cognitive states in the examples. It is important to note that these cognitive
states were Level One states of the system. The systems in question exhibit
those states only to the extent that they exhibit certain patterns of input and
output. Cognitive state realism makes a stronger claim than this. It states
that the cognitive states in question feature at Level Two, and it is those
Level Two states which are abstract in the sense introduced. If a subclass of
cognitive states can be thought to be tokens in a memory store, then cognitive
state realism requires those tokens to be medium, stimulus, response, and SR
abstract.

For example, take our hypothesised memory token, which is, by
hypothesis, modular in the sense required by cognitive state realism. That
token will be formed, in accordance with the perceptual input condition of the
grossness thesis, as a result of vastly different stimuli types impinging upon
the system, and, hence, qualify as being stimulus (and hence medium)
abstract. Similarly, that token will feature in the etiology of many different
types of behaviour, in accordance with the behavioural output condition of the
grossness thesis, and, therefore, qualify as being response abstract. That
token will also feature in vastly different stimulus-response pairs, in which
case it qualifies as being S-R abstract. That token will also be domain
abstract, which will ensure that our cognitive processes, where belief fixation
occurs, say, are not fine grained or domain specific (in the sense of the
previous chapter).

That it is Level Two states which are the states to which the
abstraction properties are attributed, is the centrally most important aspect
of cognitive state realism. It was claimed in chapter 2 that cognitive theory is,
and ought to be, committed to a Level Two decompositional level of analysis of cognitive systems. It is only under such an analysis that it can explain the capacities of those systems. So, it's not surprising that the cognitive state realist attempts to explain the Level One capacities of the system by reference to the properties of Level Two states. The important question we must address is whether Level Two representational states of the system exhibit these abstraction properties. To the extent that cognitive state realism requires that they do, and if we have reason to believe that they don't, then cognitive state realism might be committing a Level mistake. In any event, even if one opts for Level Two abstraction properties in the way cognitive state realism seems to, though, it is not obvious that anything has really been explained, since taking that option will merely push the explanatory burden back down a functional level. In such a case, the cognitive enquirer will be asking the same questions of various Level Two states such as a central processor as she would of the system as a whole. Where will the regress end? The regress must end somewhere, and I think that it should end in the first Level Two analysis of the overall system. To that extent I think cognitive state realism is mistaken. I hope to show why in later chapters.

It is also in virtue of these abstraction properties that states of a central processor, say, can be coarse grained representations of states of affairs. We saw in chapter 4 how many of the states of affairs that we take to be in the world are really abstract—they are multiply realisable by many different lower level states of affairs. It is only by being able to abstract away from the specifics of stimuli in the way that their abstraction properties allow, that Level Two cognitive states (and even Level One states for that matter) can represent the abstract states of affairs (what Fodor incorrectly described as the nonnomic properties of the stimulus) that they do.

2 Abstraction

I have been arguing that cognitive states are in some sense abstract. What does that mean? Well, for starters, it doesn't mean that cognitive states are not concrete; cognitive states are states of the brain and what could be more concrete than that?5 In the required sense of 'abstract', to be abstract is to be got before the mind by an act of abstraction, that is, by concentrating attention on some, but not all, of what is presented. A complete

5Of course psychological states might well be abstract in this sense. Dennett, for example, might say something like this, but in so doing be forced to deny that intentional psychological states are states of the brain. As we saw in chapter 1 an intentional realist should not be happy with this formulation of her theory.
material body, a shoe, ... is concrete; all of what is where the shoe is belongs to the shoe—its colour, texture, ... are all aspects or elements included in the being of the shoe. But these features or characteristics considered individually, e.g., the shoe’s colour or texture, are by comparison abstract. (Campbell 1981 p. 478)

While the cognitive state realist wants cognitive states to be concrete as per this definition, the actual act of abstraction is just what is seemed to be required in the processes leading to the generation of a concrete cognitive state. The abstracting away from the particular details of the stimulus presented to the organism in perception means that radically different stimuli can lead to the generation of the same cognitive state-type. Likewise, since various behavioural response-types can be produced by a cognitive state, response abstraction would seem to arise from a cognitive state’s being abstract—with the occurrence of reverse abstraction as it were. Since the particular details of the response are pretty much irrelevant to the operation of the cognitive state in the production of behaviour, there can be variation in the specific output emanating from that cognitive state.

You might think this talk of abstraction is highly metaphorical and metaphysical. It is. I don’t have any account, much less a psychologically plausible account, of what constitutes the process of abstraction. I suppose whatever Locke (1690) had in mind might come close to what I need in introducing the concept of abstraction. I introduce it because it precisely captures the particular properties of cognitive states which cognitive state realist inspired cognitive theorists seem to have in mind. Take the central memory store token which represents Kind of Blue’s being recorded in 1959. This seems to be a paradigm case of abstracting away from particular details of a stimulus. The stimulus which lead to the formation of that token, the record cover, is flooded with transducer detectable properties, readily awaiting to stimulate my cognitive system. But the colour of the album cover, the position of the date of recording on the cover, the font type in which the album cover was written, while all being features going to make up the stimulus, are ignored. From my merely storing that token in my memory store, we cannot discover the properties of the stimulus which lead to that tokening. It is this sense of abstraction which the cognitive state realist requires.

\[6\text{It might well be that such an act of abstraction occurs during perceptual processing. Whether abstraction occurs in the process of perception or in later cognition—if there is to be a distinction to be had here—is irrelevant to the present concern.}\]
Viewed in this way, the abstract nature of cognitive states should not strike the reader as problematic. We come across it everyday in the use of language, since sentences are abstract in just this way. Just as a cognitive state is formed by a process of abstraction, so too are the sentences in a language which express propositions; the sentence ‘Canberra is 2000 feet above sea level’ tells us nothing about Canberra’s population, distance from Melbourne or depressing lack of urban squalor.7

3 Cognitive State Realism at Work

Thus far I have given an abstract characterisation of the thesis of cognitive state realism. I have not, as yet, given any specific examples of cognitive state realism at work in cognitive theorising. In this section I want to show how the recognition of the cognitive state realism view, and the account of cognitive states contained therein, can dissolve certain problems within cognitive theorising. We look, firstly, at a problem raised by Stephen Schiffer regarding perceptual input and behavioural output conditions. Secondly, we can look at an account of the mind, Stich’s Syntactic Theory of the Mind (STM), which assumes cognitive state realism, and see how, once one recognises that assumption, certain arguments against Stich’s account dissolve. Stich’s STM is not the only instance of a theory of the mind which assumes cognitive states realism. We will also be looking at the views of Jerry Fodor in Part III. Ultimately, I want to argue against the account offered by Fodor by arguing against CSR. This means that I will also be arguing against Stich’s account which I present below.

3.1 Perceptual Input Conditions

We saw above that cognitive state realism’s grossness thesis requires certain perceptual input conditions. It is important to keep in mind what purpose these conditions serve. One might think that if cognitive states require such input conditions, but at the same time are stimulus independent, then it’s going to be quite difficult to state in any complete way those conditions. This

7The abstract nature of cognitive states was in fact suggested to me by Devitt and Sterelny’s example from the philosophy of language (1987 p. 5). The question of the abstractness of cognitive states also cuts into some deep issues regarding how it is possible for such entities as sentences and mental states to exhibit informational content. This should hardly be surprising given that cognitive states must be able to support the properties of psychological states postulated by folk psychological discourse which invokes propositional attitude talk. As to how a signal can exhibit only some information see Dretske (1981).
is because almost anything can count as a stimulus to one's forming a cognitive state. Coded messages are a case in point. Reagan might form the belief that the Iranians are coming from either CIA code inscribed on his breakfast plate, or from a headline in the Washington Post. What do these stimuli have in common that might feature as part of a perceptual input condition (other than carrying the information that the Iranians were coming)? Stephen Schiffer wants such a perceptual input condition in place if what he calls "common sense functionalism" is to get off the ground. The condition would look something like:

\[
P \text{ "If there is a red block directly in front of } x \text{ and ..., then } x \text{ will believe that there is a red block in front of } x " (Schiffer 1987 p. 31).
\]

If we construe perceptual input conditions as providing a set of necessary and sufficient conditions for a cognitive state's formation, as Schiffer seems to want, then by the stimulus independence of cognitive states these conditions are going to be impossible to formulate.

But I ask, do we really want to construe input conditions in this way? Do we really want to know in advance what conditions will have to prevail if we are to attribute, to a Martian we want to describe as having a cognitive system, the belief, say, that there is a red block directly in front of he/she/it? We should construe the conditions required by the grossness thesis to be not really such a set of conditions at all. We ought construe them as claims to the effect that for any given cognitive state generated as a result of the organism's interaction with its environment, there is some perceptual input story occurring. On this construal, the cognitive state realist is not required to even give this story. Should we really care if my belief that Orson is fat is formed as a result of seeing Orson in the flesh or on an L.A. TV chat show? All the conditions "required" by the grossness thesis amount to, is that the typical causal history of a perceptually based cognitive state will have such and such a form. Not construing perceptual input and behavioural output conditions in the way that Schiffer does, is an important issue. We will have to return to it in chapter 7.

What the abstraction properties do, is undercut this need to impose strict conditions on perceptual inputs. Given that cognitive states are abstract in the senses introduced, it seems no surprise that strict conditions of grossness will create horror for a cognitive state realist, if the perceptual input conditions are construed in the way Schiffer does. Why the need for strong conditions of grossness is undermined, is particularly evident when the abstractness of cognitive states is granted. To think that strict necessary
and sufficient conditions are required, is to ignore the very abstracting away from the mere input and output details which would need to feature in those conditions.

3.2 The STM and Cognitive State Realism

In his (1983) Stich offers a theory of the mind upon which cognitive theory might be built, which he calls the syntactic theory of the mind (STM). We first look at the theory and how it is committed to cognitive state realism, and then how that commitment affects a popular criticism of the theory originally levied by Zenon Pylyshyn.

3.2.1 The Syntactic Theory of the Mind

According to Stich's Syntactic Theory of the Mind (1983 ch. 8), the cognitive states postulated at Level Two, which are defined by causal links between stimuli, behaviour and other cognitive states, are individuated by reference to a class of abstract (in the sense of being not concrete) syntactic objects to which they are mapped. These syntactic objects are individuated according to their formal properties and relations. The mapping between cognitive states and these syntactic objects is such that the causal relations amongst the former are mirrored by the syntactic relations amongst the latter.

We can best understand what these abstract syntactic objects are by thinking of them as playing the same sort of role as representations might play in a Level Two cognitive description of a system, except that they have no semantics—they play the role of representations but are not representations. These objects are probably constructed out of syntactically simple parts by use of a set of formation rules; the class of abstract syntactic objects can be thought of as a set of wffs (well-formed formulas) in a semantics free mentalese.8

We may also think of the mapping which exists between neurological states and these syntactic objects to be a syntactic analogue of the relation between an organism and the class of mental representations postulated by cognitive theory. There may be mappings from various kinds of neurological states and these syntactic objects to the class of abstract syntactic objects, but only if one takes the sentences of the language to be uninterpreted complex syntactic objects lacking a semantics. As we shall see in 3.2.2, it is not clear that Stich can claim this since what we count as syntactic depends upon semantic interpretation.

8Stich points out that one may, if one wishes, think of this set of wffs as a language, but only if one takes the sentences of the language to be uninterpreted complex syntactic objects lacking a semantics. As we shall see in 3.2.2, it is not clear that Stich can claim this since what we count as syntactic depends upon semantic interpretation.
state types onto the class of syntactic objects depending upon how many psychological state types are required in order to formulate a syntactic cognitive theory. Stich summarises that cognitive theory *might* (under a Panglossian picture of cognitive theory) postulate at least two cognitive state types which feature in the production of behaviour: belief-like states and desire-like states (these cognitive states cannot be beliefs and desires since they are not semantically evaluable—they have no content. For more on why this is the case see Part III). In such a case both cognitive state types may be mapped to the same syntactic object. Stich calls the cognitive state types quantified over by the STM B-states and D-states, for fairly obvious reasons. I will call the class of states over which the STM quantifies syntactic states. A generic label is useful since we do not know how many psychological state types cognitive theory is going to require.

No cognitive theory is going to be of service in the development of The-One-True-Cognitive-Psychology unless it generates generalisations. In the case of the STM the generalisations involving the causal relations between cognitive states and stimuli, behaviour, and other cognitive states will be specified in terms of the formal relations between, and the formal properties of, the syntactic objects to which the syntactic states are mapped. Some examples of the type of generalisations featuring in a developed STM might be the following:9

For all subjects S, and wffs A, if S has some F-state mapped to A, then S will come to have some D-state mapped to not-A.

For all subjects S, and all wffs A and B, if S has a B-state mapped to A → B, and S comes to have a B-state mapped to A, then S will come to have a B-state mapped to B.

For all subjects S, wffs A and B, and all behaviours Z, then if S has some D-state mapped to A and a B-state mapped to Z → A, then S will do Z.10

Of course, Stich offers such generalisations in the grip of Panglossian optimism. It might turn out that the syntactic objects to which cognitive states are mapped have formal properties which require vastly different generalisations.

---

9 I will assume that the cognitive states possessed by the subject are Stich's hypothesised B-states and D-states, with F-states being the syntactic correlate of fear.

10 Z → A* should be read—but not, of course, as part of the STM—as something to the effect of "doing Z will bring about A".
According to Stich, even though cognitive states, syntactically individuated, are realised by neurological states, and hence the STM adheres to the physical realisation thesis of cognitive state realism, it is not their neurological properties in virtue of which they feature in cognitive generalisations.\(^\text{11}\) It is the formal relations amongst the syntactic objects which determine the nature of the generalisations. The main reason for this is that the STM is designed to comport with the spirit of the traditional construal of functionalism. If the generalisations were specified in terms of neurological properties then a subject would require human wetware—and, a fortiori, human wetware of a particular type—in order to possess a cognitive system satisfying the generalisations of the STM. The point is an important one for the STM, for the reason that in claiming that there is a mapping from abstract syntactic objects to neurological states, the STM does not require that the mappings be the same across subjects or even across the psychological history of a particular subject. Different neurological states—and even nonneurological states—can be mapped in the relevant ways to abstract syntactic objects.

It should be obvious from this all too brief introduction that the STM is going to satisfy the grossness thesis. Syntactic states are going to have causal powers, with causal powers mirrored by the formal relations among the abstract syntactic objects. They have causal powers because the generalisations of the STM are formulated in such a way that syntactic states are defined in terms of causal role. Syntactic states such as B-states and D-states can be caused by the environment, can cause one another and can cause behaviour just in the way specified by the grossness thesis.

The Panglossian version of the STM outlined above, is also committed to the modularity thesis. It is so committed because it attempts to follow the generalisations of folk psychological theory, and, as we have seen in chapter 4, Stich also believes that folk psychology is committed to the thesis of propositional modularity. In following those generalisations, the syntactic states postulated by the STM will have to be abstract in the sense of cognitive state realism. In order to satisfy those generalisations, B-states will be formed in the face of diverse stimuli and feature in the etiology of diverse behavioural responses. If they weren't abstract in this way, then there would have to be various B-states and D-states featuring in the different transitions from input to output depending upon the stimuli and responses involved, but that is precluded by the generalisations. We can see more clearly the way in

\(^{11}\text{Of course, we know from chapter 3 that to the extent that the formal properties of the syntactic objects specify the functional role of cognitive states, then such formal properties might well be high-level neurological properties.}\)
which syntactic states are a species of cognitive state (in our new technical sense) with the following response to an objection raised by Pylyshyn.

Before doing that, however, the reader might perceive a potential problem in incorporating the STM under the fold of cognitive state realism. Cognitive states (both the lay sense and the technical sense introduced in this chapter) seem to be paradigmatically representational states. If the states posulated by the STM are syntactic, in that they are not semantically interpreted, then how can they be representational? The answer to this question will become clear in the next section, I think. But a word about syntax and semantics should be said before going on.

3.2.2 Syntax and Semantics

A lot of fuss is made about the distinction between syntax and semantics in the philosophy of psychology. The content of a propositional attitudes is attributed according to the semantic properties of that mental state, according to some semantic theory. It is claimed that only once this interpretation has been performed can one attribute a representational status to that state. Syntactic properties of mental states, on the other, are thought to be quite independent of this semantic interpretation. States syntactically defined will be only “uninterpreted wffs described syntactically in mentalese”.

However, it seems that syntax and semantics are quite intimately related. The practice of logicians to behave as though the syntax comes first and then an interpretation is applied, puts the cart before the horse. Arguably, a syntax is a simple, if not the simplist, description of a supposedly meaning bearing system, given its intended meaning. A syntactic constituent of such a system is that which makes some uniform semantic contribution to that system. What this means is that a syntactic item is taken to be a syntactic item because it stands in a signifying relation to some semantic interpretation. The influence of semantic (and, indeed, pragmatic) considerations upon syntax can be seen in the quest for a workable relevant logic. What could be more semantic-cum-pragmatic than relevance?

If this is right then it is hard to see how Stich’s STM can be a truly non-semantic, syntactic only, theory of the mind. Some so-called semantic properties are bound to have crept in. This probably happens in the stage of individuating the syntactic tokens: that syntactic state causes that behaviour just because it was formed as a result of that kind of stimulus. Once one has isolated the inputs, the state formed, and the resultant behaviour, one has already done some of the semantic work in cognitive theorising.
With these considerations in mind we can now move on to the next section.

3.2.3 Pylyshyn's Objection

I now want to suggest that having identified the abstraction properties of cognitive state realism, we are in a position to counter an objection raised against Stich's STM. The objection is that the STM must fail to capture certain generalisations which cognitive science ought to capture. Consider the now well known case of Mary's running out of a smoke filled building. Pylyshyn claims that:

It simply will not do as an explanation ... to say that there was a certain sequence of expressions computed in her mind according to certain expression-forming rules. ... It does not show how or why this behaviour is related to very similar behaviour she would exhibit as a consequence of receiving a phone call in which she heard the utterance “the building is on fire!”, or as a consequence of her hearing the fire alarm or smelling smoke, or in fact following any event interpretable ... as generally entailing that the building was on fire ... simply leaving them as uninterpreted formal symbols begs the question of why these particular expressions should arise under what would surely seem ... like a very strange collection of diverse circumstances, as well as the question of why these symbols should lead to building-evacuation behaviour as opposed to something else. Of course, the reason the same symbols occur under such diverse circumstances is precisely that they represent a common feature of the circumstances—a feature, moreover, that is not to be found solely by inspecting properties of the physical environments. (E.g., what physical features do telephone calls warning of fire share with the smell of smoke?) (Pylyshyn 1980 p. 161)

Given the above account of cognitive state realism, if on the syntactic story, one gets the same transformation of symbols occurring across diverse stimuli, then there is something common across the varied circumstances which is capturable syntactically. To claim that commonality arises only after the semantic interpretation of those symbols, is to gratuitously ignore relevant features of the very example offered by Pylyshyn himself. Of course, some semantic work has been done in deciding that we should adopt the syntactic description of the system that we have. The point I am making is that because syntax and semantics are closely related, Stich has every right to
describe the state identified in syntactic terms. He's just wrong to think that no semantic work has been done in the process of individuating his so-called syntactic states.

Stich interprets Pylyshyn here in such a way as to give him the benefit of the doubt. He takes Pylyshyn to be claiming that the common symbols arise only after Mary has interpreted the diverse stimuli. He then goes on to defend himself by accusing Pylyshyn of begging the question. But why give him the benefit of the doubt in the first place? Perhaps I'm just less charitable than Stich, but I think we should take Pylyshyn literally here; he says that there can be common uninterpreted symbols in the face of diverse stimuli and yet denies that we have anything in common upon which to build generalisations.

This lacking of charity has some accompanying baggage, however. If the STM theorist is to take the common syntactic story and run, she needs to answer the questions which Pylyshyn thinks are crying out to be answered, viz. "why these particular expressions should arise under ... diverse circumstances, as well as ... why these symbols should lead to building evacuation behaviour as opposed to something else" (1980 p. 161). The answer to these questions can be had simply by enunciating the properties of cognitive states generally. As I have just spent a good number of pages explaining, cognitive states abstract away from particular stimuli (they are stimulus abstract). A cognitive state, whether it's a syntactic state or some other state (such as a belief or desire) can be formed irrespective of what the actual nature of the stimuli happens to be. There are various cognitive state-types which interact in certain ways, in certain circumstances to produce behaviour appropriate (if the system is working properly) to those circumstances. Cognitive states are also response abstract. So, Mary might well have fled the building in a number of diverse ways: running down the fire escape, climbing down the fireman's ladder or jumping from the highest possible window. But the one syntactic state, because it is a cognitive state, can constantly have these diverse behavioural outcomes. The upshot here is that cognitive state realism does not differentiate syntactic states from semantically interpreted states.

Within the cognitive state realist framework, within which the STM falls, we have the apparatus in place to deal with Pylyshyn's questions. Once we recognise that cognitive states have the property of abstractness, we can see how to get cognitive commonality out of the environmental and behavioural diversity in the Mary example. Perhaps Pylyshyn could be interpreted as claiming that the "interpretation" which Mary or her cognitive system (a non-trivial difference if ever there was one—I'm not sure which way
Pylyshyn would go here) performs is the act of abstraction I introduced in the previous section. I admit that I have no precise account of what abstraction amounts to, but why should we think that it was necessarily a semantic process performed after we have some syntactically individuated? It seems that the only difference between what Stich is doing, and what Pylyshyn is doing, would be that, as it were, Pylyshyn wants to stick a label onto the state identified whereas Stich does not. If syntax and semantics are closely related, then it’s hard to see how Pylyshyn’s “semantic interpretation” after the individuation of the so-called syntactic state can be anything but labelling. If it is labelling, then the syntactic-semantic dichotomy as applied in the case of cognitive states amounts to a false dichotomy.

Pylyshyn himself recognises some of the abstraction properties I have outlined in this chapter. He does so under the auspices of detailing the nature of cognitive states. What he takes cognitive states to be, are, I'm sure, what I have been calling intentional states. If this so, then Pylyshyn's objection to the STM is understandable, even though misguided. He may have thought that it was something intrinsic to intentional states that they were S-R abstract, medium abstract, etc. That would explain why he thinks that one gets commonality out of diversity of stimuli only after semantic interpretation. But as we have seen, this is a mistake. One can hold onto states featuring in psychological processes once the set of abstraction properties are recognised, without those states being full-blooded intentional states adverting to the idea of content.

Signpost

I hope it is obvious from this chapter where we are. I’ve outlined a thesis which I claim to underlie much cognitive theorising. It is the view that there are Level Two modules which quantify over states which conform to the conditions posed by the grossness, modularity, and physical realisation theses. Whatever the actual list of mechanisms cognitive theory and The-One-True-Cognitive-Psychology come up with, the cognitive state realist claims that some of those mechanisms will quantify over cognitive states in this specialised sense. Because intentional states are a kind of cognitive state, it’s a necessary condition of intentional realism’s being true that cognitive state realism is true. Whether cognitive state realism is a tenable doctrine is the subject of Part III of this work, and it is to Part III we now turn.
Part III

Intentional Realism
Chapter 7

What’s Wrong with Functionalism?

The preceding chapters are designed to provide a theoretical framework within which we may go about attempting to answer the question posed in the Prologue regarding the development of the cognitive sciences. That question, remember, concerns the propositional attitudes and their role in a mature science of the mind—The-One-True-Cognitive-Psychology. Perhaps, though, that answer can be had without recourse to the fancy theoretical framework thus far introduced. As we have seen in chapter 1, for instance, Functionalism, as commonly conceived, is designed (a) to avoid species chauvinism in the specification of mental states—if the possession of some mental state required the possession of neural states then that would exclude species without neurones from having mental states—and (b) to provide some way of specifying what the relevant mental state types are—in terms of actual and potential causal roles. There is a series of arguments in the literature which purport to show that these two putative beneficial properties of Functionalism cannot be achieved. If these arguments are sound, then it follows that any theory of cognitive structure which relies upon the Functionalist programme getting off the ground, such as intentional and cognitive state realisms, will falter in the absence of another, better theory of the mind. This might strike the reader as grist for my mill, given that I don’t believe that intentional realism will turn out to be correct. However, there are two reasons why I am not enthusiastic. First, even though Functionalism as a reductive enterprise might fail, I do believe that some Functionalist theory will hopefully provide a method for the individuation of brain states, and, second, I think that these arguments against Functionalism plainly do not work. So, this chapter is about why these arguments don’t work.
The arguments against Functionalism I wish to consider take three forms.¹ The first denies that Functionalism solves the problem of species chauvinism, since whenever it does, it lets in too many systems as systems of mentation-cognition. According to this argument, Functionalism is impaled on the horns of a dilemma: either species chauvinism or liberalism prevails. Functionalism allows too few or too many systems to count as systems with mental states. The second denies that a specification of mental states in terms of causal roles will be possible, just because the conceptual data from which the causal roles are derived—common sense or folk psychology, or even a substantive psychological theory—will fail to provide the required specification of the mental states. The final argument is similar to the second, only it is less general. It claims that Functionalism's quest for the algorithm that specifies mental state types, where that algorithm itself is specified by reference to causal roles, will not be successful. We assess each of these arguments in turn. The general strategy in confronting these arguments will be to claim that on a suitable understanding of the Functionalist enterprise the objections fail. I do not claim that this reading of the Functionalist enterprise will be immediately accepted by many Functionalists; the interpretation of Functionalism which these arguments from the literature attack does seem to be accepted by many Functionalists, though.

1 Chauvinism or Liberalism?

The chauvinism or liberalism dilemma is classically put forward by Block (1978). The example used to illustrate the liberalism of Functionalism is that of the now famous China brain (1978 pp. 279). Pretend that a billion of China's inhabitants are provided with special purpose radios that allow them to connect to each other and some artificial body resembling ours, say, but which has no brain. In addition, all the connections to the body which would normally attach to a brain are attached to transmitters which are in turn connected to a subset of the radios. When one of the body's transducers fired a message is relayed from the body to one of the radios on to various other radios, eventually to be received by a receiver in the body which initiates some motor response on the part of the body. Let's even pretend that the functional organisation of the inhabitants of China and the connections between them mimic you for a certain time, i.e. it is functionally isomorphic.

¹There are many other arguments against Functionalism as an account of mental states other than propositional attitudes, in terms of absent and inverted qualia, etc. However, for present purposes I am restricting my attention to features of Functionalism relevant to its providing an account of the propositional attitudes alone.
with respect to you. Then, if Functionalism is true, that system is describable as possessing mental states, and indeed, the belief that this is a silly thought experiment, if that's what you are now thinking! But, the argument runs, such a system surely does not possess such mental states. Therefore, Functionalism must be false.

As Block himself admits (p. 281), the claim of this argument that the China brain could not possess mental states rests on only an intuition, and an intuition which runs perilously close to being question-begging at that. We need something extra in order to secure the point against Functionalism. That point is to be had from considering the fact that our neurophysiologically based functional organisation certainly does generate mentality. It is because the China brain lacks a neurological state description, and we know that in our case such a neurological state description generates mentality, that the onus should be on Functionalists to provide independent support for their intuition that the China brain generates mentality.

Not surprisingly, perhaps, I am going to offer some independent support for the Functionalist enterprise based upon some of the considerations of the previous chapters. For now, though, we need to look at chauvinism.

Block claims that the way to avoid the problems generated by the China brain and the Bolivian economy (1978 p. 315) would be to place some constraints upon the specification of the inputs and outputs to the functionally characterised system. One could specify the inputs and outputs in terms of neural impulses, movement of limbs or stimulation of transducers. The trouble with such descriptions, claims Block (p. 316) is that they are chauvinist. Moreover, if one tries to describe inputs and outputs in species neutral terms then Block claims that will bring on liberalism since the Bolivian economy has inputs and outputs, and they might correspond to the inputs and outputs of the cognitive system. What one would do in such a case is specify the inputs, outputs and states numerically: inputs $I_1...I_n$, states $S_1...S_k$ and outputs $O_1...O_m$ related according to the Functionalist theory. There is no guarantee, though, that such a neutral description is not isomorphic to the Bolivian economy! As we shall see below, the description of inputs is crucial, so crucial, in fact, that Functionalism can be saved by their proper description (see 1.3 below).

What is going on in the Block argument? A number of crucial things are going on, the most important of which can be summed up in the following two questions: (a) in the case of liberalism, to which systems are we trying to attribute mentality? And (b) what are the criteria of attribution we are employing? We can take these issues in turn.
1.1 Cognitive Systems and Agency

The case of the China brain is interesting because there are two systems at work in the example. The first system is the artificial body which is connected to the second system, the China brain. When considering this example we must keep in mind which system it is to which we are attributing mentality. It is, of course, the conjunction of the two systems to which we must attribute mentality. Taken by themselves though, we are certainly not going to attribute mentality to them. But when we are required to make a judgement about the status of that conjunction, our intuition about that conjunction will not necessarily be the same as that for the two independent systems.

Taken by itself, the ham-radio infested Chinese populus will not be attributed with mentality, in just the same way that the Bolivian economy ought not to be. It is the intuition about the China brain in isolation, not conjoined with the artificial body, that fuels the judgement that the conjoined system is not a cognitive system. The conjunction of the two systems, though, is one to which we might plausibly attribute mentality, since the only difference between the artificial system and us is that what is doing the causal work in the stimulation of the body is not contained within that body. If it really is the roles or functions performed that is important in the generation of mentality, then we should not demur from attributing mentality in this case. To ignore this point about roles is to beg the issue against the Functionalist.

This overlooks an important point, though, when it comes to the attribution of mentality. I think there is a principle underlying those judgements, one which seems to be contravened in the China brain case. We may call this principle the principle of agency. According to this principle we attribute mentality to a system when that system's behaviour can normally be expected to be unintentionally caused by states (either Level One or Level Two) of the system itself. A classic case of the contravention of this principle would be that of our brains being mere transmitters to a superior species, controlling our actions like puppeteers. In such a case it is obviously not states of the system that are causally efficacious in the production of the system's behaviour. But what if we move the puppeteers to inside the system? That would seem to generate the China brain example and yet not contravene the principle. It's here that the unintentionality of the causes comes into play. In the puppeteering case the puppeteers are not filling the unintentionally mediated roles specified by a functional description. They themselves are deciding to make the system of which they are part perform certain actions, actions they decide the system should perform.
It is important to note that in the China brain case there is a marked difference from the puppeteering case. Although the inhabitants of China are intentional agents, themselves possessing mental states, they are not performing the roles allocated to them as intentional agents in the way that the puppeteers are. Even if the population of China knew what their own task was in the cognitive economy of the system of which they are a part, and acted out of their desire to keep that larger system running, say, they are not deciding as agents the course of the system's behaviour. By performing the role that they are, they are not deciding to make the entire system move its arm or make an utterance. If they did, then they would be performing some other functional role in that system.

The China brain case does not, in reality, contravene the principle of agency since the system in question is not the artificial body; that system would contravene the principle. It is always in principle possible to avoid contravening the principle in any given case by redefining that system so as to incorporate some extraneous elements in the etiology of behaviour. In this way a system can be made to conform to the principle. If we initially thought that the artificial body, or us controlled by Martians even, were the system in question, then that system certainly does not conform to the principle, and there is no way that we would want to attribute mentality to it or us. What this eventuality would require is that we re-evaluate what kind of system we are—we would not think that we were cognitive systems.

Having redefined the kind of systems we are, we then still have to decide upon the intentional status of the components which go to make up the system. What I want to claim is that it is the possible transgression of the principle of agency which underlies our intuitions about the mental status of the China brain. So, if it's true that the Chinese populus, themselves possessing intentional mental states, are acting not out of their own intentions, beliefs, desires etc., but performing the dumb work neurones can do in virtue of filling the appropriate causal roles, then we must conclude that the principle of agency has not been transgressed. If that's the case, then I think we should reject Block's intuition and claim that the China brain does possess mental states. In this way Functionalism will not be essentially liberal in its attribution of mentality to complex systems.

---

2This point is similar to one made by Putnam (1975 pp. 434-39) in which he claimed that systems which decomposed into parts which are ascribable of mentality, should not themselves be ascribable of mentality. I think this restriction to be ad hoc. But it is a different restriction from that implied by the principle of \( \text{agency}\). According to this principle a system can decompose into sentient parts, just so long as those sentient homunculi perform non-sentient roles, as the population of China obviously do.
Chapter 7

What's Wrong with Functionalism? 149

1.2 Criteria for Mentality

A crucial problem for any Functionalist-based psychology is the level of generality that one wishes to capture—that is, how broad should the domain of psychology be? This is evident in the discussion of autonomy and reduction in chapter 3. It is also crucial for the considerations regarding chauvinism.

The China brain example, in attempting to prove liberalism, assumed that it is supposedly sufficient for the attribution of mental states to a system that system be functionally equivalent to us. There is also a putative necessity component to the argument for chauvinism. Block's idea is that functional equivalence to us is a necessary condition for the attribution of mentality. Block admits that maybe functional equivalence is a condition on the recognition of mentality, but fails to see how it could be a condition on mentality itself. His reason is as follows. Suppose there are Martians with whom we develop extensive cultural and commercial intercourse. We learn about their science and philosophy, read their novels and go to their movies. We then discover that their underlying psychology is functionally different from ours. "Should we" therefore, asks Block, "reject our assumption that Martians can enjoy our films, believe their own apparent scientific results, etc.?" (1978 p. 311). If we don't reject our assumption then it would appear that a functional organisation of a certain type cannot be required for the attribution of mentality.

Lurking behind this example is the belief that there is some condition other than functional isomorphism with respect to us which must be met in order for a system to be described as a cognitive system. If there were no such alternative condition then Block would not be able to claim that "it would be perfectly clear that even if Martians behave differently from us on subtle psychological experiments, they nonetheless think, desire, enjoy, etc.. To suppose otherwise would be crude human chauvinism" (p. 311). The criterion lurking here is one which allows both Martians and us to be described as possessing of mentality. If some Level Two functional organisation is too specific a description, then it must be a higher level Level Two description or even a Level One description which is criterial of mentality.

Now Block gives us no idea what he takes the requisite criterion to be. Nevertheless, the Functionalist might well decide that it is that description, at whatever level, which is criterial of something's being a system possessing mentality or not. The various functional organisations such as those of us and the Martians constitute realisations—or perhaps Pylyshynian functional architecture—of some more abstract descriptions, the possessors of which are attributed with mentality.
The point I am making here is the same as that made in chapter 4 regarding the level of description at which we decide when something counts as a cognitive representational system. I think it's an important point that Block fails to treat with enough respect. He thinks that specifying the functional architecture of a system at Level Two is criterial of our judging that something is a system which possesses mental states. He then shows how that criterion is inadequate by claiming that it is chauvinist. He can only do that, though, by employing a higher level criterion, Level One, say, which captures the relevant class of cognitively described entities in its net. I want to know what that higher level criterion is, and why the Functionalist cannot employ it in her programme.

Block recognises that one might be tempted to make the move that I prefer, that, maybe, Functionalism is a Level One enterprise, but claims that a simple example counts against the move. He then goes on to run the standard argument against the Turing Test. The machine imitating a human interlocutor seemingly possesses human conversational abilities, but works according to list-search principles. He claims that because the machine works according to these principles and it seemingly possesses the same linguistic inputs-outputs as us, we must claim that it has no mental states. We have already encountered this type of objection in the previous chapter and discarded it. It is not at all clear that the range of inputs and outputs, and the relations between them, are of a sort that is evident in a cognitive system to which we want to attribute mentality.

There is another difficulty with Block's attack on Functionalism which I wish to mention in closing this section. Block seems to assume in his attack based upon necessity conditions for mentality that the concept of mentality is robust enough for us to get criteria for it which will aid us in deciding whether Martians are attributable of that concept or not. However, maybe mentality is not such a concept. It may be the case that mentality is a highly graded and pragmatic concept whose conditions of application are vague and imprecise. It might be the case that we can decide that Martians possess mentality only to greater or lesser degree, when compared to us.

The situation confronting Block can be seen in the case of infraverbal mentation. If the Functionalist paradigm is supposed to give us criteria of mentality then it should provide us with a means of deciding whether certain non-human animal species possess mentality, species with which we have some phylogenetic commonality. We don't, however, have any Functionalist-inspired way of doing this. We have yet to make the judgement as to whether or not cockroaches have beliefs and desires. There are bound to be functional similarities, to a degree, between cockroaches and ourselves—we both have
perceptual and motor control mechanisms, for instance—but how much similarity is required in order to claim that they have mental states, that infraverbal mentation is not a self-refuting concept?

1.3 Inputs and Outputs

We saw above that the description of the inputs and outputs of a functionally described system is crucial to the problems confronting Functionalism. Well, what descriptions of inputs and outputs must the Functionalist employ? There is a possibility of being species chauvinist in the specification of the Level One description which we take to be characteristic of mentality. Does a system have to possess linguistic capacities, mobility, reproductive capacities, etc.? As I have claimed in chapter 4, the properties of a complex system we take to be the mark of the cognitive are rather more abstract than a certain range of actual behaviours. They rather have to do with the different ways of responding to various ranges of stimuli. The trouble with the abstraction properties of relations between inputs and outputs is that those very properties which I have taken to be a mark of the mental might well be too liberal as well. One might think that the transitions of inputs and outputs evident in cognitive systems which are S-R abstract, say, can be exhibited by the Australian economy. This seems a fair bet since a Japanese import might well be related to a variety of outputs from the country: an export or an international bill of exchange.

Block and Owens (1983) rightly point out that this is a major problem for Functionalist-based accounts of the mental. How can its impact be reduced? One way would be to invoke the principle of agency. The intentional agents which of necessity make up an economy act out of their intentional states; somebody decides upon receiving a Japanese import that an export or bill of exchange gets output.

A better way would be to specify the inputs and outputs of a system that avoids both liberalism and chauvinism. Consider the case of rocks. Why don't we attribute mentality to rocks? Quite often, rocks are deemed not to have mental states just because they don't behave—see Fodor (1987 p. 69). Complex systems such as rocks do, however, have outputs: erosion and heat radiation, for instance. The point is: why don't those outputs count as behaviour? Consider some not very complex organism, say a paramecium. Running the same kind of Block and Owens liberalism line should get the critic of Functionalism to say that the paramecium has some description in terms of inputs and outputs that makes it functionally isomorphic to us. However, that line is never run. Why? Because we know that the paramecium
is an organism responding to an environment in certain ways, ways that don't allow us to attribute it with mentality—due, I claim, to its not possessing the collection of abstraction properties to the relevant degree. In describing the paramecium in these terms, we have fixed a certain level of abstraction at which to describe its transitions from input to output. It is that level at which we judge that the paramecium is not functionally isomorphic to us. There may be some other level of description of the inputs and outputs of the paramecium in terms of its absorption of sunlight and chemicals, and its outputting of waste and oxygen, such that the processes of photosynthesis are isomorphic to the processes of cognition. But that is not the level at which we make psychological judgements about paramecia. If it were, then they too would count as cognitive systems.

It is at some level similar to that at which we describe the inputs and outputs of the process of photosynthesis at which we make the judgement about the transitions from input to output of rocks eroding. However, even if there is an isomorphism between the story we tell about the rocks, we are not going to judge that the rocks have mental states just because we realise that the level of description of those state transitions are not at the level of abstraction at which we make judgements about us or paramecia. It is for such reasons that we do not claim that rocks and paramecia are functionally isomorphic. Similarly, it is the reason why we demur from attributing mentality to economic systems or the Milky Way. Perhaps there is a gas cloud of galactic size which moved with such slowness that its time scale would be extremely slow by our standards (Putnam 1987 p. 88). We can grant that such a system might be a cognitive system not because we describe its inputs and outputs in astronomical terms, but in terms relevant to psychological theorising, as when we compare ourselves to paramecia and rocks.

What is this so-called level of abstract description at which we describe the inputs and outputs of a cognitive system? That's the really hard question in the present discussion. Whatever it is, it's the difference between describing the paramecium as functionally isomorphic to us, and describing it as functionally distinct from us. One might claim that it is the description of the inputs and outputs that are psychologically relevant: inputs that count as stimuli and outputs that count as behaviour. In short, the relevant level is the one at which folk psychology applies. When we are willing to describe the inputs as perceptions, then we have arrived at the correct level. In explaining the outputs of rocks and paramecia we have no need to appeal to beliefs and desires, in order to frame the explanations or predictions of the outputs. We would describe the inputs to a rock as perception and its outputs as behaviour only if we were forced to attribute folk intentional states to the system. But in
such cases we, obviously, do not have to: chemistry and geology will provide the level of description at which we can state these explanations. Describing the inputs in this way will not commit the Functionalist to any form of chauvinism, since the general class of perceptions does not have to include the visual or auditory perceptions of our species.

I admit that this is not much of an account of how inputs and outputs should be specified. As I said, this is the really hard question, and at the moment I don't have a worked out answer to offer. However, I think what I've said is enough to placate Block and Owens.

### 2 Schiffer

Perhaps the easiest way to generate an argument against Functionalism is to stipulate that it meet certain *prima facie* plausible, but in effect, unreasonable desiderata, and then decry it for failing to meet those desiderata. In essence this is what Block has done in his argument that Functionalism is chauvinistic. This style of "argument" is also employed against Functionalism by Stephen Schiffer (1987). Schiffer also claims that Functionalism (again whether one is dealing with common sense or scientific Functionalism) is required to postulate some necessary conditions in order to generate the specifications of the mental states quantified over by the Functionalist theory. Such necessary conditions will be, claims Schiffer, perceptual input conditions and behavioural output conditions. An example of an uncompleted perceptual input condition might be Schiffer's own example already mentioned in chapter 4:

\[
\text{[P]} \quad \text{If there is a red block directly in front of } x \text{ and } ..., \text{ then } x \text{ will believe that there is a red block in front of } x.
\]

It was argued in chapter 4 that it was the mark of a cognitive system that there were indefinitely many ways in which that system might come to be in a cognitive state. If that's so, then we should expect that no such perceptual input conditions are going to be forthcoming, if the system in question is truly a cognitive system. To insist that Functionalism must be able to come up with such conditions is to insist upon the impossible. Now it might well be the case that Functionalists *thought* that they might be able to come up with conditions, and if it is this (what I claim to be a mistaken) belief which Schiffer is calling into question then he is correct. The question is, though: does the Functionalist have to come up with those strict conditions? If
Functionalism does have to provide such conditions then the considerations of chapter 4 suggest that the Functionalist's task is an impossible one.

The demand that Functionalism must provide Schifferian perceptual input conditions is the demand that there be definitions of the mental states a Functionalist theory of the mind quantifies over—for our purposes, the propositional attitudes. This can be seen in Schiffer's attack upon what he calls commonsense Functionalism. Commonsense Functionalism is the view that our propositional attitude concepts are defined by reference to the common knowledge (either explicit or implicit) of the agents that possess them. In other words, it is commonsense platitudes regarding propositional attitudes which give propositional attitude concepts their meaning. Schiffer complains about this conception on two fronts. The first has to do with who has the access to these platitudes. He says: "If the meaning of 'believes' is determined by a folk psychology expressed by its use, then that theory must be one implicitly held by everyone who has the concept of belief" (1987 p. 31). But, claims Schiffer, it is clear that those who possess the concept of belief have no idea about defining that concept, however implicit the theory might be. This is especially evident, he claims, in the case of machines, extraterrestrials, and even Helen Keller and Ray Charles. How could they possibly have any idea of the perceptual input conditions defining belief?

The second front has to do with the likelihood of coming up with conditions which will be strong enough to define mental state concepts. Schiffer claims that even if there were some corpus of knowledge possessed by all those with the concept of belief, that knowledge would not be of a kind to generate definitions.

On both fronts, Schiffer imposes standards that Functionalism ought not to have to meet. As to the first front, take the case of ordinary grammatical competence. Many speakers of a language have no idea of the formal arrangements of the language even though the grammars descriptive grammarians generate are determined by the use of the speakers of that language. We don't say that there is no theory of the language just because Bruce Layman cannot articulate such a theory. The same goes for Functionalist definitions. The response to Schiffer's second front of objections is to deny that Functionalists must, in offering their theory, provide "definitions" in terms of sets of necessary and sufficient conditions. We saw in chapters 4 and 6 that in the cognitive system case to which Functionalist theories are going to apply, there are not going to be any such necessary and sufficient conditions because of the abstraction properties. So, whatever account the Functionalist is going to give, it won't be in terms of the definitions alluded to by Schiffer. It's impossible; we cannot get them.

Nor
should we even contemplate for a moment that we could get them. Remember
that the causal roles alluded to by the Functionalist are not only actual roles
but potential roles as well. The trouble with specifying counterfactuals in the
roles determinant of mental states is that they give us another reason to deny
that there will be a necessary and sufficient set of roles which are constitutive
of mental statehood, since there are indefinitely many counterfactual roles
that could so feature.

An alternative strategy for dealing with Schiffer’s objections is to re-
examine the way in which Functionalism hopes to individuate mental states.
In the discussions of Functionalism so far, and in the discussion of Putnam in
the next section, it is assumed that the Functionalist wants to give a very fine
grained taxonomy of mental states where the belief that p is differentiated
from the belief that q. It might be the case, though, as we saw in chapter 1,
that the Functionalist wants to taxonomise mental states more coarsely only,
so that we individuate believing that p as opposed to the desire that p. In
effect the Functionalist would be giving identity conditions (in the loose sense
of ‘condition’ I have just been advocating) for the so-called “intentional boxes”.
We can think of these intentional boxes as boxes in a flow chart representing
what Pylyshyn (1984) has called the functional architecture of the cognitive
system. If that’s right, then it will avoid Schifferian style of objections. There
simply will not be a Functionalist individuation of the belief that there is a
red block in front of me. Consequently, there will be no need for a perceptual
input condition of the kind required by Schiffer.

As mentioned in chapter 1, Fodor recognises that there are potential
problems for Functionalism’s going fine-grained, and steers clear of a
primarily Functionalist-based semantics for those reasons (although he
thinks that functional role might have some minor part to play in the
determination of content). Of course, having some way of fixing the contents
of the intentional boxes is crucial. For that one is going to need some sort of
semantic theory which will secure the intentional status of the boxes’
contents. I am not going to have much to say regarding what is the right
semantic theory; perhaps a causal theory similar to Fodor (1987) or Dretske
(1981) or Millikan (1989) will suffice. In this work I am not crucially
concerned with that enterprise, for reasons cited in chapter 6. However, I will
have a little something to say about semantics and content, as promised, in
the next chapter.
3 Putnam and Multiple Realisability

The most recent attack upon Functionalism can be found in Putnam's recent *Representation and Reality* (1987). Putnam has finally betrayed the doctrine he helped spawn; he has aborted his own conceptual child. Putnam offers three lines of argument against any Functionalist programme. The first is an argument from meaning holism. We considered that argument in chapter 1. The second is what we may call the argument from broad content. That will be discussed in the next chapter. The last argument, which is the subject of this section, we may the call the argument from the multiple realisability of functional states. Whether one believed in some form of "Turing machine" formalism of the computational states quantified over by Functionalist theory, or relied upon a David Lewis (1972) style (or even Putnam style—as of "Philosophy and Our Mental Life" (1975f)) of formalisation in terms of an implicitly held theory such as folk psychology (or in terms of a substantive psychological theory in Putnam's case), it is required by Functionalist theory that each mental state reduce to a computational state. Putnam claims, in a similar vein to Block, that it is false that there is one computational state shared by all physically possible systems to which we want to attribute the same collection of mental states.

3.1 The One True Algorithm and Levels

Putnam's idea is this. Some Functionalists (although not Fodor and Pylyshyn) think that a functional description of a system will generate a taxonomy of mental kinds fine-grained enough to differentiate those states according to their content. If system A is in some computational state x, and another system is in a computational state y, those systems are in the same mental state provided there exists some synonymy relation between x and y such both x and y mean p. If it's functional role that determines the meaning of mental states, then any two agents which differ in their belief set will turn out to mean different things by any state defined by the overall computational-functional model (1987 pp. 85-87). Or consider an agent within some linguistic-cultural context. Putnam claims that the "functional organisation" of any two individuals may not be exactly the same (p. 82). Suppose that there is "belief fixation" component to an inductive logic which is hardwired into us. It might be part of our functional architecture. Inductive logics can differ in their assignment of probabilities, and so belief fixation can vary across individuals. The upshot is that when two agents are deemed to believe
that there are kangaroos in the neighbourhood, there will not be anything functional-cum-computational-cum-physical in common (pp. 81-84).

There is an important addendum to Putnam's argument. As he argues in an Appendix, it turns out that there are, in effect, too many realisations of any system, the workings of which we specify by some computational formalism—machine tables or (folk) psychological theories (pp. 121-25). If this is true, then it follows that

the assumption that something is a "realisation" of a given automaton description (possesses a specified "functional organisation") is equivalent to the statement that it behaves as if it had that description. In short, "Functionalism", if it were correct, would imply behaviourism! If it is true that to possess given mental states is simply to possess a certain "functional organisation", then it is also true that to possess given mental states is simply to possess certain behaviour dispositions! (p. 124)

This is an important result for Functionalist theory. Where as Putnam takes it to be a reductio, I will argue below that the conclusion should be embraced. For now, though, back to the main argument.

Putnam goes on a great deal about "interpretation". Functionalism is described, in Putnam's words, as the Master Algorithm for Interpretation (p. 91). In this role attributed to it, Functionalism is a panacea for all one's psychological and even semantic ailments (p. 92), designed to provide not only an account of propositional attitude types (such as believing and desiring) but also a general account of meaning and reference! While I think that this is imposing too much theoretical work onto Functionalism, we can grant, for the sake of argument, the Functionalism-as-interpretation metaphor.³ Even though Putnam does not believe that there is such a master algorithm for interpretation, he is certainly no eliminativist with respect to propositional attitudes either (see 1987, chapter 4). He thinks that we do make propositional attitude ascriptions, and we make those ascriptions in the course of some form of interpretative practice.

³Having said that, I think contemporary philosophy of language does depend upon the Functionalist programme bearing a fair degree of theoretical weight. If the meaning of our words derive their meaning from the semantic properties of our psychological states, then the psychological story takes on a responsibility reaching further than mere psychological interests. However, I'm not sure that any Functionalist has thought that there was going to be a Functionalist theory of reference.
With this interpretative practice in place, Putnam is then able to claim the multiple realisability of computational states postulated by Functionalism:

...we are not going to find any physical state ... that all physically possible believers have to be in to have a given belief, or whatever. But now it emerges that the same thing is true of computational states. ... Physically possible sentient beings just come in too many “designs”, physically and computationally speaking, for anything like “one computational state per propositional attitude” Functionalism to be true. (p. 84)

We have seen in Part I that multiple realisation is a relation which holds only between levels—between a lower level and a higher level to be exact. Since functional organisations are multiply realisable with respect to Putnam’s interpretative practice, that interpretative enterprise constitutes a level of explanation-description. This point is crucial for two reasons. Firstly, we need to know at what level of analysis this putative interpretative analysis is supposed to be; and secondly, if some level of explanation is being employed in order to decide when to attribute propositional attitudes, one should be explicit about it, because, as we saw in examining Block’s argument, maybe it is a level that can be employed by the Functionalist so as to avoid the current objection. We examine these points in turn.

I can only presume that the explanandum of the interpretative practice of Putnam’s is the action of agents. In other words, the systems being interpreted are human agents, or intentional systems, as Putnam himself calls them—no doubt following the lead of Dennett (1987 & 1979). These individual intentional systems might form, through their interactions, some larger system—a society or culture—which might even include the natural kinds of the agents’ environment. However, it is the individual agent within that larger system which is the object of study under Putnamian interpretation. We might well have to locate that system in its environmental and cultural context, but it is that individual to which we are going to attribute mental states such as propositional attitudes.

Now the computational states—either boxes featuring in the description of the functional architecture or a fine-grained taxonomy of mental states which distinguished various beliefs, say, from each other—which the Functionalist wanted to attribute to such a system in order to explain the presence of propositional attitudes were presumably the highest level Level Two states of the system. If the system were a black box, the functional organisation would be postulated to explain the capacities of
the system. But as Putnam has just argued, that kind of state attribution cannot account for the propositional attitudes. Those states capturable by some Level Two flow chart representing the functional architecture are multiply realisable. The only alternative left, therefore, is that Putnam's interpretation must be a Level One analysis.

This result is significant, I think for two reasons. The first is that if the arguments adduced by Putnam are sound, and to possess propositional attitudes is to be capable of being attributed with states under a Level One analysis, then those arguments support the conclusions of chapter 4, in that the criteria of the cognitive, where we attribute cognitive states of which propositional attitudes are a species, are had from Level One. The second reason why the result is significant is that the result Putnam proves in the Appendix is just the result that it is the Level One properties of a system that count when one is considering whether the system realises some functional model. Putnam thought the results of his Appendix constitute a reductio of the Functionalist position. I think what that result shows is that many functional specifications are really Level One analyses of complex systems, when Functionalists mistakenly thought that they were doing Level Two analyses. That's a Level confusion if ever there was one.

Putnam assumes that his result could not be embraced by a Functionalist because it would make such functional descriptions behaviourist. However, behaviourist analyses of complex systems really are just a species of Level One analysis. Not all Level One analyses need be riddled with the problems associated with behaviourism. What Putnam's result shows is that Functionalist analysis is a species of Level One analysis, not that it is Behaviourist analysis. Describing the kind of analysis the Functionalist should be pursuing as behaviourist is mere mud slinging, trying to prove theoretical guilt by association only. Why is Level One analysis not mere Behaviourist analysis? For a start, Logical or Analytical Behaviourism wanted to define mental state concepts. I've argued in the previous section that a Level One Functionalist analysis of mental states will not be seeking definitions in terms of the inputs and outputs of cognitive systems. These forms of Behaviourism also wanted to define the mental state concepts in terms of behavioural dispositions. In so doing behaviourists wanted to eschew any reference to states of the system in their analyses. So, they would, for instance, analyse attributions of the form 'A believes that it is raining outside' as equivalent to 'If A were to go outside then she would carry an umbrella'. Level One analyses do not demand that states of a system be analyses away in this way. As we saw in chapter 2, one can advert to states of
a system in Level One analyses provided that they are individuated at Level One.

We should now look at Putnam’s argument applied to David Lewis’ version of Functionalism. Remember that on Lewis’ account, the mental states get specified by reference to causal roles between sensory stimuli, motor responses and mental states, where the causal roles can be gleaned from the platitudes of folk psychology. Even though mental states are included in this specification, and they can even be “internal states”, it does not follow that the specification is not made under a Level One analysis. As claimed in chapter 2, a system can go through state transitions in the production of output; but these are Level One state transitions. The Lewis story should, I think be interpreted as a Level One description of a system. If a Lewis style functional specification gets realised just when its predictions about the system’s behaviour come out to be true, as Putnam claims (1987 p. 96), which would be the case if Lewis’ Functionalism were pitched from Level One, then that is the notion of realisation with which the Functionalist is going to have to live (despite Lewis’ supposed reluctance).

Stephen Schiffer interprets, correctly I think, Functionalism in the way I have been advocating. He could not be clearer as to his interpretation:

We might have a black-box problem: we are given an input/output system (the black box) whose outputs are a function of its inputs and it’s internal, physical states; we have access to the inputs and outputs but know nothing about the nature of the internal states or of the causal laws governing them. Nevertheless, we seek a theory that will be explanatory and predictive of the outputs. To provide such a theory is to solve the black-box problem.

We might be able to solve the problem by devising a correct functional theory of the system: we might theorise that there are so many internal state-types the system might be in, which are related to one another, to inputs, and to outputs in such-and-such causal or transitional ways. If this theory is correct and detailed enough, it could enable us to predict the system’s outputs on the basis of its inputs, just as knowledge of a computer program may provide us with the ability to predict its outputs, even though we know next to nothing about its internal hardware. (1987 p. 24)

\[4 I think also that the minimal Functionalist account of the propositional attitudes employed by Jackson and Pettit (Forthcoming) reads Lewis and Functionalism in the way I am suggesting here. According to them folk psychology is a higher level of explanation than The-One-True-Cognitive-Psychology or the neurosciences since the various ways these lower level enterprises might turn out will be consistent with the Functionalist account of beliefs and desires.\]
According to this interpretation, a functional theory employed by a Functionalist will be unashamedly Level One. Given this interpretation, it is not surprising that Schiffer does not employ the multiple realisability of algorithms argument used by Putnam against Functionalism.

That just about concludes my criticisms of the Putnamian objection to Functionalism. Before moving on to some of his minor objections, I want to look at what I think are some of the implications for the reading of Functionalism I am adopting. If Functionalism with regard to propositional attitudes should be interpreted as a Level One enterprise in order to avoid the multiple realisability and chauvinism arguments, then it will follow that The-One-True-Cognitive-Psychology, which is a self professed Level Two enterprise, is going to be chauvinistic. This is, I think, an advantage, since it comports well with the considerations of chapter 3. We saw there that maybe domain-specific reductions might take place between various (Level Two) levels of explanation-description. If The-One-True-Cognitive-Psychology has already contravened the Functionalist’s ideal of nonchauvinistic multiple realizability, then a chief obstacle in the path of reduction is blocked, and the way is opened for a domain-specific reduction. Also, we saw that The-One-True-Cognitive-Psychology should not be thought of as either developmentally or confirmationally autonomous from lower enterprises such as the neurosciences. Since cognitive psychology is already chauvinist then allowing the neurosciences to play some developmental or methodological role will not detract from the multiple realisability desideratum of the Functionalist account of mental states such as the propositional attitudes.

Having said all of this, though, there still might be the theoretical possibility that the Level One Functionalist story I have been advocating is chauvinistic. This, again, is a point about the robustness of concepts such as MENTALITY. Block and Putnam want to attribute mentality to a system which could be said to possess mental states such as beliefs and desires. Now it might well be that systems to which we attribute beliefs and desires are only a subclass of all the possible “intelligent” creatures to which we might want to attribute mentality, or, at least, “intelligence”. What goes for Level Two states, in their being only one possible realisation of a Level One belief-desire model, might well happen to the Level One belief-desire model. Maybe there is some higher level than my proffered Level One analysis at which we attribute concepts such as mentality. It might well be the case that we would want to attribute mentality or intelligence to creatures to which we would not normally attribute propositional attitudes.
What this chapter shows, I think, is that traditional conceptions of the Functionalist programmes are wrong. If cognitive state realism and intentional realism require that mental states are conceptually analysed at Level Two, then I think the objections presented in this chapter have some force. But what I hope to have shown is that the functional description of mental states can be interpreted as being at Level One. It is then an empirical matter as to whether or not there is an isomorphism between the states at Level One which are definitive of mental states, and the states at Level Two over which The-One-True-Cognitive-Psychology quantifies.

That leaves us in the position of having to decide whether a Level Two analysis of our cognitive systems are going to turn out along the lines envisaged by cognitive state realism. That thesis, now, amounts to the view that there is such an isomorphism between Levels One and Two. Functionalism might be a Level One analysis of propositional attitudes, but nevertheless, there will have to be a functional Level Two analysis of cognitive systems in order to explain the capacities of those systems. The question, therefore, is whether that functional decomposition of a cognitive system matches up with the Functionalist Level One analysis usually employed by philosophers. Our task, then, is to take a look at some candidate Level Two decompositions. That's the task in chapter 9.

Where we are in Part III is this. We have just employed some of the results of the previous sections in order to stave off some quick ways of cutting the theoretical ground from under the cognitive state realist—we now know that it doesn't depend upon Functionalism as popularly construed. We have also seen in this chapter that the question of the so-called semantic properties of mental-cognitive states is a hot one, and many of the problems for intentional realism seem to stem from this very property. So before proceeding to chapter 9 and our look at Level Two cognitive architectures, we must first address the topic of intentional semantics.
One traditional conceptual argument against intentional realism has its roots in the philosophy of language. Putnam (1975a), in his now famous "The Meaning of 'Meaning'", argued for the slogan that "Meanings ain't in the head". What this amounts to is the claim that what our words mean depends upon the way the world is rather than just, say, the descriptions associated with those words in the minds of the speakers of those words. Two speakers could be in the exact same collection of physical states (or brain states, even), and yet refer to different things. Since terms with the same meaning must refer to the same things, the meaning of those words of the two psychologically identical speakers must have different meanings.\(^1\)

How does this relate to intentional realism? On the intentional realist story, intentional states are representational in just the same way as linguistic items. Intentional states have content, and it's in virtue of how they connect up to the world that they have the content that they do. In short, beliefs and desires have a "meaning" in just the same way as linguistic items. However, the intentional realist posits that intentional states are states of one's brain. If those states are in one's head, as it were, and meanings are not in the head, as it were, then how can intentional states be in one's head given that they have meaning? That, in a nutshell, is the challenge to intentional realism in terms of the semantic properties of the propositional attitudes.

\(^{15}\)Putnam (1975a) uses his now famous "twin earth" experiment in arguing for this position. The idea is that there is a molecule for molecule replica of earth somewhere except in one respect: where water (H\(_2\)O) is present on earth, some other substance (XYZ) is present on twin earth, called 'water' by its inhabitants. This means that two molecule for molecule (except for H\(_2\)O and XYZ) replicas of Putnam could be in exactly the same brain states, and use the exact same set of descriptions (the story is set before the advent of an advanced chemistry) for picking the clear liquid filling their oceans while Putnam on earth refers to water whereas his Doppelgänger refers to XYZ. Since words with the same meaning must refer to the same thing, Putnam concludes that the meaning of 'water' on earth and twin earth must be different.
The efficacy and influence of this challenge has been entrenched by, most notably, Burge (1979 & 1986). As well as explicitly extending the Putnam line to intentional states, Burge also argues that social facts such as one's community's linguistic practices in part determine the content of one's intentional states.

We may call the Putnam and Burge type argument the argument from broad content. Why 'broad'? Because linguistic items and intentional states refer, mean, or represent in terms of their relations to the world "outside the head or skin" of the speaker or agent. Things going on within the speaker or agent are, contrastingly described, "narrow". More precisely, there are really two types of broadness, what we may call synchronic broadness and diachronic broadness. The classic cases of broadness are synchronic. In these cases, some representational item only counts as a representation in virtue of its synchronic relation to the environment. An example of this might be an organism's perceiving a red apple. One can also be related to the environment diachronically. One's belief that there is a beer in the fridge, although its truth value depends upon how the world is now, is a belief about that beer because its diachronic connection to that beer, just like my belief that there is water in the lake is diachronically connected to water rather than Putnam's twin water. Quite often diachronic broadness consists in a causal history of some meaning or representational item, and an example to follow in latter sections of this chapter will employ this form of broadness.

There are basically two responses to the argument from broad content. The first is to reject the intentional realist enterprise on the grounds that its commitment to mind-brain supervenience fails. This is the tactic adopted by Burge, the later Putnam, and Pettit, among others. The second response comes from the intentional realists, most notably Fodor (1987). He accepts the broad challenge but answers that broad content is of no use to scientific psychology, and advocates the adoption of what has become known as "narrow content". Roughly, the narrow content programme urges that the contribution of the context determining environment upon mental states should be ignored for the purposes of scientific psychology.

In this chapter I want to argue that even to the extent that the argument from broad content correctly hits a target, the hitting of that target in no way counts against the intentional realist programme. To that extent, I don't think that the intentional realist needs to opt for the narrow content programme. For that reason, an examination of the narrow content programme is just redundant from the point of view of this work, and so I don't care whether that programme can be coherently spelled out. Basically, I am going to argue that the argument from broad content and its slogan
“beliefs aren’t in the head” does not count against one of the relata of the meaning or representational relation being “narrow”. That is, although the mental representations which bear the meaning relation possess that meaning according to some broad criterion, it does not follow that that bearer of the meaning is broad. Or, again, in terms of intentional states, contentful beliefs might be the beliefs they are just because they feature as relata in a relation to a broad environment, but it does not follow from that, that as relata they cannot be “in one’s head”. Spelling out this idea is the content of section 1. In section 2 I distinguish between two roles that broadness might play when it comes to taxonomising representational or meaning bearing entities. I argue that one of them is assumed by the argument from broadness, but it must be the other role that is required in the current debate. Finally, in section 3, I offer an analysis of the putative puzzle about how broadly individuated states can seem to supervene upon narrow states. It is here that the issue of the causal relevance of broadly individuated states is raised. I argue that the considerations of section 1 will provide a means for dealing with this issue.

I hope it will then be all too clear why the narrow content programme is unnecessary, and why the slogan “beliefs ain’t in the head” is an obscuring red herring.

1 Can “Beliefs Be In The Head”?

Words have semantic properties. In that way they are substitutable for ‘x’ in the relations: x means y, x refers to y and x represents y. When Putnam tells us that “meanings ain’t in the head” what he must mean is that any theory of such relations, as opposed to either of the relata, cannot be given solely in terms of the properties of one of the relata. That would seem to be impossible, since it would seem to misidentify what is to be explained; it gets the expIanandum of our theory wrong. Semantic properties such as meaning, reference, and representation are paradigmatic relational properties, and relational properties are specified in terms of the relation and its relata (how many relata being dependant upon the adimit of the relation). If this is so obvious, why does the argument from broad content strike so many as a surprising and new point in the philosophies of language and psychology? It should be just obvious that it could not be solely the intrinsic properties of one of the relata of an intentional relation that could provide an account of its relational properties. Putnam says:
reference is socially fixed and not determined by conditions or objects in individual brains/minds. Looking inside the brain for the reference of our words is, at least in the cases of the kind we have been discussing, just looking in the wrong place. (1988 p. 25)

This seems obviously right; but I don’t think anyone has ever claimed that one could look inside the brain for the reference of our words, precisely because reference is a relation.

1.1 Relational Properties

What’s gone wrong here? Perhaps people have mistakenly thought that the properties which are in fact relational are not relational. It’s hard to imagine that one could have thought that given the seemingly obvious relationality of semantic properties. It does seem obvious that a word’s meaning, for instance, does not depend upon the intrinsic properties of the word: different words, after all, can have the same meaning. However, it might have been thought that if representational mental states are identical with or supervene upon the states of one’s nervous system, then those mental states might have been thought to be intrinsic nonrelational properties, since the states of my nervous system are presumably intrinsic states (whether or not this is so is the content of section 3).

Perhaps, though, this is to mistake the object of the surprise. Even though an account of the representation relation is undoubtedly broad, the slogan “meaning ain’t in the head” might be interpreted as the claim that the bearer of the meaning, the ‘x’ in the relation, cannot be narrow. This is most clearly seen in the intentional state case, the slogan being “beliefs ain’t in the head”. To ensure the isomorphism with the meaning case, we must read the slogan as claiming that beliefs have the content they do because of some of their relational properties; the slogan tells us that some mental state x represents some state of affairs y. That seems fairly uncontroversial. However, there seems to be an alternative reading of the slogan in which it might be thought that the broadness of a belief’s representational status implies that that belief qua relatum of the representation relation—the bearer of the meaning or representational property—must also be broad, and hence “not in the head”. Now this is a surprising and interesting claim, and one that I think is assumed to follow from the former claim about the broadness of the representation relation. If you think that the former claim implies the latter, then there really is going to be something for the intentional realist to worry about, since not only must she give a broad
account of the semantic properties of intentional states, the bearers, *qua bearer*, of those semantic properties themselves cannot supervene on, or be identical with states of one's brain.

Well, does the implication hold? I am going to argue that it does not, and, consequently, the intentional realist should not be overly worried. However, one might object here that the implication must hold, since if the states are individuated broadly according to their relational properties, then the bearers of those relational properties must also be relational entities. This is a move made by Garfield (1988 ch.5), for instance. Because, for example, linguistic tokens have their meaning essentially, in terms of their relational properties, those tokens *are* "things in a conventionally constituted context" (Garfield 1988 p. 100). The thought seems to be that the implication *does* hold because there is only a bearer of the meaning because of its relationality. This seems to be the claim that the theory of meaning underlying the argument from broad content also determines the *ontology* of the linguistic or the representational domains.

I think that in one sense this is right, and yet in another sense it is wrong. There would seem no doubt that it is relational or broad properties which determine the ontology of the *representational* or the linguistic. However, to the extent that the tokens we describe as linguistic, and hence having meaning, or the mental representations which we describe as representational are individuated linguistically or representationally in terms of their relational properties, the bearers of those intentional properties might be ontologically individuated by reference to some of their *other* properties. How is this possible? The considerations which follow suggest the answer.

### 1.2 What Can One Dollar Buy The Intentional Realist?

We can learn something about the intentional case by looking at an uncontroversial non-intentional relational case. Consider the case of dollar coins. Suppose that I am visited by A who comes from a Central American tribe where there is no monetary exchange or the like. She sees me put a dollar coin into my pocket and asks: "What is that in your pocket?" I of course answer that it is a dollar coin. "Well, what is a dollar coin?" A persists. A couple of answers are going to be possible here. One answer will describe the physical properties of the dollar coins: shape and markings, chemical analysis, weight *etc.* We may call this the "narrow" explanation. But something can possess all those narrow properties and *not be a dollar coin*: replicas for display or forgeries, for instance. An alternative answer would be
to list the properties of the narrow explanation (perhaps not in as much
detail, but at least enough to differentiate a dollar coin from other coins and
random pieces of metal) and, in addition, a history of the proper production
(by mints) of units of monetary exchange, combined with some historico-
economic story regarding the practices of economic exchange. We may call
this the "broad" explanation, where the additional features contained therein
constitute the context. It seems that the broad answer to A's question is the
better of the two, since it will enable A to use a dollar coin within the
monetary system, and when she comes across the special edition dollar coin
with a different set of intrinsic properties, or indeed some other unit of
monetary exchange, she will also have an explanation of why that object is
also a dollar coin.

It's easy to see that there will be no strict narrow account of coins;
ye differ vastly in size, shape and constitution. In fact, it might be that it is
not even coins, or notes that function as the appropriate units of monetary
exchange. Perhaps locks of hair could have been used.

Given that our answers to A's questions are appropriate, the kind of
question A was asking was a conceptual one. It tells us what makes
something a dollar coin, or, more generally, a unit of currency. In these cases
it is certain relational properties of that something which make it the
something that it is. The broad explanation is intended to give some idea of
the criteria of application of the concept DOLLAR COIN or UNIT OF
CURRENCY to some entity. Now each of the entities to which we apply these
relationally construed concepts seems to possess a set of narrow or intrinsic
properties—it's that entity determined by narrow properties which we
describe as a dollar coin. It is because of these properties that A was able to
ask about the object, the dollar coin, in my pocket. Even though the
determinants of that object's being a dollar coin are broad, it does not follow
that that coin is not in my pocket, or that A's question is nonsensical. Why?
Because the location of the coin which is the bearer of the relational
properties is determined solely by another set of properties of the coin which
are not constitutive of its being a dollar coin.

My claim is that the case of intentional states is just like the case of
the dollar coin. Both possess relational properties, but these relational
properties can be born by some entity which is specified by other, often
nonrelational, properties.

It might be thought that since a dollar coin is a unit of currency in
virtue of its relations to The Royal Australian Mint and our community wide
socio-economic practices, the coin is not really in the pocket, since the dollar
coin somehow literally spans the intrinsically identified thing in my pocket.
and the rest of the surrounding context. I think this is a crazy view, and I
think that for same reasons that I think that a bearer of a relational property
can possess certain properties (such as “being in the head”) in virtue of their
intrinsic properties. Let’s see why.

1.3 Contextual Conferment

How an intrinsically specified entity can possess broad properties can be seen
once the doctrine of what I call context conferment is understood. Contextual conferment is the bestowing of certain properties upon an
individual entity or entities in order to facilitate certain goals or ends.

To see why dollar coins are contextual conferees consider a slightly
different economic practice. In this practice the unit of exchange must be
taken back to its point of origin or to a “point of exchange” in order to be
verified as genuine. Since it can be verified, possession would not constitute
ownership since the token unit of exchange’s exchange history could be kept.
For this reason one might as well not have some token which one transports
to possible cites of exchange (where someone might want to buy or sell
something) since one would have to take the token to a point of exchange
anyway. This system would undoubtedly be very secure; it would minimise
forgery, and it would hard to establish a currency black market. However, the
system would be extremely slow and cumbersome. It would only make sense
if it were possible to have the verifying and recording points of exchange just
about everywhere. Under this scenario, the point of having dollar coins
would be lost, since they were designed to allow exchange to occur wherever
they were exchanged.

The main point to notice about this tale is that the units that get
exchanged possess their properties of genuine units of exchange only in virtue
of being processed by an official point of exchange. It is these processes
occurring that provide a context in which those units are legal units of
exchange, and for that reason the system of exchange becomes cumbersome
and slow. The system can be speeded up if one confers upon the units, in this
case what are to become known as dollar coins, the powers of exchange
independent of being processed by the mint or points of exchange. In other
words, if one confers the properties which the processing system used to apply
to an individual or entity, when those processes are absent—ie. one claims
that dollar coins are “legal tender”—then the system can be freed up.

—Who knows, maybe the electronic banking revolution will enable these practical
problems to be solved in the manner of the spread of EFTPOS (electronic funds
transfer at point of sale).
It is this conferring of the properties originally provided by the context onto some entity which I call contextual conferment. What it does is to confer certain relational properties onto some object which is individuated, in part, by certain intrinsic properties such as being a coin of a certain size shape etc.. It is such requirements for contextual conferment, and it is a mere practical requirement, that makes it possible that there are forgeries. To make a forgery all one needs to do is create a fake conferee which does not have the relevant contextual connections. This possibility is one of the costs associated with contextual conferment, but it's a cost worth living with given the advantages of the context conferment programme.

One should notice that context conferment of relational properties upon an entity seems to confer a certain type of relational property; not all relational properties can be context conferred. Consider the relational property of location. An object has its location just because of its relations to other objects or a set of space-time co-ordinates. It would seem that location is not a relational property that can be context conferred since that would mean that an object could be at some location when, in fact, it was not in that location. Whatever the characteristics are of the relational properties which can be context conferred which distinguish them from relational properties such as location, I don't know. The point is that there is a difference between various relational properties, and we must keep that in mind when considering the relationality of intentional states. I will be claiming, of course, that intentional states possess context-conferrable relational properties.

One important difference between non-conferrable and conferrable relational properties is that the need for conferring certain properties upon an object in some sense determines some of the intrinsic properties of the object conferred upon. I possess the height that I do for reasons not much to do with my relational property of being the same height as Bruce. However, in the dollar coin case, the objects must be easily transportable and easily exchanged. A ten tonne rock will not satisfy the demands required by the conferring of context. Similarly, when a collection of objects have context conferred upon them, normally there will be a relatively small number of types of objects conferred. If the number were too great, then that would also militate against the reason why conferring was introduced: if all dollar coins were different, then we would have trouble determining if we were being offered legal tender or not.
1.4 Beliefs and Context Conferment

I want to claim that context conferment is ubiquitous. Being a dollar coin, a Holden car, made in Japan are all conferred relational properties. Linguistic tokens also possess conferred relational properties. Consider a linguistic analogue of the above fictitious dollar coin story, in which the language is an "os:ension only" language or, if that's impossible, a language in which tokens can be used only the presence of their referents. Such a language would be clumsy and unworkable, and really misses the point of what a language does. It's because the items in a language refer, that the referents of the terms need not be around when the term is employed. In this case, the linguistic item refers or means because of the context in which it is used, where that context must always be immediate. In this language one could not refer to Australia unless one were in Australia. In order to free up the language, in the same way that exchange needed to be freed up, the tokens of our language are context conferred. In this way, it is not the intrinsic properties of a token that determine what it means, but the relational properties conferred upon the token which determine its intentional status. Notice again that some of the intrinsic properties of the token have been especially chosen with an eye to the ends desired by context conferring: the tokens are constructed out of a limited alphabet, for example.

And ditto for intentional states. Imagine a complex organism which literally does not represent its environment—it has no memory capacity. Perhaps all it could do is believe and desire what it was currently perceiving. I suspect that a species with such limited capacities would not survive long. There is a limited sense in which such an organism does represent its environment: it represents its occurrently perceived environment. But it's a form of representation which is inadequate since its "representations" are literally tied too closely to their contextual objects. In order to count as a true representer the organism's cognitive structure would have to have context conferment upon its internal states. In this way some internal state of the organism will represent according to certain broad or contextual determinants when those determinants are not in evidence. I stress here that it is the internal states which are the relata of the representation relation. That this can be so is guaranteed by the presence of context conferment.3

3Each of the earlier examples I have cited has had the intervention of an intentional agent in the process of deciding that context conferment is required in order to achieve the desired ends. In the intentional state case the forces guiding the development of cognitive systems which possess intentional states will have to play the intervening role. Those forces, I presume, will be fitness enhancing or Divine.
As with the previous examples of context conferred relational properties, the intentional state case will also have certain of the conferred upon object's intrinsic states determined by the relational properties. Let's assume, contrary to the best evidence available, that individual neurones are context conferred. That is, each neurone represents some thing in the environment—one's grandmother, say. In the process of being context conferred some of the intrinsic properties of that neurone might change. The excitatory threshold of its synaptic connections might change; they change so that it fires in the presence of grandmother or when she is thought about.

This apparent effect of context conferment upon the entities which are conferred upon is crucial to the current concern. The theoretically puzzling feature of the intentional realist programme, which the argument from broadness plays on, is the one of determining how can context get into one's head or onto a dollar coin. What the discussion of context conferment is supposed to do is give some idea as to how this is possible. We will return to the explication of how context conferment achieves this in section 3.

2 Individuation vs Constitution

I think that the intentional waters have been muddied by a mistaken view of what the argument from broad content shows. One way to take the conceptual data evident in that argument is to think that context plays a role in the individuation of certain entities. It is only after consulting the context that we decide that we have dollar coins, since those hitherto intrinsically specified objects now have a set of additional relational properties. On this reading, the entities are individuated just because they have been context conferred. I will call this broad individuation.

The second reading of the argument from broadness takes the context as somehow constituting part of the relationally specified object. Broadness generates ontology, in other words. This is roughly the view attributed to Garfield above. The view seems to be that entities with relational properties “carry their context around with them”, in much the same way that an object with location carries its context around with it, except in our case that context is somehow part of the constituency of the object. This reading ignores the distinction between context conferred and non-context conferred relational properties.

By construing broadness as an individuation criterion I am claiming that intentional states, and, a fortiori, content bearing mental states generally, are, and can be, states within individual organisms. At the same time, these states are not individuated by a non-broad narrow semantical
theory. This view, which postulates a narrow ontology but broad individuation, has been dubbed, rather pompously, Naturalistic Individualism by Garfield. He describes the view thus:

Although the facts responsible for determining interpretations are on this account naturalistic (comprising relations of the organism to distal stimuli and objects), the phenomena that get interpreted (only internal states and processes), and the generalisations over them, obey the individualistic supervenience principles of the Individualistic [narrow—J.F.] Theory of Meaning. The meaningful phenomena that are the theoretical entities of such a psychology are still, on this account, individualistic states and processes within individual organisms. ... But Naturalistic Individualism tries to have its cake and eat it too. Having established the role of naturalistic evidence in the interpretation of events or states as representational, the Naturalistic Individualist turns around and asserts that, despite the essential role that the nonindividualistic properties of these phenomena play in their individuation, their nature—their identity qua psychological phenomena of particular types—can be specified individualistically. (1988 p. 94)

This is a fair description of the way I think the intention realist should spell out her theory. The only questionable point in the description is the claim that the individualistic supervenience base is adhered to by the view I am advocating. Garfield’s point is that if intentional states are individuated broadly, then it would seem that those states must supervene on the conjunction of the intrinsically specified bearers of the relational properties and the context. Dollar coins, it might seem, supervene not only the intrinsically specified bit of metal, but upon context which has been conferred upon it. I think this is not quite right.

Garfield goes on to argue that this balancing act cannot be maintained, since the arguments against intentional states being individuated narrowly also count against Naturalistic Individualism. The reason he gives is the very reason I objected to his description: broadly individuated states will supervene on context. From this he draws the ontological moral that “the correct way to describe these phenomena ontologically would be as relations between their bearers and their environments” (1988 p. 106). This is the exact move advocated under the constitutive broadness strategy. Notice, though, that the intentional state is now described by Garfield as a relation. Constitutive broadness ends up confusing intentional relata with the intentional relations in the way
described in section 1.1. Broad individuation with its commitment to context conferment is designed to avoid this confusion.

There is a puzzle here though. If an entity which has been conferred with context supervenes upon that context, then why isn't that entity ontologically or constitutively broad? The whole point of introducing context conferment was to get an ontology that was not broad so that certain practical goals could be achieved. The sense in which such an entity supervenes on its context is that qua representer, one of the relata of the representation relation, it supervenes upon that context. An individualistic or intrinsic ontology combined with conferred broad properties gives us those goals. But how do we generate that individualistic ontology? How can a broadly individuated entity ontologically supervene on the individualistic?

3 Supervenience

As mentioned in the introduction to this chapter, the argument from broad content is supposed to cause problems for mind-brain supervenience. Since X supervenes on Y iff there is no change in X without a change in Y, then psychological states can only differ if brain states differ. But the twin earth cases show that agents with identical brains can possess different propositional attitudes. The argument from broad content, as stated, is an argument about how we type-individuate mental states. If mental are broadly type-individuated as the argument claims they must, then it follows that any mental state token will be individuated broadly, just in virtue of being tokens which fall under that type. That means that the argument from broad content may be construed as an argument against token mental states being individuated narrowly or non-relationally. Supervenience is often thought to yield token identity. Therefore, the argument from broadness is actually an argument against mental state tokens being supervenient upon, or token identical to, brain state tokens, since brain states are thought not to be relationally individuated.

3.1 Levels of Explanation

Fodor asks (1987 p. 30) if there really is a problem about supervenience to be solved in the face of the argument from broadness. Ultimately, he thinks there is some challenge to answer, but thinks that the problem arises only if we 'believe in the narrow or nonrelational individuation of brain states. So, the seemingly easy way to avoid the problem for supervenience would be to
argue that brain states are relationally individuated. He also claims that relationally individuating brain states is just plain silly (p. 31). Now Fodor does not give any explicit arguments why he thinks this line is that silly. Even ignoring the possibility (mentioned in chapter 3) of functionally individuating brain state types, where functional individuation is a paradigm case of relational individuation, in fact, the relational individuation of brain states follows from the mind-brain token identity: if a mental token is a relationally individuated type of mental state, and that mental token is identical with some brain state, then we have relationally individuated a brain state token state.

This is still just plain silly, according to Fodor, since it means that the individuated brain states will then fail to supervene on molecular states. If one then makes the same move as in the mental state case, it will follow that molecular states must also be individuated relationally. And ditto for all the other levels of description down to the very basic level (if there is one).

Where the hell does it all stop?

The answer is that it both does stop and doesn’t depending upon what one is talking about. Relational individuation stops at the psychological level (or some neurofunctional level) when it comes to the individuation of the kinds at that level. Mental states types are individuated relationally with respect to context, if the argument from broad context is correct. However, neurobiology, molecular biology, chemistry, and physics might not individuate its kinds relationally with respect to context (they may, of course individuate kinds relationally, but not relationally with respect to context).

Relational individuation does not stop when we have a token mental state which is token identical to a state of the brain, a collection of molecules, and collection of the basic stuff of the universe. To the extent that this mental state token consists of, say, molecules, then that collection of molecules is individuated relationally for psychological purposes. Of course, that collection of molecules does not constitute a chemical kind, but does fall under a psychological kind just because the token identities which exist.

We can see this more clearly in the dollar coin case. That clump of minerals does not itself form a chemical kind, nor does the collection of physical particles constitute a physical kind. However, those collections do fall under one of the kinds of units of currency. It is these units of currency as a type that is individuated relationally. So, the collections of molecules or particles which make up the token unit of currency have been individuated relationally.
3.2 Causal Powers

If the claims of 3.1 are correct then the puzzle regarding the supervenient base of broadly individuated entities has been dissolved. There are, however, two more stumbling blocks in the path of broad intentional realism, both of which have to do with the causal powers of mental states. The first has to do with how mental states can cause behaviour as result of possessing the intentional content that they do, given that intentional content is individuated broadly. We can understand how putatively narrow brain states cause behaviour, but if mental states are necessarily broad, how can those contextual features affect the causal powers of internal causes of behaviour? In other words, how can broadly individuated states be causally relevant?

The second stumbling block has again to do with causal powers. Fodor claims that individualism is a thesis about individuating mental states only in ways that affect causal powers. Both me and my Doppelgänger perform the same actions, so we should attribute to both of us mental states with identical content. But the argument from broad content tells us that we should type the contents of our mental states as distinct.

We take these stumbling blocks in order.

3.2.1 Intentional Causation

The reply to this problem has been partially covered already in this section. Remember that what gets individuated according the principles of the argument from broad content combined with the idea of context conferment, is a state of the brain. Stated in this way it's difficult to see where the problem lies. Perhaps the problem is more evident with some different terminology. Block (forthcoming) describes the situation as a paradox, generated by the following three premises:

1. The intentional content of a thought (or other intentional state) is causally relevant to its behavioural (and other) effects.

2. Intentional content reduces to meanings of internal representations.

3. Internal processors are sensitive to the “syntactic forms” of internal representations, not their meanings.

The most obvious response to this set of premises is to claim that mental representations have their syntactic properties just because of their semantic
properties. We saw in chapter 6 that to a large extent semantics affects syntax. As we have also seen, Fodor, for instance, recognises that in order for intentional realism to work, syntax must at least mirror or "mimic" semantics. But it's no accident that this mirroring takes place in a mental representational system. The syntactic properties of the representations derive from the links to the object of the representational system, viz. the world which provides the context. We have already seen above that context conferment can bestow certain properties on the relatum of a broad relation, in much the way of the grandmother neuron's firing potentials being so affected. The broad/semantic-narrow/syntactic dichotomy is surely a misleading and false dichotomy.

To see why there should be no problem here, consider again the dollar coin. There is no analogous problem in the dollar coin case. The properties of the dollar coin that are analogous to that of a mental representation are its being used in legitimate economic transactions. In the same way that context ensures that the mental representations feature in the etiology of behaviour, the dollar coins feature in exchange transactions just because of their relations to context. However, in the case of the dollar coin we see no mystery, so we should not see any mystery in the case of intentionality either.

3.2.2 Broad Individuation

There is still a problem lurking in 3.2.1. The problem arises in cases where mental representations seem to have different meaning but their syntactic properties are identical. Both me and my Doppelgänger have mental states whose syntactic properties must count as identical since our behaviour is identical, but we represent different things. We thus arrive at the second stumbling block.

If the line I am running here is correct, then it follows that the intentional realist should adopt a broad conception of content in her programme. There is a problem with this though, inasmuch as intentional realism, to the extent that it is a view of how scientific psychology will develop, attempts to map the etiology of behaviour. However, the etiology of behaviour seems to be the same across me and my Doppelgänger, whereas accepting a broad criterion for the individuation of content leads us to distinguish between etiologies. How can such a tension be resolved?

One way to dissolve the tension would be to argue that the behaviours of myself and my Doppelgänger are different. It is possible to deny that our behaviours are the same, by claiming that I pick up a glass of water, where as my twin picks up a glass of XYZ. So, the mental states will have different
causal powers since they elicit different behaviours. However, as Fodor objects (1987 pp. 41-2), that would mean that context could affect the causal powers of our mental states without affecting our brains since token identity is assumed to hold, and, ex hypothesi, our brain states are the same. The claim here is that to type-individuate behaviours as distinct would be to give up the local supervenience of causal powers. Asks Fodor: “How could differences of context affect the causal powers of one’s mental states without affecting the states of one’s brain?” (1987 pp. 41-2).

I take it that “affecting the states of one’s brains” is to alter the properties of those brains. Now does my brain and my Doppelgänger’s share all their properties in common. Obviously not, since my brain is possessed by me and not my Doppelgänger. There is another difference: the information bearing states of my brain have transacted with water where as the states of my Doppelgänger’s brain have transacted with XYZ. That’s surely a difference, but the issue is whether or not that difference constitutes a difference in causal powers. The following suggests that it does.

I said above that we could think of the causal powers of a cognitive state as being given by its set of syntactic properties. Now as we saw in the outline of Stich’s STM above, the syntactic properties of a state will be derived from a functional specification, where that specification will be spelled out in terms of causal roles. Something has causal powers to the extent that it can fill certain causal roles. Now the causal roles that determine something’s causal powers should include both actual and potential causal roles (as should be the case with functional role semantics from chapter 1). While me and my Doppelgänger would seem to possess mental states which have the same causal powers in 1750, or whenever the pre-molecular time was in which the Putnam story was set, once our chemistry was developed and it was discovered that water was on earth but XYZ was on twin earth, then the causal roles of those states across me and my twin come apart. I would say: “When I said in 1750 that there was water in my cup, I was talking about H2O” whereas my Doppelgänger would say: “When I said in 1750 that there was water in my cup I was talking about XYZ”.

The immediate response to this is to claim that this does not show that the causal roles are different since if you switched me and my Doppelgänger just as the chemical discoveries were made, then I would have uttered what my Doppelgänger uttered, and vice-versa. All one has to do here, though, is to alter the counterfactual case to one in which the switch has been made but I and my Doppelgänger know of the switch. The causal roles of a state depend upon other states with which it interacts such as my believing that I am on
the same planet that I was one minute ago. The situation seems to be this. There are too many stories one can come up with where the causal powers do not differ. However, if we find an instance where they do, the supporter of broad content and individuation can maintain his view in the face of the Focorian objection, since she will have a case of causal powers differing.

Moreover, I think there is a strong independent motivation for going against Fodor, when one considers the issue of the meaning change of contentful states. Remember that in the case of the original Putnam story, we might have been willing to grant that me and my twin had different beliefs after the state of chemical knowledge increased. What Putnam had to convince us of was that the meaning of ‘water’ did not change subsequent to those discoveries. One of the advantages of the broad individuation is that we can see how the meaning of ‘water’ in both my language and the language of Doppelgänger remains the same across the chemical discoveries post 1750.

Let us consider the question of meaning change under the narrow individuation of Fodor. Looking at me and my Doppelgänger synchronically, as it were, in 1750 there does seem to be a pull to the view that the causal powers are the same (although I have just given some reasons why I don’t think they are). So we can claim that they have the same narrow content. However, what happens to the narrow content of our respective mental states over time? When the chemical discoveries were made it seems that the causal powers of one’s earlier mental states have changed, since we will now be disposed to make different utterances about the referents and truth values of our long held mental states. What this means is that Fodor can have his individuation if he wants, but if he does, he will be forced into claiming that the narrow content of beliefs and desires change depending upon certain discoveries regarding context. Since contentful mental states have their content essentially, this means that the mental states over which scientific psychology quantifies will change. Fodor has said nothing about how the narrow content of mental states can change over time. He needs to tell such a story. In sum: it might well seem counterintuitive to some that we type the broadly individuated mental states of me and my Doppelgänger as distinct in 1750, but it is going to be equally as counterintuitive that some of our mental states change meaning after 1750, according to the narrow content programme.

While I think these considerations ought to be telling against the narrow individuation opted for by the intentional realist, there is another line which also seems compelling. Fodor gives the broad individuator two options: she can either give up mind-brain supervenience (yuk!) or go broad on the
individuation of brain states. What he does not allow is the broad individuator to split the difference:

If supervenience be damned for individuation, it can't be saved for causation. Burge says that "local causation does not make more plausible local individuation" (p. 16), but he's wrong if, as it would seem, "local causation" implies local supervenience of causal powers. Local causation requires local individuation when so construed. You can have contextual individuation if you insist on it. But you can't have it for free. Etiology suffers. (1987 p. 42)

I suppose it is the point of the doctrine of context conferment being advocated here that one can, as Fodor puts it, split the difference. The reason is that context conferment can affect some of the locally supervenient causal powers of the object conferred upon, or more generally, in context conferring, we choose an entity that has a range of actual and potential causal powers that suits the need from which we context conferred in the first place. There is also a sense in which there are two senses of causal power at play here. In the brain state case, we know that the states conferred upon will have certain causal powers at the neurological level of description. We can even assume that they have those powers because of their intrinsic properties. However, because those brain states are token mental states, they also possess certain causal powers with respect to the psychological level just because they fall under a psychological kind. Now it's context conferment that can bestow such psychological level causal powers onto mental states which are token identical to brain states, and those brain states are chosen just because of some of the causal powers they possess.

Signpost

The thrust of this chapter is the claim that since there is no broad-narrow currency distinction in the case of the dollar coin, to the extent that the two cases are analogous, there should also be no need for the broad-narrow content distinction. To that extent there is going to be no problem for intentional realism in terms of mind-brain supervenience. To be sure, broad individuation will lead to a taxonomy of mental states different from one's

---

I'm assuming here that causal relations can be exhibited at many different levels of explanation-description. That in itself is a controversial thesis, and argued for by Peter Menzies (1988) and argued against by Braddon-Mitchell (Unpublished). For the sake of argument I'm assuming it here since Fodor also assumes it in his (and Pylyshyn's) (1987).
initial expectations. However, if certain counterintuitive results are to be avoided, and we are not going to confuse intentional relations and relata, this unexpected taxonomy generated by broad individuation will be all too easily tolerated.

Having, hopefully, accomplished that task, it is time—at last—to find out what I think is wrong with intentional realism. Basically, I think it is the cognitive state realism component of the view that is the culprit. So, it is an examination of that thesis to which we now turn.
Chapter 9

Two Theories of Cognitive Architecture

If the traditional objections to intentional realism don’t work, as I have claimed, then what line of criticism should we adopt? The strategy we should employ is to argue against intentional realism by rejecting the thesis implied by intentional realism viz. cognitive state realism. What kinds of argument might be summoned against cognitive state realism? I don’t think there are any apodeictic a priori arguments that will work. Ultimately, the truth or falsity of this thesis, as with any in the discipline, will be empirically established. So, as well as examining some of the more conceptual style of arguments against cognitive state realism, we may also look at what might constitute empirical arguments against it, i.e. what type of evidence might count against cognitive state realism.

Cognitive state realism is a thesis about cognitive architecture. It postulates cognitive mechanisms which range over representational states with the properties listed in chapters 1, 4 and 5. One way to cast doubt on whether the cognitive state realist account is the correct form of architecture is to survey some other theories of cognitive architecture. In this chapter I wish to compare and contrast the architecture of cognitive state realism with another theory of cognitive architecture. The alternative theory we are to examine is based upon the idea of a vertical faculty as introduced in chapter 5; we may call it vertical faculty theory (VFT, for short). In order to get the clearest idea of the underpinnings of that programme, I think we should ideally examine in detail the work of the originator of the idea of a vertical faculty, Franz Gall. That, however, is a basically historical task, tangential to our direct concerns. I, therefore, direct the reader to Appendix A for those underpinnings. Instead, we can move on to the major tenets of vertical faculty theory, some of which we have come across in earlier chapters.
1 Vertical Faculty Theory

Jerry Fodor has dubbed Gallean faculties "vertical faculties" and the traditional philosopher's faculties—ones consistent with cognitive state realism—"horizontal faculties". Gall agrees with the philosophers that the mind is composed of functionally isolatable subsystems. He disagrees as to how the division should be drawn. As has already been intimated, Gallean vertical faculties are distinguished by reference to their subject matter—i.e. they are *domain specific*. As we saw in chapter 5, domain specificity, as such, will not provide a taxonomic edge in the way Fodor envisages. Instead, vertical faculties are distinguished by their taking as their domain of operation, domains which are relatively *fine* grained. Cognitive state realist architecture, on the other hand, ranges over mechanisms whose domain of operation is coarse—they answer a wide range of questions in the performance of their functions. Vertical faculty theory and cognitive state realism are, then, alternative ways of slicing the same cognitive pie.

By way of formalising this alternative approach to cognitive architecture, we may follow Fodor and summarise the main ingredients of Gallean slices of the cognitive pie thus: vertical faculties are *domain specific* (have a fine grained domain of operation), *genetically determined* and realised by distinct neural structures. Fodor adds a fourth ingredient: he claims that they are *computationally autonomous*. That vertical faculties are realised by distinct neural structures is central to the historical Gallean programme; but is *not* a component of vertical faculty theory. The same goes for the innateness of vertical faculties. For present purposes domain specificity and computational autonomy are the crucial properties of vertical faculties we need to explore. As we have already examined domain specificity in some detail (in chapter 5), we may take a look at computational autonomy in the next section. Following that I will say why the questions of neural realisation and genetic determination are not relevant to our current concerns.

1.1 Computational Autonomy

What Fodor means by computational autonomy seems to be that not only are vertical faculties distinct in the function they perform, but they are also independent in the performance of those functions (Fodor 1983 p. 21). He says: "That we can, most of us, count and chew gum at the same time would have struck Gall as a fact that offers significant perspectives upon our mental organisation" (p. 22). But Gall also wanted vertical faculties to *interact*. He says: "We have to discover the fundamental powers of the mind, for it is only
these that can have separate organs in the brain" (Quoted in Marshall 1980 p. 23). That there are separate powers should not lead us to think that they cannot interact. Non-fundamental powers or capacities result from the interaction from the fundamental and distinct faculties. He says: “Indeed, we believe that the total nervous system is a combination of many; that all these individual systems differ in their office; ...that more or less of a bond, and therefore of reciprocal influence, exists between all the individual systems” (Marshall 1980 p. 24). The crucial point here is how the individual faculties “differ in their office”. Gall would not want the destruction of one faculty to impede the performance of any other fundamental faculty, and yet he wants interaction between the faculties.

We can take the idea of the autonomy of vertical faculties two ways. The first is the claim that vertical faculties are autonomous in the sense that a faculty is an actual functional mechanism, with some specific function to perform. In Pylyshyn’s terms, the vertical faculty would have to appear as part of the functional architecture of the system under analysis. Or, in terms more in keeping with the terminology of this work, the vertical faculty will feature as a module within the Level Two functional description of the system.

I suppose ultimately we will require more complete individuation conditions for what constitutes a cognitive mechanism than these very abstract characterisations. Such conditions, though, are required not only by the vertical faculty theorist but any theory of cognitive architecture, including cognitive state realism. So, not much hangs on the formulation of those conditions for our current concerns. I read Gall as claiming that faculties’ “differing in their office” amounts to autonomy in this sense. That vertical faculty theory is committed to this sense of autonomy is incontrovertible; any cognitive theory—including cognitive state realism—must be so committed. However, the other sense of autonomy, the sense intended by Fodor, should not be thought of as part of the vertical faculty programme, for the following sorts of reasons.

The sense in which a mechanism can be “independent in the performance of its function” is grossly unclear. Suppose that some vertical faculty processes outputs from some perceptual system and our faculty fails to operate because of damage to either that perceptual system or the faculty’s connection to that system, then should we claim that our putative faculty is not computationally autonomous? The functioning of our faculty would seem to “depend” upon the perceptual system. Or perhaps computational autonomy amounts to a vertical faculty’s not having to “share computational resources” whatever exactly they are. Such resources might be information about the
world, where two faculties compete for that information stored in some memory. The case for this sense of computational autonomy of vertical faculties is not prima facie obvious given certain psychological data. The kinds of data I have in mind are attentional studies. The idea is this. If faculties really are computationally autonomous, then attentional studies in one modality should generate results independent of results in other modalities. This, however, is far from being obviously the case. According to some studies there is a correlation in the ability to execute auditory and visual attention demanding tasks (Earl Hunt 1985 p. 20). There also exist data from studies of interference patterns affecting simultaneous task execution. Even when tasks would seem to involve different faculties, interference is almost always observed. Admittedly, there are some exceptional cases. Balancing, for instance, interferes with visual memory but not verbal memory. Interference would also seem to be reduced by learning. 

1.2 Localisation and Equipotentiality

Central to Gall's programme was a commitment to cerebral localisation. Historically (see Appendix A), the debate between vertical faculty theory and the antecedents of cognitive state realism took the verticality-horizontality dispute to be if not the same dispute as the localisation-equipotentiality dispute then, at least, to be dependent upon it. If it turned out that equipotentiality of the brain ruled the day, then the vertical faculty theory was a goner. Hence, a great deal of time and effort was, and still is, devoted to determining the extent, if any, to which cognitive functions are realised in particular regions of the brain. However, these two disputes really are independent disputes; and for that reason VFT is not committed to the sort of localisation of function its historical antecedents had in mind.

The crucial thing to realise here is that there are various localisation theses. The most general is that which claims mind-brain supervenience or identity. Gall certainly needed to push this line (see Appendix A). Another thesis is that cognitive mechanisms are localised in neuroanatomical states. Language comprehension is sometimes thought to be localised in the first temporal gyrus (Wernicke's area) and language production localised in the third frontal gyrus adjacent to the Rolandic motor strip (Broca's area). It's this localisation thesis to which VFT is uncommitted. If the contention of chapter 3 is correct, then, although vertical faculty theory is construed as a psychological level model, it is going to be strictly realised in some high level

\[^{1}\text{Hunt 1985 p. 20.} \] Actually, these studies concern input systems or Fodorian modules and not faculties as such. These modules will be described below.
neurofunctional states—or perhaps even organs—of the brain. However, these states are not anatomical states in the mould of the traditional localisation-equipotentiality debate.

The reason why VFT is not committed to anatomical localisation of function is that whatever story one comes up with regarding cognitive architecture, that story, is going to be, in principle, consistent with the equipotentiality of the brain holding. That's because the localisation-equipotentiality dispute is a dispute at the implementation level. What holds for VFT here, also holds for cognitive state realism. The cognitive state realist's metaphorical intentional boxes might well be localised in particular anatomical regions of the brain as well. This is in fact just what the old faculty theorists claimed, although they localised horizontal faculties in the ventricles rather than the cortex. What is at issue between vertical faculty theory and cognitive state realism is the nature of the cognitive mechanisms which may or may not be realised in anatomical organs, the former opting for fine grained mechanisms, and the latter coarse grained mechanisms.

Granted the conceptual independence of these disputes, just what is the relation between them? I take it that there is an evidential relation between the disputes. For suppose it turned out that the localisationists win the day, then the functions localised will support either VFT or cognitive state realism, depending upon the functions. Suppose, on the other hand, that the equipotentialists win. That won't in itself support cognitive state realism over the vertical faculty account. It just means that the implementation level is not going to provide a way of deciding between the two theories of cognitive architecture. The reason for this is that if localisation turns out to be true then it follows that we must have some idea of what the functions are that are localised, while it's not clear that, assuming equipotentiality, we have to have a completed story of the functions of the brain.

So, the fuss about localisation consists in the hope that by examining the disruption of neural structures we might get some idea as to what functions should feature in our cognitive model. This is exactly the task of the cognitive neuropsychologist. Of course, the task of the neuropsychologist is not an easy one. A correlation between a lesion in one area of the brain $x$ and loss of a particular cognitive function does not imply that area $x$ is the neural centre responsible for that cognitive function. $X$ might perform some necessary but not sufficient role in the utilisation of a cognitive capacity (Patricia Churchland 1984 p. 143). For example, one might have the capacity to verbally communicate even though one cannot do so due to the impairment of the motor control mechanisms which feature in speech production. It might also be the case that cognitive functions are not correlated with any one area.
of the brain. So a lesion to one area automatically causes deficiencies in other cognitive abilities. Hence, the brain lesion-behavioural deficit correspondence does not simply guarantee that we are determining the locality of a cognitive capacity. Despite these theoretical constraints upon the cognitive neuropsychologist’s programme, all seems well for the disconfirmation or confirmation of our two theories of cognitive architecture.

1.3 Innateness

It is assumed by many proponents of VFT style theories that vertical modules must be innate. The following considerations suggest that they ought not be. Referring back to footnote 7 in chapter 5, we saw that Fodor changed his mind about there possibly being a chess playing module. In his (1983) he claimed that there is no way that a modularity theorist would want to claim that there is a chess playing module. One reason for why there might not be a chess playing module is that it is hard to see how a mechanism specific to chess playing could result from innate structure. In his (1987a), however, he allows for there being “brute force” modularity in addition to modularity “in the nature of things”. If there is a module (“in the nature of things”) for chess, then it is not a standard Fodorian module. In addition to not being innate (as standard Fodorian modules are supposed to be), a chess playing module is not a perceptual mechanism at all. For that reason, there not even a prima facie case to be made for its being a module of the standard Fodorian type. If anything counts as a higher cognitive process, chess playing would seem to. Chess board perception might well have an accompanying specialised mechanisms but chess playing would seem to require more intelligent processing than mere recognition of positions.

In virtue of what might the chess playing module develop? The answer to this question lies somewhere in the expert-novice literature from psychology—eg. Larkin, McDermott, Simon and Simon (1980) and Chi, Glaser and Rees (1982). Perhaps what separates the expert from the novice in some cognitive task (playing chess or solving physics problems) is the emergence of certain complex productions consisting of condition-action pairs. When that condition is met then the action is automatically triggered without the need for vast amounts of processing which would slow down the agent’s processing time of the problem under consideration (Larkin et al. p. 1337). More metaphorically, is suggested that the expert’s cognitive system executes instructions in a compiled rather than an interpreted manner. If this is right there is no need for an interpreter—whatever it is—to execute one instruction at a time; execution after compilation has the advantage of running many
instructions together rather than separately. These productions might be the mechanisms underlying the phenomenon of *chunking*. Chunking is thought to be in part responsible for expertise in domains such as problem solving. If a chunk is generated in STM (short term memory), then that one symbol stored actually represents many pieces of information about that task, which in the novice would have to be stored separately—that's why it is a chunk (Newell and Simon 1972 p. 781). The advantage of chunks for the expert is that reaction time is decreased and amount of information stored is increased.

Assuming this conclusion, at what level of grain do chunks feature? Now chunking is a task relative phenomenon—in fact, extremely so. The mechanism of chunking only operates in the domain of expertise, to the extent that while the expert’s performance is enhanced by chunking in typical situations, (remembering the positions of chess pieces from an actual or possible game, say) that performance drops off dramatically when the situation becomes atypical (the chess pieces are arranged at random) (Larkin, McDermott, Simon and Simon 1980 p. 1336). This suggests that there is no general cognitive capacity evident here that cuts across various problem domains, contrary to what one might think if there were generalised cognitive mechanisms (a central processor, say) responsible for higher cognitive processes. Since problem solving ranks as a higher cognitive process, and if there is the close connection between information and cognitive mechanisms claimed in chapter 5, then the facts about chunking suggest that the mechanisms responsible for those cognitive processes can be fine grained in the manner required by VFT.

Fodor himself suggests that it is a form of informational encapsulation which generates this “in the nature of things” sense of modularity:

Chess playing, by contrast, is modular in the sense that only a restricted body of background information (call it chess theory) is relevant to rational play even in principle. This second kind of modularity...is interesting to the engineer, however, since informational encapsulation makes for feasible simulation regardless of what the source if the encapsulation may be. (Fodor 1987a p. 36)

There is no requirement that the process of the formation of vertical modules goes the way Fodor suggests; although it is consistent with the example model from the expert-novice literature, where the formation of productions in chunking allows processing without other parts of the cognitive system needing to be consulted; and that amounts to encapsulation.

All this suggests that there are non-perceptual cognitive tasks that can be performed by mechanisms which can be described as modular; and
that is all the VFT requires. The mechanisms are modular to the extent that they are required in our functional decomposition of the cognitive system in order to account for the data of cognition, in this case the differences between experts and novices. Because such mechanisms are not brute force, they are not innate. Even if it turns out that there is no chess module, the above discussion has provided the conceptual possibility of non-innate vertical modules; since I am currently listing the essential elements of VFT, that possibility is all we currently require. The claim of VFT is simply that whatever the set of mechanisms postulated by cognitive theory, those mechanisms will not be coarse grained.

1.4 VFT and Fodorian Modularity

We have seen that in The Modularity of Mind Fodor offers a tripartite functional taxonomy of cognitive processes. Transducers provide the interface between an organism and its environment. Distal environmental objects cause the impingement of energy at various surfaces of the transducers. These patterns of proximal stimulations are then converted in some lawful way into neural code—the transducer's output. Input systems "mediate between the transducer outputs and central cognitive mechanisms by encoding the mental representations which provide domains for the operations of the latter" (Fodor 1983 p. 42). What this means is that input systems perform an inference making role: they take the transducer outputs which specify proximal stimulation patterns and "deliver representations that are most naturally interpreted as characterizing the arrangement of things in the world" (Fodor 1983 p. 42). However, the output from input systems is only going to give the organism the way the world looks, sounds or feels, rather than the way the organism believes the world actually is. The organism's beliefs about how the world is should be constructed out of (inter alia) these input system outputs in conjunction with background information about how good the seeing, hearing or touching is. So the operations of the input systems should not be identified with belief-desire fixation since what we believe depends upon this very background information. According to Fodor, then, it is central processes, those mechanisms which access the informational outputs from input systems which perform belief fixation.

The properties of these input systems, or "modules" as Fodor calls them, need not be spelled out in detail here.\footnote{For the record they are: domain specificity, informational encapsulation, mandatoriness, speed, being associated with distinct neural structures, exhibit characteristic and specific breakdown patterns, their ontogeny exhibits characteristic}
they make input systems a member of the class of vertical faculties. To the extent that Fodor is committed to the verticality of modules, he is an adherent of VFT. However, as we have seen Fodor also thinks that there are central cognitive processes whose domain of operation is cognitive states. Now central processes are domain inspecific; so to this extent Fodor adopts a mixed cognitive taxonomy: partly vertical and partly horizontal. VFT, as I envisage it, in fact committed to the verticality of much more than perceptual input systems; VFT claims that higher cognitive processes are performed by domain specific cognitive mechanisms, that is by mechanisms which do not operate over cognitive states. The arguments for this position feature in the section 3. By way of rounding off our discussion of the nature of vertical faculty theory, I wish to give some idea of what the implications of accepting the VFT are going to be.

2 Implications of VFT

Let's return to Fodor. He says:

We have suggested that the characteristic function of modular cognitive systems is input analysis and that the characteristic function of central processes is the fixation of belief. If this is right, then we have three ways of taxonomising cognitive processes which prove to be coextensive:

FUNCTIONAL TAXONOMY: input analysis versus fixation of belief.

TAXONOMY BY SUBJECT MATTER: domain specific versus domain neutral.

TAXONOMY BY COMPUTATIONAL CHARACTER: encapsulated versus Quinean/isotropic.

I repeat that this coextension, if it holds at all, holds contingently. Nothing in point of logic stops one from imagining that these categories cross classify the cognitive systems. If they do not, then that is a fact about the structure of the mind. Indeed, it is a deep fact about the structure of the mind. (Fodor 1983 p. 112)

pace and sequencing, consciousness has limited access to them and they have shallow outputs.
If this coextension does hold, then that would be a deep fact about the mind. VFT is committed to the coextension not holding. The VFT claims that there is no actual interesting taxonomy by subject matter: all cognitive processing can be accounted for by domain specific mechanisms. VFT can remain fairly neutral about how interesting the taxonomy by computational character turns out; maybe there are encapsulated input systems. Now if the VFT is right in all of this, then the functional taxonomy cannot be made. Why? Because in order to be made, that taxonomy conceptually depends upon the other two taxonomies holding, and, as we have just seen, VFT denies that taxonomy by subject matter holds.

Why is there a conceptual dependence? It is here that we can discharge the results from chapter 6 regarding the abstractness properties of cognitive states (cognitive states in our technical sense). Remember that cognitive states, of which beliefs are a species, are stimulus, S-R, and domain abstract, etc.. What that means is that cognitive states can take just about anything as their content. So, the mechanisms responsible for the generation of cognitive states must be domain inspecific, or rather take domains of operation that are very coarse grained. This means that one can only get the functional taxonomy off the ground once one has a taxonomy by subject matter flying; and VFT denies that the latter taxonomy has wings. If VFT is correct, then cognitive state realism is in trouble, therefore, so is intentional realism.

A lot of work is being done here by the notion of the degree of grainedness of domains of operation and the abstraction properties. Whether that work is interesting from the point of view of cognitive theory is dependent upon how plausible VFT turns out to be. It is time to turn to the consideration as to whether VFT is to be preferred over cognitive state realism. Ultimately, the decision we, or rather future cognitive theorists, make will be an empirical one. However, I think we should begin looking at the evidence now. In that way, cognitive state realist cognitive architecture might be seen to be “not the only game in town”.

3 A Priori Arguments Against VFT

We have seen that VFT is committed to higher cognitive processes being performed by domain specific, or more precisely, fine grained cognitive mechanisms. This claim has two components: the first is an a priori claim to the effect that it’s conceptually possible to account for higher cognitive processes under VFT. The second component is an empirical claim to the effect that in all likelihood, the data can support VFT. We tackle the a priori
in this section, and the empirical in section 4 where evidence from neuropsychology will be examined.

Why might one opt for a cognitive state realist, or at least a mixed cognitive architecture, part VFT and part cognitive state realist? One might think that we are conceptually compelled towards cognitive state realism. Both Fodor and John R. Anderson think this way. We may look at Anderson first and then Fodor.

3.1 Anderson

In *The Architecture of Cognition* John R. Anderson says:

The most deeply rooted preconception guiding my theorising is a belief in the unity of human cognition, that is, that all the higher cognitive processes, such as memory, language, problem solving, imagery, deduction, and induction, are different manifestations of the same underlying system. This is not to deny that there are many powerful, special-purpose “peripheral” systems for processing perceptual information and co-ordinating motor performance. However, behind these lies a common cognitive system for higher-level processing. ... The view that the mind is unitary is certainly not universally held; it may not even be a majority opinion. ... This faculty approach holds that distinct cognitive principles underlie the operations of distinct cognitive functions. The unitary approach holds that all higher-level cognitive functions can be explained by one set of principles. (Anderson 1983 pp. 1-2)

He gives three arguments for accepting the cognitive state realist style architecture over that of VFT.

The first is an argument from evolutionary development. The VFT postulates special faculties or organs whose operations are restricted to a particular content domain such as mathematics, chess, computer programming or sculpture. But these are content domains for which there was no possibility of anticipation in our species evolutionary history. So, we shouldn’t postulate domain specific, innate mechanisms but rather a domain cross-cutting mechanism that while being operative in particular domains such as chess and computer programming do so only due to their general properties consistent with the species evolutionary history (Anderson 1983 p. 3).
The main problem with this argument is that no post-phrenological faculty theorist is going to postulate innate faculties that are as content specific as computer programming or chess. Not even a rabid nativist like Fodor claims that there is a chess module. Obviously, these cannot be anticipated in our species’ evolutionary history, and hence no innate genetically determined faculties ought be postulated for these abilities. Just because there seems to be some isolatable cognitive ability, it does not follow that there is an actual mechanism to be postulated by VFT. Even Gall thought that there were a collection of fundamental faculties which had to be discovered.

This objection also assumes that VFT is committed to the innateness of vertical modules. As argued in the previous section, VFT requires only that there be isolatable cognitive mechanisms. How we come to have them is another issue. My bet is that many of the mechanisms identified by VFT will be innately specified; but this does not follow from VFT itself. If the contention of section 2 is correct, then it might well be the case that there are chess playing or computer programming modules.

The second argument concerns the localisation of function. He says:

There is a tendency to regard the existence of “language areas” and other localisations of function in the brain as strong evidence for faculties. However, there is nothing necessary about this inference, as shown by a computer analogy: two programs can occupy different areas of computer memory, much as two different cognitive abilities might lie in two separate areas of the brain. However, the programmes may have identical principles. For instance, I can have one ACT simulation doing language and another doing geometry. Thus, there need be no connection between distinct physical location and distinct cognitive principles. The real issue concerns the uniqueness of the structure and processes underlying cognitive functions, not their physical location. (Anderson 1983 pp. 3-4)

Two points need to be raised regarding this passage. Firstly, Anderson is correct in claiming that it is not the physical location of cognitive function that is important. However, the important issue does not involve the “uniqueness of the structure and processes underlying cognitive function”. Instead the issue is about the grain of domain of operation of whatever structures and processes we want to postulate. The point VFT is committed to is not one of localisation of function, but that a function corresponding to a task domain gets operated upon by its own mechanism which performs that function rather than some other(s).
Chapter 9  Two Theories of Cognitive Architecture  194

The second point follows from the first and shows what I take to be Anderson's chief misunderstanding regarding the VFT programme. Anderson's use of 'distinct cognitive principles' hints at his mistaken belief that the "states and processes" underlying a domain specific mechanism will vary across domains and hence faculties; but this is not what the VFT requires—although it may well contingently be the case. Domain specificity does not imply distinct cognitive principles as Anderson seems to assume. All that is implied is the distinctness of the content of the information being processed by the faculties—and not the distinctness of the principles by which that information is processed.

If we take Andersen's own analogy between programmes and faculties then just as two programmes can have identical principles (built up out of list structures etc.), two faculties might possess similar principles (input systems, memory or storage capacity etc.). This is the idea behind vertical faculties each possessing general horizontal properties such as memory and judgement mentioned when discussing Gall above. Thus, if Anderson has one simulation doing language and another geometry, although the underlying states and processes are the same, they are different applications of the one system of states and processes. Different applications of the one set of principles are just two mechanisms which process different informational content but with similar structural features. In effect, then, Anderson himself has content specific mechanisms operating which are different applications of certain cognitive principles which is perfectly consistent with the VFT. In short, in this argument it is assumed that the VFT requires cognitive domains to exhibit disparate cognitive processing principles. But this is a false assumption—as was the case with the idiosyncratic processing strategies of Fodor (see chapter 5). Anderson seems to confuse the question of mechanism with the question of process when he speaks of "distinct cognitive principles". Higher functions might be performed by one big central processor in which the one processing strategy that processor uses is common to the performance of all those functions. Alternatively, there might be a collection of vertical mechanisms each operating according to their own idiosyncratic principles; this seems to be what Anderson is arguing against. However, there is a third alternative viz. that there is a collection of vertical mechanisms each operating according to more or less the same principles. This scenario is also consistent with VFT.3

3If the processing strategies of the vertical mechanisms are the same then why have them instead of the one big processor? See chapter 1.
The final argument is that faculties should not be postulated for the reason that the boundaries between these organs cannot be drawn \textit{a priori}. He says:

It is pretty clear where the activity of the lung leaves off and that of the circulation system takes over, but this really cannot be said for cognitive faculties. The lung and the heart are both involved in an activity such as running, but it is possible to identify their distinctive contributions. It has been proposed that there is a language faculty, a number faculty, a deduction faculty, and a problem solving faculty, but if there are such faculties, their activities are terribly intertwined in a task like computer programming. When we look at an expert programmer creating a programme, we cannot separate the contributions of the various faculties. Indeed, if we applied any reasonable criterion for individuating faculties, we would have conclude that computer programming was a separate faculty. This is because some of the core principles for this skill organisation, such as strategies for creating recursive programmes, apply across the entire range of programming behaviours and are seldom if ever evoked elsewhere. Since it is nonsense to suggest a programming faculty, we should be more skeptical of other proposed faculties. (Anderson 1983 p. 4)

I'm not sure why Anderson thinks that there is an \textit{a priori} difficulty here. It would not appear obvious that distinguishing the respective contributions of the lungs and circulatory system in running is in any sense \textit{a priori}. Surely, such a distinction is an \textit{empirical} matter. I grant that it is difficult to make such a distinction in the case of which faculties contribute to an activity such as computer programming. Of course, at this stage we are nowhere near to coming up with anything like a complete list of faculties. So determining what mix is required in order to perform computer programming is just not possible. This is an epistemic shortcoming, given our present state of knowledge. The same state of ignorance would probably have been the case for the lungs and circulatory system with respect to running a few hundred years ago. It is not an \textit{a priori} shortcoming.

What this objection should be claiming is that there is no way to individuate actual cognitive mechanisms in the way VFT requires. Now that \textit{would} be a problem; but it is general problem for all cognitive theories which require the individuation of cognitive mechanisms. When does the activity of input systems leave off and central processes start? We need some answer to that question (at the boundary between perception and cognition Fodor must say, but has no account to give of how to draw \textit{that} boundary), but if so, then
that is not an objection against VFT; it is instead a general methodological problem for cognitive science.

3.2 Fodor

In considering higher cognitive processes, Fodor asks two questions: Are there domain inspecific cognitive mechanisms? and if there are, are they modular (i.e. also informationally encapsulated, mandatory, fast, etc.)? His answers are that, yes, there must be domain inspecific mechanisms which are informationally unencapsulated. The arguments for the two answers are independent, so I will take them in turn.

3.2.1 The Argument for Domain Inspecificity

The argument for the necessity of domain cross cutting cognitive mechanisms is an a priori one. He says:

The general form of the argument goes back to at least Aristotle: the representations that input systems deliver have to interface somewhere, and the computational mechanisms that effect the interface must ipso facto have access to information from more than one cognitive domain. (Fodor 1983 pp. 101-2)

Three instances are given of this general argument. (a):

To a first approximation, we can assume that the mechanisms that affect this process work like this: they look simultaneously at the representations delivered by the various input systems and at the information currently in memory, and they arrive at a best (i.e., best available) hypothesis about how the world must be, given these various sorts of data. But if there are mechanisms that fix perceptual belief, and if they work in anything like this way, then these mechanisms are not domain specific. Indeed, the point of having them is precisely to ensure that, wherever possible, what the organisms believes is determined by all the information it has access to, regardless of which cognitive domains this information is drawn from. (Fodor 1983 p. 102)

(b) The mechanisms responsible for speech production must have access to what we see, hear, remember or think. Hence, such mechanisms effect an
interface amongst the vertical faculties and are not domain specific. (c) One of the points of perception is that the world often turns out to be other than the organism expects given background information in the possession of the organism; but an interface between this background information and perception must take place somewhere if that information is to be used in the fixation of belief and production of consequent behaviour. Again, it seems that there must be mechanisms which cut across the domains of the input systems (Fodor 1983 pp. 102-3).

The above fail to support the necessity of domain inspecific mechanisms. Remember that Fodor himself says:

> domain specificity has to do with the range of questions for which a device provides answers (the range of inputs for which it computes analyses); whereas encapsulation has to do with the range of information the device consults in deciding what answers to provide. A system could thus be domain specific but unencapsulated (it answers a relatively narrow range of questions but in doing so it uses whatever it knows); and a system could be nondenominational but encapsulated (it will give some answer to any question; but it gives answers off the top of its head—i.e., by reference to less than all the relevant information).

> If, in short, it is true that only domain-specific systems are encapsulated, then that truth is interesting. (Fodor 1983 pp. 103-104)

The argument deriving from Aristotle makes no mention of the questions that mechanisms processing the outputs from input systems must answer. Each instance of the argument concerns the amount of information used in the processing of the outputs from those perceptual systems. So, these arguments cannot yield the conclusion that higher mechanisms are domain inspecific. At most, they support the conclusion that these mechanisms are unencapsulated. It is still possible given these examples that the questions posed are quite specific, but are answered with respect to all the information available to the organism. The only way this argument would work was if there were a premise to the effect that all encapsulated mechanisms are domain specific; but there is no such premise. While VFT can agree that the mechanisms responsible for higher cognitive processes need to be unencapsulated, VFT can be compromised only if it is shown that there need be coarse-grained mechanisms at work.
3.2.2 The Argument for Informational Unencapsulation

The structure of this argument takes the form of an analogy between scientific confirmation from the philosophy of science and the nature of belief fixation in cognitive science:

central systems look at what the input systems deliver, and they look at what is in memory, and they use this information to constrain the computation of “best hypotheses” about what the world is like. These processes are, of course, largely unconscious, and very little is known about their operation. However, it seems reasonable enough that something can be inferred about them from what we know about explicit processes of nondemonstrative inference—viz., from what we know about empirical inference in science. (Fodor 1983 p. 104)

Fodor goes on, firstly, to suggest that nondemonstrative fixation of belief in science is unencapsulated; it is unencapsulated because it exhibits two properties: scientific confirmation is both isotropic and Quinean. He then goes on to suggest that central processes are also isotropic and Quinean, and hence, unencapsulated. He has independent reasons why central processes should exhibit these properties: assuming central processes to be unencapsulated, then a major problem confronting AI modelling of higher cognitive processing, viz. the “frame problem,” is exactly what we would expect under the assumption that those higher processes involve isotropic and Quinean computations (Fodor 1983 pp. 105-15). There are various conceptions of the frame problem. It can probably be best summed up as the problem of how to update an AI system’s database in the course of the system’s operation, such that only the data structures relevant to those operations are affected. A robot, for example, in moving from one room to another should update data relevant to only its position, and not data regarding, say, the value of pi, the name of its creator or the colour of the walls in the room.4 More will be said about the frame problem in 3.2.4.

Notice that these arguments attempt to show that central systems are informationally unencapsulated only. They are not designed to show the domain specificity of those systems. That, at this stage, is assumed to have been argued for under the Aristotelian argument considered in the previous section.

---

4For more on the frame problem see Dennett (1984a) and McCarthy & Hayes (1969).
Chapter 9

Two Theories of Cognitive Architecture

Having thus outlined Fodor's argument, I do not intend to examine it in any detail. As might be already obvious, I agree that higher processes are plausibly considered to be unencapsulated. It is their domain specificity that is at issue when considering VFT. In particular, I want to show how domain specificity can be made consistent with the properties of being isotropic and Quinean. In order to do this, in the remainder of this section I will consider these two properties in some detail.

3.2.3 On Being Isotropic and Quinean

If the fixation of belief is isotropic, that means that the information relevant to the processes of fixation may be drawn from anywhere in the organism's cognitive economy. By “anywhere” is meant information from any content domain. In the case of scientific confirmation, then, “everything the scientist knows is, in principle, relevant to determining what else he ought to believe. In principle, our botany constrains our astronomy, if only we could think of ways to make them connect” (Fodor 1983 p. 105). It's fairly obvious that isotropic systems are also unencapsulated systems par excellence. By referring to any and all the information available to the organism, there is no question of the belief fixation occurring as a result of access to limited information.

Because the information available to an isotropic system is domain insensitive, it should not be thought that isotropic systems are domain insensitive as well; isotropic systems may well utilise information from a variety of content domains, but only answer a very limited range of questions in which case it must be domain specific.

Closely related to a system's being isotropic is a system's being Quinean. In scientific confirmation it might be the case that the isotropic system puts up two theories that generate predictions about every relevant thing covered by the theories, but that one of the theories is better confirmed than the other on grounds of simplicity, conservatism or whatever. The point about being Quinean is that these grounds for confirmation “are properties that theories have in virtue of their relation to the whole structure of scientific beliefs taken collectively. A measure of conservatism or simplicity would be a metric over global properties of belief systems” (Fodor 1983 p. 108). Again, the theories generated by a Quinean cognitive mechanism may be in answer to a very limited range of questions even though which theory is

---

5I refer the reader to Philip Cam (1987) if s/he is interested in Fodor's argument.
accepted is determined by such global properties. There is no reason why
those questions need be in any sense global.

One might want to object against VFT that it's just obvious that our
higher cognitive processes answer a wide range of questions, and are
therefore domain inspecific. I find this not at all obvious. What I do find
obvious, though, is that our cognitive system as a whole answers a wide range
of questions. That, however, tells us nothing about the mechanisms at Level
Two which go to make up that cognitive system; to make that inference would
be to commit a Level mistake—see chapter 1. In order for VFT to be correct, it
will have to account for this apparent obviousness of our being coarse-grained
representers. What the supporter of VFT will have to do, therefore, is make a
move analogous to that which we saw in chapter 4. Remember that in the
case of the at worst connectionism scenario, we had to retreat to Level One in
order to count the system as a representer. In the present case, the adherent
of VFT will have to retreat to Level One also, so that systems such as
ourselves can count as coarse-grained representers.

One reason why the issue of informational encapsulation has been run
together with domain specificity, is that the two might not be as independent
as Fodor assumes. As we have seen, content would seem to be an inform-
atival notion. That is, after all, the presumption behind works such as
Dretske's Knowledge and the Flow of Information (Dretske 1981). So the
content domain specificity of a cognitive mechanism might well be
conceptually linked to questions of encapsulation.

3.2.4 The Frame Problem and CSR

As we have seen above, Fodor's argument that central processes are Quineian
and isotropic rests to a large extent upon an analogy with scientific
confirmation. We also saw that he has independent motivation for thinking
that central processes have those properties: if they are isotropic and
Quineian then the frame problem is just the sort of problem we should expect
to arise. Now the frame problem (and, indeed, the problem of scientific
confirmation) is considered by Fodor to be a fairly intractable problem. In
fact, Fodor casts doubt upon many of the approaches by AI theorists to get
around the former problem (1983 pp.114-17). Both "heuristic" approaches and
the employment of scripts or frames fail to get around the problem, at least,
according to Fodor.

From this supposed intractability, Fodor then casts a net of gloom over
the prospects for a cognitive science of central processes. He claims that all
the progress made in cognitive enquiry thus far has been about the modular
input systems, and its the holistic and global nature of belief fixation which has led to a relatively dead end in the case of central processes. Consequently, Fodor characterises what he hopes will one day be called “Fodor’s First Law of the Nonexistence of Cognitive Science”. This “law” states that the more global and isotropic a cognitive process, the less we can understand it, and very isotropic processes can’t be understood at all (1983 p.107).

This gloomy conclusion might strike some readers as surprising, given Fodor’s stance on intentional and cognitive state realism. Since, central processes are where belief fixation occurs, there will be a cognitive science utilising intentional/cognitive states, only if there is a cognitive science of central processes. So, if Fodor’s pessimistic epistemological story is correct, then so much the worse for intentional realism.

Actually, it seems that either Fodor’s pessimistic scenario does not follow, or intentional realism is going to be in trouble. Consider the status of the frame problem with respect to us. The frame problem seems to be a problem only from the point of view of AI, and not cognitive psychology proper. Why? Because, if Fodor is right, although we possesses central processors which are inherently Quineian and isotropic, we do not suffer from the frame problem; we don’t suffer the sorts of problems confronting robots of AI research. Two alternatives follow from this, it seems to me. First, since nature, seemingly, has solved the frame problem in our case, then there should be, in principle, no reason why we should also not be able to solve it in the case of AI; the frame problem would not, then, be as intractable as Fodor suggests. In this case, Fodor’s pessimistic epistemological conclusion does not follow. Or secondly, if it is true, as Fodor claims, that central systems are Quinean and isotropic, and that these properties do invariably lead to intractable frame problems, and it is also true that we do not exhibit the frame or similar problems, then that must constitute evidence that we do not possess central processes. If we do not possess central processes then CSR fails. Either way, it seems to me that Fodor’s argument employing the frame problem is going to be problematic for him.

4 Empirical Arguments

In the previous section we took a brief look at some arguments which purported to show why an alternative architecture to that of cognitive state realism will not be feasible. Since cognitive state realism is an empirical thesis we now look at some empirical data to look for support for either it, or the alternative architecture I have been advocating viz. vertical faculty theory. I stress again that this section should be considered more as a
preliminary step to the disconfirmation or confirmation of cognitive state realism rather than a final word. The present work should be considered as the theoretical foundations for a substantive empirical research project. With this in mind we turn to the data.

4.1 The Neuropsychological Argument

What does the evidence presented in Appendix B tell us about the modularity of cognitive mechanisms? Most of this evidence could be used to support the existence of Fodorian modular input systems. We, however, are interested in the architectural principles which generate the higher cognitive processes underlying behaviour.

Remember that the cognitive neuropsychologist makes the following sort of inference: the impairment of some function F in isolation from other functions is evidence for that function's being performed by a module (in the brain state sense from chapter 6), and hence being a psychologically confirmed function. Now, what about the converse inference, from lack of impairment to lack of modularity? One might think that in the absence of such impairment we should seriously entertain the possibility that there is no such module. The evidence from neuropsychology would tell us that there are no coarse grained impairments; all the impairments cited are fine grained. This suggests the possibility of there not being coarse grained modules that the cognitive state realism requires to account for higher cognitive processes.

Martin Davies and Max Coltheart (Unpublished) cast doubt upon this inference since there are two possible explanations of the lack of dissociation: one is that modularity is lacking, while the other is that certain nonpsychologically important factors are coming into play. Such factors might be that the brain states in question which realise the psychological functions are not modular in the brain state sense of modularity introduced above in chapter 6.

The D & C line is correct as far as it goes; lack of impairment might just mean that central processes, say, are not modular in that brain state sense. However, we are interested in the other sense of modularity in which a task is modular if it features as part of the Level Two functional analysis of the system. So, there is another stage to the neuropsychologically based inference: lack of impairment implies lack of modularity which, further, implies lack of psychologically real function. Davies and Coltheart take as their example two hypothetical tasks, A and B, and ask if they are modular or not. Now if either breaks down or continues in the absence of the other task, then we would say that they were modular tasks. However, how many actual
tasks do we want to claim are performed? If it's the case that the nonpsychologically interesting factors ensure that these tasks never get performed in isolation, then why do we want to claim that there are separate modules within our Level Two analysis of the system? It would seem that we ought to say that there is some general task being performed here, a task one might describe as the conjunction of those two subtasks. We ought do this unless we have some other reasons, independent of the neuropsychological evidence, to suggest that there are two tasks here instead of one, even though those tasks are not modular in the brain state sense. In other words, the neuropsychological evidence does not support, in this hypothetical case, the postulation of two tasks being performed, and we must look elsewhere for evidence to support the existence of our two psychologically real tasks.

Where might such alternative evidence come from? Well, from anywhere in the psychological disciplines, really. We have already seen in chapter 1 Pylyshyn's criteria for psychological reality. It will be data from areas such psychometrical studies which will provide evidence for the Level Two modularity of some psychological function. In the absence of such alternative evidence, then the lack of modularity in the brain state sense should lead one to infer lack of modularity in the functional description sense.

Memory is probably the choice example of a higher cognitive function which, seemingly, fails to count as a module in our sense, just because it would appear to be non-modular in the brain state sense. Says Gross:

No one has ever selectively destroyed memory and only memory. Even lesions in the hippocampal region, which leave a devastating effect on many different types of memories, leave memories of motor skills intact. (Corkin 1968)

To see the importance of this type of claim we can look at Stich's now famous example of Mrs T., an elderly woman who suffered a progressive loss of memory due to the degeneration of brain tissue (Stich 1983 pp. 54-6). Mrs T. presumably lost the concept of ASSASSINATION since she would continue to utter "McKinley was assassinated" but would not realise, inter alia, that this would mean that McKinley must consequently be dead. Now the usual conclusion drawn from this example is that it is unclear that Mrs T. believes that McKinley was assassinated. If we grant that she has lost the concept ASSASSINATION just what cognitive differences does that make to Mrs T., and are those differences the result of some coarse grained disruption to her cognitive system?

Now although Mrs T. has lost the concept ASSASSINATION something within her cognitive system is playing various roles in the running of
cognitive processes. For instance, she can still utter the sentence 'McKinley was assassinated'. Moreover, she utters the sentence in response to questions about McKinley. So, because she lacks the concept but can still exhibit behaviour taken to be partly indicative of belief, then it follows that coarse grained cognitive states are not necessary for the production of those behaviours.

But what does that tell us about vertical modules? There would seem to be two options. The first is to claim that there are no central processes and that although Mrs T. has somehow lost the concept she still possesses some fine grained modules that have stored information about McKinley—thus explaining the continued ability to answer questions about him. On this option it is possible to explain the various deficits and continued competences of Mrs T. Strictly speaking there are no concepts featuring at Level Two according to VFT. What Mrs T. has lost is perhaps the meaning of 'assassination' (from her language module's lexicon of course!).

In claiming that according to VFT there really are no concepts at Level Two, I am not saying that we do not possess a conceptual repertoire. The claim is, rather, that there are no Level Two states that play the causal roles that cognitive state realism requires. Perhaps concepts are attributed to a system at Level One. Alternatively, perhaps concepts are just lexical items featuring in a language module. If either of these options turns out true so much the worse for cognitive state realism.

Even in the face of this ignorance of the mechanisms at work according to the VFT programme, things get worse if one opts for the following second option. Assume that there is a central processor at work which accesses concepts. If Mrs T. has lost that concept then it is pretty difficult to see how she can believe that McKinley was assassinated. Okay; but if she didn't believe that McKinley was assassinated it's also difficult to see why she would respond to questions about McKinley in the manner she does given the loss of the concept. It would be no good for the believer in central processes to claim that Mrs T.'s continued capacities resulted from, say, the continuance of some sentence token in a belief register; since it is causal role that is constitutive of something's being a belief, and it is exactly the causal role that has broken down in Mrs T.'s case, by hypothesis, then it does not make sense to postulate a mental sentence. So, on this option we are going to be at a loss to explain Mrs T.'s behaviour.

The Mrs T. example was used by Stich to bring out the vagaries associated with the attribution of content to mental states. I think the

---

6Well, I really don't have the faintest idea what Mrs T. has lost, since the range of cognitive mechanisms postulated by VFT is unknown.
example shows us something about causal roles. What Mrs T. shows us, like the rest of the neuropsychological data, is that breakdowns fractionate the mind at a fine level of grain. There is obviously some state of Mrs T. that plays some causal role in her answering of the interlocutor's questions. Given that this causal role is a partial constituent of beliefhood, the state in question would have to be a partial belief of some sort. I'm not sure of what such a partial belief could be—a mental sentence of the form "McKinley was #$%*!". That would still not explain, though, why Mrs T. gave the answer that she did. We can have degrees of belief that form the basis of our actions according to one's favourite decision theory, but that cannot be the sense of partial belief required here. In order to postulate coarse grained Level Two cognitive mechanisms a cognitive state theorist will have to account for such cognitive anomalies as well as the normalities—this being the claim the Churchlands have been making for years—by given some account of the etiology of Mrs T.'s response. I don't see where such an account is going to come from on the cognitive state realist story.

What the neuropsychological evidence and Mrs T. tell us is that really there are a number of smaller tasks that are performed, and most importantly that due to dissociations and breakdowns, the modularity of many of those tasks can be inferred. Since those modules are fine grained mechanisms, we get support for the VFT over cognitive state realism. Now this seems to be the state of play as seen from the neuropsychological literature only. The argument as just presented depends crucially on there not being other independent reason for postulated coarse grained mechanisms. Granting that my knowledge of the empirical data leaves quite a deal to be desired, I haven't seen any data that should lead us to confirm the existence of coarse grained mechanisms, which is necessary in order to confirm cognitive states realism. That might be surprising to some given some well known experimental data, which we should now examine. I don't think, however, that the data shows what it is purported to show.

4.2 Propositional Representation

It is without doubt that from the perspective of Level One, our cognitive processes seem to range over coarse grained representations in the way that cognitive state realism (and intentional realism) requires. This is most easily seen in our use of language: in our production and understanding of linguistic items we manipulate coarse grained information. One way to interpret this capacity which is consistent with cognitive state realism is to claim that rather than these coarse grained representations featuring within the
language module, say, there is another representational system which not only translates into a natural language but which also underlies cognitively mediated behaviour, with linguistic utterances being examples of such behaviour. Such propositional representations as they have been called (Stillings et al. 1987 p. 21) can be construed as cognitive states. The question is: what kind of evidence is adduced in support of the existence of such states?

4.2.1 Introspective Arguments

One style of argument is based upon the introspected distinction between having an idea and putting that idea into words. We often, so the story goes, have a concept “clearly in mind” but cannot retrieve the word that stands for the concept. This is the so-called “tip-of-the-tongue” phenomenon, and is supposed to indicate that concepts are not identical with linguistic items. Similarly, there is also the problem of giving the definition of a some natural language word which stands for a concept. Very often we find that the giving of the definition of a word takes a considerable amount of processing and mental effort. This is supposed to suggest that the definitions of those words are expressed in a separate internal representational medium which gets associated with the word (Stillings et al. p. 22).

It seems, though, that the former example’s working depends upon how we might possibly “have a concept clearly in mind” without being able to retrieve the word for the concept. Maybe one cannot have a concept in mind in the way required without being able to come up with the word. Such euphemistic expressions, which often feature in such anecdotal evidence, don’t do much to clear the already muddied cognitive waters. Of course, a proper cognitive theory is going to have to account for such introspected experiences. If one assumes that words cannot be concepts, then this problem can be explained by that fact. However, if words are concepts, then one can equally explain the problem by reference to occasional difficulties in accessing lexical items by the language module, say. Why there are such occasional difficulties we would hope to eventually explain.

Far from showing that the definitions associated with linguistic items are expressed in a separate propositional system, the second example can be easily explained by questioning whether there is a strict definition associated with words. If there are not, then it will take some cognitive processing to come up with the criteria of application for the word. Perhaps we are not very good at giving definitions because we have only a rough prototype to draw from and when asked to give an actual definition we try to flesh out the prototype as exhaustively as we can. The point is that the processing that
Another reason offered for postulating a propositional representational system separate from the language system is that the existence of a propositional network putatively explains how children can begin acquiring concepts even though they were not "born knowing a natural language" (Stillings et al. p. 22). As stated, this reason will fail because it might be that being born with a natural language is irrelevant, just so long as one acquires concepts when acquiring that language. That would count against there having to be separate coarse grained representational network. The response to this is to claim that children can learn concepts that they are unable to explain until they have been talking for quite some time. Again, this does not show that concepts are formed independently of language, but just that in the acquisition of concepts we do not acquire strict definitions of those concepts so that we can explain what the concepts are.

4.2.2 A Recognition Memory Test: The Retention of Meaning

All of us at some stage have had the experience of remembering the overall point of something said to us, while not retaining the actual words used to communicate that point. It is argued that this phenomenon supports the view that there is an underlying propositional representational system independent of language. There is, supposedly, experimental evidence in support of this anecdotal evidence. Here is a description of a classic experiment by Sachs (1967).

In the experiment subjects listened to paragraphs on various topics, such as Galileo's work on the telescope. The passages were interrupted, and subjects were given a recognition memory test for a sentence that had occurred 0, 40, or 80 syllables earlier in the passage. In the test the subject was presented with a sentence and asked to judge whether it was the same as a sentence in the passage or changed. For the Galileo passage the subject could be presented with any one of the following four sentences. The actual sentence was sentence 1.

1. He sent a letter about it to Galileo, the great Italian scientist.
2. He sent Galileo, the great Italian scientist, a letter about it.
3. A letter about it was sent to Galileo, the great Italian scientist.
4. Galileo, the great Italian scientist, sent him a letter about it.

Notice that sentences 2 and 3 have the same meaning as sentence 1, although they have different grammatical forms. Sentence 4 has a different meaning. At a delay of 80 syllables (20-25 seconds) subjects rarely made the error of thinking that sentence 4 has occurred in the passage. However, they frequently and mistakenly thought they remembered sentence 2 or sentence 3. Thus, under the conditions of Sach's experiment subjects were able to remember the meaning of a sentence without remembering its linguistic form. A reasonable explanation of the finding is that the meaning was represented in an underlying nonlinguistic form that could be translated into a variety of linguistic paraphrases, which subjects confused in the recognition test (Stillings 1987 pp. 22-3)

This case certainly shows that certain components in the analysis of linguistic items, syntactic and lexical details, can be lost very quickly. However, what I fail to see is that it shows us anything about there being some nonlinguistic propositional representational system. I think it's fairly uncontroversial that in this case there has been an abstraction away from certain linguistic properties of an utterance; that should not be surprising given the distinguishing marks of cognitive systems put forward in this work. However, there does not seem to be any requirement that the abstract entities which somehow carry the meaning of the utterances be nonlinguistic. Fodor has called these meaning bearing entities "interlevel representations" (Fodor 1983 pp. 55-60), and they are representations which may be internal to the language module. In that way, they will not be nonlinguistic. Whatever the mechanisms are which mediate the processes of verbal comprehension they will have to deliver the meaning of linguistic utterances, where that meaning abstracts away from the details of syntax in just this way. Perhaps the claim being made by the proponents of propositional representation is that just as soon as the particular details of syntax are lost then the system has gone propositional. If that's so, then there cannot be an argument here for cognitive state realism. That's because the representations which are now described as propositional feature within the language module, and that means that they will not be coarse grained enough to qualify as cognitive states in the sense of the act.

---

7There must be a huge caveat hanging over this example to the effect that very often such details are not lost in comprehension. The passing of dictation tests at school and the existence of stenographers points to this fact.
4.2.3 Scripts

The trouble with the arguments presented so far is that, even if they do show that there is a propositional representational network, and so there are coarse grained representations, that still does not provide evidence for cognitive state realism's being true, since those representations might not possess the causal roles that partly constitutive of cognitive states. We now turn to arguments from cognitive psychology which purport to show the existence of coarse grained representations which play the kinds of causal roles outlined in chapter 5. We turn then to questions of scripts.

A script is hypothesised to be "a declarative knowledge structure that captures general information about a routine series of events or a recurrent type of social event, such eating in a restaurant or visiting the doctor" (Stillings et al. p. 31). They are thought to be constituted by such things as names or themes (visiting the doctor), typical roles (doctor, nurses and patient) entry conditions (feeling ill) and goal directed scenes (checking in the receptionist, waiting, undressing).

The idea behind scripts seems to be this. We act out of our knowledge about the world, given certain desires we possess. Now scripts are supposed to provide the representational states which interact with those desires to bring about action. So, it might be scripts which will contain some of the cognitive states which will play the appropriate causal roles. Two questions are raised by the literature on scripts: are there scripts? and if so, are they structures composed of cognitive states?

As to the first question, the evidence for the existence of scripts is especially weak. The main argument for their existence runs something like this: a knowledge structure such as scripts are postulated. One then hypothesises certain behavioural consequences of their presence that should show up under experimental conditions. It then turns out that the behavioural consequences occur, and that constitutes evidence for the scripts. The weakness in the argument is that very often that behavioural consequence might be explicable by some other feature of cognitive processing.

A classic experiment used to support the existence of scripts was performed by Bower, Black and Turner (1979). They hypothesised that subjects would confuse material they had read in the experiment with material they had inadvertently filled in from their script knowledge. Their prediction was "borne out". The trouble with such experiments is that the evidence vastly underdetermines the conclusions drawn. Perhaps in the face of pressure to recall accurately some story which they do not totally
remember may well incorrectly substitute information from one's personal memory store either consciously or unconsciously. However, such a phenomenon is consistent with there being no scripts at all.

The second question regarding whether the existence of scripts constitutes vindication of cognitive state realism is most important. One problem with the script experimental paradigm is that the evidence regarding the roles that they are supposed to play in the causation of behaviour is grossly unclear. The script literature properly considered really has to do with complex concepts such as eating at a restaurant or visiting a doctor. The point in common is that these complex concepts happen to be activities which agents perform. There are many such concepts relevant to human action: marching, swimming and riding a bicycle. One may then think of scripts as the specification of a conceptual prototype. The trouble with this interpretation is that it gets us back to the problem of linguistic versus nonlinguistic knowledge representation. Scripts so construed might still be a form of linguistic representation, which would mean that it will fail to fill the types of causal roles required by cognitive state realism. To an extent, this problem is foreshadowed by Bower, Black and Turner when they claim that there is split between enacting and verbalising scripts in the collection of the script data (1979 p. 214).

Bower, Black and Turner also worry that the data collected about scripts through report and recall methods will give access to script knowledge that is consciously introspectable (p. 214). I'm not sure just what kind of knowledge unconscious nonintrospectable declarative knowledge would be like. Whatever it is like, it must be shown to get into the causing of behaviour in the way cognitive state realism requires. There is bound to be a great deal of unconscious and nonintrospectable knowledge about the concepts associated with human activities. Consider riding a bicycle. We list the requisite actions one must perform in order to successfully ride a bicycle. What one cannot do is to relay the information or declarative knowledge that would actually gain a listener the relevant skills. There's a distinction philosophers have long made between knowing how and knowing that. The information encoded in scripts would certainly seem to be knowledge that. Once one begins to go unconscious and nonintrospectable one would seem to be moving to knowledge how. Certainly knowledge how gets the kinds of causal roles required by cognitive state realism into our picture. However, we now have the wrong kind of representational information filling the causal role: it's skills encoded informationally and not the matters of fact about states of affairs in the world which cognitive state realism wants. The moral seems to be that you can have your scripts with all the propositional
information regarding human activities you like, and yet still not have that information playing a role in the causation of behaviour; some aspects associated with visiting a doctor might well just be like riding a bike.

At this stage there appear to be no arguments in support of cognitive state realism from script theory within cognitive psychology. At the moment, we are offered mere metaphor in place of theory.

Final Signpost

It is indeed an understatement to describe this chapter as a tentative start to determining what would have to be empirically the case in order for cognitive state realism to be judged true or false. To that extent the major component of this work is considered to be the theoretical foundations of the earlier chapters which allow one to pursue this empirical line of enquiry as a means to evaluating the thesis of intentional realism. There is a vast literature in the fields of neuroscience, neuropsychology, cognitive psychology and social psychology which may be surveyed in order to determine the correctness of the cognitive state realist thesis. The completion of that task awaits us in the future. Then, perhaps, the true topography of intentionality will be evident to all.

One trouble with playing the “only game in town” is that it can lead one to ignore underground swells of illegal games. One does not need to be a political or legal theorist to realise that what is considered legal often changes over time. I think cognitive legislators are going to pass some bills to legalise some other game in this town.
Let us assume that intentional realism is false, that intentional states won't feature in The-One-True-Cognitive-Psychology, our best science of the mind. One advantage if it turned out that intentional realism were true would be that it would automatically bestow upon folk psychology an ontological and explanatory role for states such as beliefs and desires. Indeed, if one believed in the singularity of explanation thesis (from chapter 1) as do the intentional realists and eliminative materialists, then intentional realism's being true would be the only way that intentional psychology could have such explanatory or ontological role to play: intentional realism's being false would force us to dispense with folk or intentional psychology as an explanatory exercise. Thankfully, the singularity of explanation thesis is wrong—or so it has been argued above.

With our assumption still in place, what kind of role might intentional psychology play in order that we do not dispense with it in face of The-One-True-Cognitive-Psychology's not requiring intentional states? I think there are at least two possibilities regarding roles other than those presented by The-One-True-Cognitive-Psychology. The first has to do with the possible non-psychological roles intentional state attributions might play. As we shall see below, this strategy actually divides in two. As part of its non-psychological usefulness, intentional psychology might be utilised by disciplines other than psychology viz. the social sciences. That utilisation might well be enough for us not to dispense with folk psychology. The second has to do with taking a broad conception of psychology so that even though intentional states might not coincide with states postulated by The-One-True-Cognitive-Psychology, there nevertheless remains a psychological role for folk states to play. We begin with some considerations regarding the non-psychological uses of folk psychology.
1 Is Folk Psychology Too Fine Grained in its Explanations?

It is often suggested that one reason why intentional realism is false, i.e. that common sense belief-desire explanations will not feature in a future science of the mind—The-One-True-Cognitive-Psychology—is that the states and processes over which The-One-True-Cognitive-Psychology quantifies will be too rich and fine grained to count as folk intentional states and processes. My present thought is that perhaps this charge has things the wrong way around. Perhaps it's intentional psychology which will turn out to be too fine grained and rich in the kinds it postulates in order to be quantified over by The-One-True-Cognitive-Psychology. Let me explain.

Intentional psychology is normally taken to be modelled on sets of two attitude types: belief and desire. Since Hume (1748) there has been a prevailing requirement for at least these two attitude types; beliefs to represent the world and desires to motivate the agent. Fodor, for one, is very explicit about central role of these two attitudes, as evidenced by his calling intentional psychology “belief-desire psychology”. But as we know, there are multifarious propositional attitude types: there are beliefs, desires, hopes, rememberings, thoughts, fears, etc. Clearly, when The-One-True-Cognitive-Psychology is complete there is going to be a determinate number of attitude-types quantified over if intentional realism is true. As to how many of the attitudes types are quantified over is, I take it, an empirical matter. However, some of the attitude types the users of intentional psychology attribute to agents seem to do more work than psychological work. Let's take a look at this in some detail. While I think that the considerations below show that the failure of the intentional realist programme will not mean the demise of intentional psychology, they can, in addition, be used as arguments for rejecting intentional realism—provided the criteria for vindication of the attitudes spelled out in chapter 1 (section 4.4) are in place.

1.1 What is a Psychological State?

Assuming that we often do frame explanations of an agent's behaviour in terms of more than just beliefs and desires, why should we be required to do so from a psychological point of view? Consider the following explanations of Reagan’s ordering the destruction of the Iranian gun boats:

---

1For example, Schiffer (1987 p. 41) and Loar (1981 pp. 17-18) use this putative feature of intentional psychology in their discussions.
EX1 (from chapter 1): Reagan believed that the Iranians were pirating Hollywood westerns and that the boats were smuggling copies across the Gulf. Reagan desired that the pirating of Hollywood westerns stop. He also believed that the ordering of their destruction would be an effective means towards ending pirating.

EX2: Reagan knew that the Iranians were pirating Hollywood westerns and that the boats were smuggling copies across the Gulf. Reagan desired that the pirating of Hollywood westerns stop. He also knew that the ordering of their destruction would be an effective means towards ending pirating.

Such explanations are commonplace. Our current question is: what does EX2 give us over EX1 in terms of the explanation of Reagan's action? I think EX2 does give us something more in the way of an explanation, but not anything more of an explanation of why Reagan did that particular action.

EX2 tells us that the Iranians really were pirating Hollywood westerns and smuggling the copies across the Gulf. But what has this have to do with Reagan's deciding to perform his action? Let's suppose that Reagan claimed to know about the Iranian's gambit. Even if it is true that Reagan did know at that time, he did not know that his claim was justified. Claims to know something when that something turns out to be false get downgraded to the status of beliefs. So, what over and above mere believing do we require in this explanation? Nothing I claim, since irrespective of whether Reagan merely believed or knew, he would have performed that action. Describing Reagan's psychological state as that of knowing that p seems somewhat redundant in intentional psychological explanation.

There is an obvious way in which EX2 is not redundant. It is that EX2 establishes why Reagan's action was efficacious in bringing about the satisfaction of Reagan's desire. If the Iranians were not running the pirated movies across the Gulf, then destroying the gun boats would fail to stop the distribution of the westerns. The attribution of knowledge to Reagan explains the success of his actions. Whether or not Reagan knew that the bombing would be an effective means of ending the pirating will determine the likelihood of Reagan's performing that action. Perhaps there are two candidate ways of stopping the pirating: bombing the gun boats or writing a letter of protest to the Ayatollah Khomenei. If he claimed to know that the former would work but only believed the latter would then we get a better idea of why he performed that action rather some other.
This, however, commits us to nothing more than differing degrees of belief on the part of an agent. An agent might claim that he knows that p. But that is only a claim. The question of the agent's actually knowing that p remains open.

Typically, explanations such as EX1 and EX2 are third person and post hoc. The reason why there is a redundancy in these cases is that while we possess information relevant to why an agent performed some action, that information does not seem relevant from the agent's point of view in whatever psychological processes are going on within him or her at the time s/he decides to perform or performs the action. From the agent's point of view, there can only be a claim to know that p before the event.

This question of the redundancy of certain propositional attitude ascriptions raises the questions of just what is a psychological state, and which of the attitudes correspond to psychological states. From the above discussion it would not seem that propositional attitudes ascriptions involving factive verbs such as 'know', 'remember', 'perceive' and 'regret' are going to count as psychological states on the intentional realist story. According to intentional realism, to have a psychological state of the type postulated by intentional psychology is to bear some relation to an internal mental representation. Since an agent's knowing or remembering that p requires that some state of affairs to be the case, then the requirement of intentional realism that an organism knows that p iff it stands in some relation to a mental representation cannot be met.

Fodor recognises that there are these problems for intentional realism to sort out. As early as The Language of Thought (1975 p.75fn) he suggests that some propositional attitude ascriptions might be problematic from the point of view of psychological theory. In a footnote he says:

Clearly, no organism knows that a is F unless it is the case that a is F. ... It follows that there can be no computational relation to a formula such that (an organism knows that a is F) iff (it stands in a relation to that formula). ... one could not expect more than a rough correspondence between the inventory of propositional attitudes that we pre-theoretically acknowledge and the ones which psychological theories prove eventually to be about. (1975 p.76fn)

Assuming that only those states quantified over by psychological theory will count as psychological states, then attitudes involving factive verbs would seem not to be purely psychological states—according to intentional realism.

\[^{2}\text{For a discussion of factive verbs see Kiparsky & Kiparsky (1971).}\]
at least. We saw in chapter 1 that intentional realism might well have to employ a mixed taxonomy of intentional states in order to account for the various attributions of propositional attitudes which do not consist in a relation to a mental representation. Propositional attitude ascriptions involving factive verbs will, presumably, have to feature as another variant in this mixed taxonomy.

The upshot of these considerations seems to be that intentional state ascriptions, while undoubtedly having some psychological salience, also have epistemic significance. What EX2 presents to us is not merely Reagan qua agent but Reagan qua knower. Presenting Reagan as a knower adds a further justificatory component to Reagan's acting on his desires: he knew that the Iranians were infringing Copyright, and so, granting his desires, the action he undertook was most justified. If he had been wrong then we would demur from our judgment that he was justified, other things being equal.

It seems that in the case of propositional attitude psychology, not only does the philosophy of psychology meet the philosophy of language, but also epistemology. Presuming that epistemology is an important theoretical enterprise, then these considerations constitute a start at determining the extra-psychological work performed by intentional psychology. The next section builds upon this start.

1.2 Explanatory Work and Justification

We can further explore the issue about justification just raised by considering yet another description of the etiology of Reagan's action framed in terms of intentional states. Consider:

EX3: Reagan feared that the Iranians were pirating Hollywood westerns and that the boats were smuggling copies across the Gulf. Reagan preferred that the pirating of Hollywood westerns stop. He also hoped that their destruction would be an effective means towards ending pirating.

How does EX3 differ from EX1 and EX2? The first difference is one's possible surprise if it turns out that Reagan's action is efficacious. After all, from the description of Reagan's epistemic condition given in EX3 it is not even clear that the Iranians are in fact pirating Hollywood westerns. While it also does not follow from EX1 that the Iranians are in fact pirating the westerns, at least if one claims to believe something, one must have reasons for doing so. Fears on the other hand, are often not accompanied by justifying reasons, at
all. It's also not clear that the destruction of the gun boats will put a stop to the pirating; the Iranians might have other channels by which to move their celluloid contraband. It seems that EX3, just like EX1 and EX2, has built into it an epistemic evaluation component. Irrespective of whether or not Reagan's action is efficacious, if we explain his action by EX3 rather EX1 or EX2 then we imply that there is a potential epistemic objection to whether Reagan was justified in taking his course of action. I want now to take a look at this idea of the epistemic justificatory component of intentional explanations.

There are a couple of ways in which intentional explanations contribute to the praise or blame of an agent's actions. The first, but not very interesting way from the point of view of our current concerns, is that an attribution of a desire or preference of the agent is made. Now if we think that the agent's desire is not worthwhile—as Reagan's would seem not be worthwhile (well, to most of us anyway)—then any consequent action based upon that desire is going to be deemed not worthwhile or even irresponsible. The second and interesting way takes as a given that the desire of the agent is worth acting in accordance with, but questions the epistemic condition of the agent and how that condition affects the satisfaction of that desire. If it is not clear or perhaps guaranteed that destroying the gun boats will satisfy Reagan's desire, then we will assess his action as irresponsible, or whatever. This epistemic evaluative role is what EX3 makes clear where the other candidate explanations do not.

To the extent that EX3 plays a part in the epistemic evaluation of an agent's actions, it has some legitimate explanatory work to perform. However the question requiring answering here is: is that explanatory work performed by EX3 work required from a psychological (read: intentional realist) perspective?

Forgetting for the moment about epistemic evaluation and its role in the assessing of an agent's action, from a psychological point of view, it would not seem that EX3 would be required as an alternative to EX1 or EX2 as an explanation of Reagan's action. In so far as we want to explain the brute behavioural output of Reagan (which remains constant across the explanations, I should point out), EX1 is going to perform all the explanatory work we require. The reason why is much the same as in the case of why EX2 might not be required as an explanation by the intentional realist. In the case of EX3 there is no factive component of the intentional states cited in the explanation. But just as the question of the truth or falsity of the belief seemed irrelevant to the intentional realist's explanatory enterprise, so too the epistemic justification would seem irrelevant, since the behaviour of Reagan would be the same across EX1 and EX3.
One might object that there is an obvious way in which EX3 does more psychological work. It will explain Reagan’s surprise if his gambit pays off, since he was unsure as to whether the course of action he undertook would be efficacious. Perhaps when making his original decision, Reagan might have sniggered to himself and Shultz because he thought it would be fun to upset the Iranians even if the information upon which he was acting was not epistemically justified. In this case, however, Reagan’s behaviour is different from the case described above where EX1 or EX2 seem relevant. Reagan’s action, as initially described, was the simple ordering of the attack, and not ordering the attack with a snigger to Shultz. The behaviour to be explained needs to remain constant if this point is to hold.

If the claim that explanations such as EX3 are not going to be required by psychology seems counterintuitive, it shouldn’t. What I am not saying is that explanations such as EX3 do not have a role to play in our describing and assessing the actions of agents. Of course EX3 is not redundant in this respect. The point I am making is that it might be the case that The-One-True-Cognitive-Psychology will not require explanations such as EX3. What I am suggesting is that perhaps intentional psychology does more work than mere psychological work. Intentional psychology might be bound up in our assessments of an agent’s action—whether that action is justified, right or wrong etc.. Given that intentional psychology is basically a folk theory, and such folk theories developed in conjunction with our epistemic and moral conceptions, then this extra non-psychological work performed by intentional psychology should come as no surprise.

A problem for the intentional realist, therefore, is to decide how much of this extra work s/he thinks The-One-True-Cognitive-Psychology is going to buy into. What decision the intentional realist ultimately makes will depend upon views regarding issues such as whether our psychological conceptions of sanity and rationality are inherently evaluative in ways other than those we might think a purely scientific enterprise should be. Whether or not the intentional realist is going to easily decide upon such issues, or whether the issues are decidable at all, is not my purpose in raising these issues. Our concern has been to point out that any extra-psychological component is going to pressure one to accept that intentional psychology may well have uses even when intentional realism fails.

2 A Broader Conception of Psychology

We have thus far been considering non-psychological uses for intentional psychology that might ensure that folk psychology will not be dispensable.
There are, however, denials of dispensibility on purely psychological grounds. I am thinking of Kitcher (1984) and Jackson and Pettit (Forthcoming). These works, though, are silent on the question of the status of intentional psychology given that intentional realism fails. Kitcher merely attempts to play down the supposed shortcomings of folk psychology as argued by the eliminativists (see chapter 1). Jackson and Pettit attempt to give an account of intentional folk psychology which would make it consistent with whatever ways the One-True-Cognitive-Psychology or the neurosciences turn out—whichever of these two comes up being the proper scientific leader in the explanation and prediction of behaviour. This belief in the consistency thesis from chapter 1, claim J and P, consists in intentional states reducing to say neuroscientific states or states postulated by the One-True-Cognitive-Psychology; they allow that intentional psychology might taxonomise and organise states of agents differently from that of its more scientific cousins. From Part I we know that in so doing these enterprises will constitute different level of explanation-description. I have preaching throughout this work that intentional psychology does so constitute a distinctive level of explanation description.

From this sensible position, Jackson and Pettit then go on to argue that “beliefs and desires will very likely be found within what completed neuroscience tells us” (Forthcoming p. 20), and again, “Completed neuroscience will indeed provide a complete story about when and why we do what we do, but will incorporate rather than eliminate beliefs and desires in this complete story” (p. 23). The idea seems to be that intentional psychology will “show which part of any likely complete neuroscience story is the part which says (though not in so many words) that there are beliefs and desires” (p. 21). They seem to be saying that because we are dealing with distinct levels of explanation then neuroscience tells us that the roles we identify through intentional psychology really are filled. That is, neuroscience might tell us that our brains are just receivers of signals from Mars, where Martians controlled us (Jackson & Pettit, Forthcoming p. 14), thereby contravening the principle of agency (from chapter 7). In such a case we would not possess beliefs and desires even though we acted as if we did have them. This keeps intentional psychology at a different level from the One-True-Cognitive-Psychology or the neurosciences, but allows the possibility of there being evidential and/or supervenient relations between them, the former relation being allowed by the considerations of chapter 3.

Now, assuming that the intentional psychological level is a different level from that of the successful explainer and predictor of behaviour, how might we utilise such a level of explanation-description within psychological
The easiest way would be to conclude that the domain of psychology really encompasses various levels of explanation-description. Just as there are many levels to the neurosciences—neuroanatomy and neurophysiology—there may well be many levels to the psychological domain. There is neuropsychology, as we have seen above, cognitive psychology and social psychology, and even psychiatry, maybe. A clinical psychologist, for example, might find it useful to draw from elements of each of these fields.

Where does intentional psychology fit into the picture? If intentional realism is false, and intentional states do not feature in The-One-True-Cognitive-Psychology, then there is always the possibility that intentional psychology might find its home in one of the higher level areas falling within the purview of the psychological domain, social psychology or psychiatry, say. Indeed, these enterprises make use of intentional states now, and may well make use of them in the future, despite intentional realism’s being false. This assumes, of course, that we will still be pursuing social psychology and psychiatry in the face of The-One-True-Cognitive-Psychology. The-One-True-Cognitive-Psychology might well turn out to be a powerful weapon in the explanation and prediction of human behaviour, but I have as yet to see any arguments that it will exhaust the enterprise of psychology. If it doesn’t, then there still may be some psychological role for intentional psychology to play assuming intentional realism to be false.

3 The Social Sciences

There is a final means by which intentional psychology may be salvaged: we may appeal to even higher level disciplines than any so far mentioned. At an even higher level of explanation-description there exists economics, sociology, and anthropology. The social sciences have at times a need to attempt explanations and predictions of behaviour. Because these enterprises are explanatory-descriptive levels, they will quantify over kinds that suit their explanatory ends. To the extent that they do appeal to states of agents in forming explanations and generating predictions of their ensuing actions, then the social sciences will also appeal to the states of agents that will do the explanatory work required of the particular levels in question. It might be the case that the social sciences will only require to advert to Level One states of agents in order to frame the generalisation peculiar to their discipline. The Level Two states being of no consequence from the point of view of those high level disciplines.
The failure of intentional realism is regarded as a massive catastrophe for the social sciences according to some of the literature on folk psychology.\(^3\) I think this view is fuelled, in part, by the assumption that the singularity of explanation thesis holds true. If one assumes that there is some other calculus for generating explanations and predictions of human behaviour other than The-One-True-Cognitive-Psychology such as intentional psychology, then even though that alternative calculus is not as explanatorily and predictively accurate as The-One-True-Cognitive-Psychology in that its generalisations are less than perfect, nevertheless those generalisations may be adequate from the social scientific point of view. The mistake behind the pessimists regarding the status of the social sciences with respect to a beliefless and desireless One-True-Cognitive-Psychology is they assume that those social sciences must take their psychological ontology from The-One-True-Cognitive-Psychology. As with the rest of their ontology, high level disciplines such as the social sciences are free to taxonomise the world in the way they want, to choose ontologies that best suit their explanatory-descriptive ends.

---

\(^3\)See for instance Stich (1983 p. 228-9).
In 1802 Franz Joseph Gall (1758-1828) was thrown out of Vienna by decree of Emperor Francis I, he was excommunicated by Pius VII in 1817 and generally denounced by the medical establishment throughout Europe (John Marshall 1980 p. 24). Gall was, after all, the founder of what was to become known as phrenology: the doctrine that important traits of character can be determined from a study of the bumps on the skull (Robert M. Young 1970 p. 9).

As a young school boy growing up in the middle part of the eighteenth century, Gall observed that classmates with prominent eyes tended to have good memories. Subsequent observations of fellow students at university led him to conjecture that he had a genuine hypothesis relating physical features to a cognitive capacity, a hypothesis which he believed was capable of being tested. How he went about attempting to prove this hypothesis provides both Gal’s major contribution to psychology and his methodological downfall.

He says:

I could not believe, that the union of two circumstances which had struck me on these different occasions, was solely the result of accident. Having still more assured myself of this, I began to suspect that there must exist a connection between this conformation of the eyes, and the facility of learning by heart. (Quoted in Young 1971 p. 13)

Gall then proceeds to generalise:

Proceeding from reflection to reflection, and from observation to observation, it occurred to me that, if memory were made evident by external signs, it might be so likewise with other talents or intellectual faculties. From this time all the individuals who were distinguished by
any quality or faculty, became the object of my special attention, and of systematic study as to the form of the head. (Young 1971 p. 13)

However, Gall was later to qualify these remarks by the following:

I had in the interval commenced the study of medicine. We had much said to us about the functions of the muscles, the viscera, etc., but nothing respecting the functions of the brain and its various parts. I recalled my early observations, and immediately suspected, what I was not long in reducing to certainty, that the difference in the form of the heads is occasioned by the difference in the form of the brains. (Young 1971 p. 13)

So Gall transcends his purely physiognomical beginnings to become a propagator of functional neuropsychology.

Today, we (although not everyone, unfortunately) dismiss phrenology as utter nonsense and qua phrenologist Gall deserved any scorn heaped upon him. But Gall lived a double life; he was at once a craniologist interested in cranioscopy and at the same time a skilled anatomist using advanced dissection techniques. Under this latter guise Gall achieved early justified eminence as a neuroanatomist, setting the groundwork for much of modern neuroscience. For instance, it was Gall who distinguished cortical grey and white matter and differentiated projection, associative and commissural fibres (C.G. Gross 1985 p. 17). Gall also leaves us with a psychological legacy: the beginnings of VFT. Where Gall’s credibility comes unstuck is the false physiognomical suppositions accompanying his neuropsychology, and poor empirical methodology. We can summarise the main tenets of Gall’s programme thus:

(1) Mental capacities develop differentially across both inter- and intra-species individuals.

(2) Mental capacities depend upon fundamental innate faculties.

(3) Faculties are localised in specific organs of the cerebral cortex.

(4) The prominence of the faculties is a function of the size of the cortical organ.
(5) The size of the cortical organs is represented by the shape of the skull or cranium.¹

From the point of view of the history of neuropsychology (1) through (3) are perfectly respectable theoretical hypotheses. That history also tells us that (4) and (5) are pure physiognomy; (4) has proved to be wrong and (5) is little better than patently untenable. In fact Gall actually says:

The moral and intellectual dispositions are innate; their manifestation depends upon organisation; the brain is exclusively the organ of the mind; the brain is composed of as many particular and independent organs, as there are fundamental powers of the mind—these four incontestable principles form the basis of the whole physiology of the brain. (Herrnstein and Boring 1965 p. 219)

It is, therefore, possible to interpret Gall as believing (1) through (3) to be the central hypotheses of his programme. Gall, in fact explicitly rejected the craniological label attached to him: “They call me a craniologist, and the science which I have discovered, craniology. I rather think that the wise men have baptised the child before it was born. The object of my researches is the brain” (Marshall 1980 p. 24). We can see the Gallean antecedents of VFT by taking a look at (1) through (3) from a contemporary perspective.

1 Individual Differences and Competence

Gall’s interest in individual differences takes three forms. The first is the degree to which a particular individual’s cognitive capacities differ according to the task at hand. Thus, Smith might be particularly good at remembering numbers, but hopeless at remembering faces. How the deployment of intra-personal faculties can vary across content domains in this way plays a central role in Gall’s arguments for how faculties are localised. Much more on this later.

Secondly, Gall is interested in the failure of cognitive capacities to correlate across individuals. Jones is good at remembering things but cannot add or subtract whereas Smith can perform immense calculations in his head but cannot remember his telephone number. According to Fodor (1983 pp. 18-20), Gall has the annoying tendency to run intra-individual differences with these differences across individuals. To be sure, there are two distinct

¹This list is a summation of Gall’s hypotheses given by Gross (1985 p. 17) and Young (1971 p. 12).
motivations for attributing faculties here, and they ought to be kept distinct. As we shall see in the next section, I think both motivations can be retained within a VFT framework.

The third set of differences are inter-specific. Gall proposed that innate propensities and faculties are "unequally shared by different species of animals" (Marshall 1980 p. 24). Although Gall's reflections on inter-species differences are pre-evolutionary, he does recognise some cognitive capacities across species. He says: "The same organ, which in the nightingale produces singing, in the beaver the faculty to build, produces correspondingly in man music, architecture... the arts and sciences were not invented because of the necessities arising for them, but because of our innate disposition" (Marshall 1980 p. 24). Gall frequently analogises faculties to instincts in this way. Fodor claims this to be a problem for Gall since it does not sit very well with the argument from degrees of individual differences. Why? Because while both instincts and faculties are going to be genetically determined, the presence of an instinct is going to be inferred from competences undifferentiated across subject populations of species; but Gall's arguments from individual differences require differences in competences across subject populations of species. Or to put the problem another way, Gall wants individual differences to be inherited in the same way in which instincts are inherited; but that is to confuse issues regarding genetic determination and species specificity. What is instinctive is genetically determined, but not all genetically determined capacities need be instincts. Faculties attributable because of the evidence individual differences provide can be genetically determined without their being instinctual.

Although Gall's interest in intraspecific differences does not sit well with the instinct analogy, VFT has the option of restricting the instinct-faculty analogy to the discussion of inter-specific differences and not intra-specific differences. In fact, as evidenced in the last quotation, it is not, strictly speaking, inter-specific differences which are of interest to Gall at all, but inter-specific similarities. What VFT must do is factor two distinct arguments out of Gall here: inter-specific similarities and intra-specific differences.

At this point I should clarify just what it is that Gall is attempting to explain here. At various stages Gall refers to the explananda in question as "powers", "qualities", "instincts", "propensities", "aptitudes" and "talents". This is a decidedly mixed bag, the elements of which are not obviously coextensive. We can marshall the general idea of cognitive capacity or competence to aid Gall here. Particular faculties are postulated as a contribution to the explanation of a particular cognitive capacity such as
language competence. The set of faculties ultimately postulated will contribute to accounting for overall cognitive competence. How coarse or fine grainedly Gall and VFT carve up competences is crucial; we shall examine this question below.

2 Explanation by Faculty

Gall’s postulation of particular faculties responsible for mental capacities ought not, strictly speaking, be taken as an attempt to explain those capacities. If Gall was doing that then he would suffer the circularity inherent in the old faculty psychologies. To postulate a faculty of amativeness to explain amativeness is no better an explanation than that offered by Moliere’s physician who explained that opium produces sleep because it has a soporific tendency (Young 1971 p. 22). Instead, Gall’s postulation of faculties ought to be seen as an assumption in a larger scale explanatory programme, in terms of an evolutionary account of the existence of Level Two faculties and the detailed functioning of the brain. Unfortunately, Gall does not go much beyond this assumption. He has a lot to say about what faculties there are—that is what the functions of the brain are; but little to say about how the brain functions (Young 1971 p. 22). For Gall, knowledge of the functions of the brain must precede knowledge of the structure of the brain (his empirical methodology was to suffer because of this belief).

Having put the question of circularity aside, two features of faculties should be noticed. The first is that Gallean faculties are innate. By innate we must take Gall to be claiming that faculties are genetically determined. His nativism, though, is far from naive. As Marshall summarises: “The claim that the ‘elementary qualities of the mind are innate’ would nonetheless require, of course, that these qualities must be ‘drawn out and cultivated’ by the environment” (Marshall 1980 p. 23). This allows for the possibility of triggering in the development of these “fundamental powers of the mind”. Possession of a particular innate faculty implies a basic competence, but unlike instinctive competence Gall must have differentiation within a species if he is to account for individual differences.

How differentiation in competence is to be explained is going to be a problem for Gall’s project. While he can account for individual differences by appeal to difference in the performance of the underlying faculties across those individuals, he is not going to be able to account for differences in

---

2Although for how much one can explain in this way see chapter 2.
Appendix A

performance within the individual. In short, he required a performance-competence distinction.

This brings us to the second feature: the fundamentality of faculties. Gall requires some principled way of determining just how many faculties there are. As we shall see in the next section he goes about this determination on the basis of categorising individual differences in cognitive capacities from anecdote collection. The trouble is though that there are too many individual differences to go around; a one-to-one correspondence between striking behaviour and faculties is psychologically implausible. So Gall's faculties are fundamental in the sense that there are a few faculties out of which varied behaviour can arise depending upon the "mix" of the basic or fundamental faculties.3

Fodor claims that Gall denied the existence of horizontal faculties such as memory, judgement and reason. But as we saw in the previous section, this is a misrepresentation. What Gall is denying is that these horizontal categories can constitute faculties. It is possible on Gall's programme for there to be memory, judgement, sensation, etc., but these categories would apply within each vertical faculty associated with a particular cognitive domain rather than cutting across those domains.

That the faculty theorist must be careful about individual differences can be seen in the subsequent identification of more faculties after Gall. Spurzheim, Gall's collaborator, extended Gall's original list of twenty-seven to thirty-five. By the end of the century there were over one-hundred phrenological faculties postulated (Gross 1985 p. 17).

Why might cognitive architecture be arranged vertically? Gall's reasoning was something like the following. If there is only one faculty cutting across different domains then the performance of that faculty should remain constant across task domains. But there is no such constancy. Individuals can be good at remembering faces but hopeless at numbers. Fodor claims this reasoning to be fallacious. All Gall has shown is that facial memory is distinct from numerical memory which is, he claims, compatible with facial and numerical memory being exercises of the self same faculty with respect to faces in the one instance and numbers in the other. As he says: "there is no obvious reason why the same faculty should not be strong in one employment and weak in another, so long as the employments are not themselves identical" (Fodor 1983 p. 17). But this won't do. Fodor needs to cash out his use of 'employment' here. One reading seems to be that different employments result from different mechanisms operating. This is exactly what Gall needs to show. He would be quite happy to show "merely" that

---

3The question of faculties mixing and interacting will become important below.
facial memory is distinct from numerical memory since each form of memory is restricted to a particular cognitive domain. Gall does not, despite Fodor’s belief to the contrary, have to eschew the term ‘memory’ from each domain provided that there is no domain cutting psychological mechanism corresponding to that horizontal faculty.

But another reading is possible. Different employments might mean differential inputs: number representations on the one hand and facial representations on the other. If there is one memory with different inputs then we can expect different outputs. This, however, does not explain the differential performance of the mechanism involved. If it is a truly horizontal mechanism the representations (be they of faces or numbers) are going to be uniformly processable by that mechanism, no matter what they are representations of. That is, no matter what the representations are representations of, those representations will be processable by the mechanism due to some properties—syntactic properties, say—of the representations that allow them to be so processed. Whatever those properties requisite for processing are, both representations of numbers and faces are going to possess those properties if there is a horizontal faculty in operation here.

Two other options are available. It might be the case that it is the memory’s accessing the different modalities from which information is to be stored that affects how easily different types of information are stored. But that is another problem, separate from the question of the employment of the memory mechanism. This option does, however, call into question Gall’s argument from intra-individual, cross domain, differences. The differences in intra-individual memory performance might result not from there being different memory systems, but from the different performances of, say, different domain specific perceptual mechanisms along the lines of Fodor (1983). It follows then that the argument from individual differences is not sufficient for Gall’s claim here, since the data provided by individual differences does support various cognitive architectures. Anyhow, this remains a different objection from that introduced by Fodor since it’s not the employment of the mechanism per se that causes problems for Gall.

The final option is that the one mechanism processes various information differently. This will automatically give Gall his point since there will be distinctive processing involving different types of information (numbers and faces, say) in accordance with the domain specificity of vertical faculties. So much for individual differences.

Fodor also questions Gall’s methodologically important arguments from inter-individual differences. As we have seen Gall cites cases in which Smith and Jones might differ in musical ability. Fodor claims that this should
not lead one to postulate a special vertical musical faculty. Musical ability might just result from a certain “mix” of various horizontal faculties—or, for that matter, interactions between other vertical faculties. There is going to be an optimal mix for each of the vertical categories stipulated by Gall and there will be differences in the extent to which individuals approximate possessing that optimal mix (Fodor 1983 p. 19).

At the level of individual differences across a single cognitive domain Fodor’s response seems appropriate enough. But in the case of explaining individual differences across a variety of tasks Fodor’s reply breaks down. If Smith and Jones have approximately the same musical ability due to having similar “mixes” of horizontal faculties, it would seem to follow that they should also have roughly equivalent competences in other cognitive domains as well. Why? Because whatever the relevant properties of horizontal faculties are that contribute to a particular mix those properties will presumably remain constant across whatever cognitive domain is involved. What I have in mind here is this. Say the mix for a particular cognitive ability is two parts perceptual acuity, one part sensitivity with a dash of judgement. Now the strength or weakness of these individual faculties is fixed for the individual; that’s why some of us are more or less better at some cognitive skills than others for our entire lives. So no matter in what other mixes these faculties take part in they will be weak or strong. Suppose that some other ability consists of one part perceptual acuity, three parts sensitivity and lots of judgement. If Smith and Jones are equal in musical ability then we should expect that they would also be equal in this other ability as well, just because those very same horizontal faculties contribute to that other ability.

There’s a reply to this. It claims that we should not assume that Smith and Jones’ equal musical ability arises from equally strong or weak horizontal faculties. Maybe Smith is very “perceptual” and not very “sensitive” whereas Jones is just the opposite: not very perceptual but very sensitive; their respective equality of musical ability is just the averaging out of these differences. If this is the right story as to why Smith and Jones have equal musical ability then what guarantee do we have that this same averaging out will give equality of ability with respect to the other cognitive domain I have claimed they should also have equal abilities in? Perhaps differences in perceptual acuity and sensitivity will combine to generate different abilities in this other cognitive domain.

The trouble with this response is that although it might be the case that different proportions of horizontal faculties might affect the overall mix given constancy of competence within those horizontal faculties one needs to
show that such constancy will produce a different competence in the resultant mixed domain. Equally, I suppose, the defender of Gall would need to show that competence would remain constant for this other domain. But too little is really known about the psychological reality of these mixed non-fundamental faculties to decide which result would be forthcoming. I think the burden of proof lies here with Fodor. My challenge is that his argument against Gall only works when certain results obtain from comparisons across cognitive domains. Since it is not clear whether those results favour his case or Gall's, I think his original argument falls short of the mark.

Just in case the reader thinks that the defender of Gall should be burdened with some of the proof, some actual cases from neuropsychology seem to provide arguments which count against objections based upon mixes of horizontal faculties such as Fodor's. Take, for example, a hyperlexic. Such a child is characterised by developmental lags in which the onset of speaking and walking are delayed. The distinguishing feature of the hyperlexic is that by the age of three or four, despite slowness in other cognitive areas and without appreciable help from environmental factors such as parental intervention, he or she has mastered oral reading. Gardner recounts:

A boy aged four years and ten months ... could read a third grade level passage fluently despite woefully inadequate speech and understanding and an IQ in the mentally defective range; so much he enjoy this activity that he sought to read all materials in sight. A three year old youngster seen by these investigators could read newspapers aloud and also recited everything in sight including dictionaries and telephone books. (Gardner 1977 p. 135)

Gardner also quotes from a case study in which a six year old could not desist from reading:

Reading had the features of a compulsive ritual: he could not be distracted or easily deterred. If he were given a book, he had to start at the beginning, reading the entire title page, including the publishing house, date of publication, Library of Congress number, etc. He showed great resistance to interruption and would return later to this point and carry on the exercise. Gardner 1977 p. 135)

There is no evidence to suggest that what the hyperlexic reads aloud is comprehended:

Reading for him consists strictly in the oral decoding of visual graphemes: his prosody and accentuation do not vary with the sense,
nonsense words are read as effortlessly and emphatically as meaningful ones, jokes are not laughed at, commands are not honoured, and even references to the present situation may be missed. (Gardner 1977 p. 136)

Another case I want to consider is that of L. Again the quote is from Gardner.

This boy, eleven years of age at the time when he was most intensively studied, had an IQ of 50, placing him in the severely retarded range. Deficiencies of information and reasoning capacities notwithstanding, this freakish youngster could perform all sorts of numerical feats. He was able to remember endless series, such as railroad timetables and newspaper financial columns. He could immediately state the day of the week for any date between 1880 and 1950. Given twelve two-place numbers to sum, he came up with the total the instant the presentation was completed. His speed and accuracy at other arithmetical challenges were equally impressive. (Gardner 1977 p. 231)

L.'s numerical abilities were detected early; at the age of five years he could count by 2's, 4's, 8's and 16's. Another famous calculator was Fleury who, while being severely retarded, provided the cube root of 465,484,375 (=775) in thirteen seconds (Patricia Churchland 1986 p. 232).

What is peculiar about both the hyperlexic and idiot savants such as L. is the other cognitive capacities they exhibit. Whatever horizontal cognitive mechanisms we might want to attribute in such cases, the hyperlexic and idiot savant display extreme cognitive deficiencies in all other cognitive domains. If one wanted to explain the hyperlexic's proficient reading in terms of the mix of some set of horizontal mechanisms, one needs to show how that mix can be produced in the apparent absence of competence in any horizontal domain. There is no problem in explaining the mix in the case of individuals with an adequate spread of cognitive capacities since there are capacities out of which such a mix can eventuate. This need is especially evident in the case of hyperlexia. Any of the horizontal capacities we might want to attribute are just lacking in these cases. To attribute the child's ability to its "remembering" all the phoneme-to-grapheme group correlations necessary for the production of this incessant reading (in the face of the general language deficiencies evident) is exactly what we do not want to do. Again, we seem to have exhibited a range of cognitive abilities that varies across individuals,
where those abilities are unique in the cognitive economy of the subject. Hence some underlying mechanisms specific to those abilities.

3 Cerebral Localisation

Crucial to Gall’s programme is the idea that faculties are localised in specific organs within the brain. First organs and then localisation.

3.1 Mental Organs

Central to Gall’s thinking here is the analogy between cognitive functions and anatomical organs. Chiefly, Gall wanted to, and needed to given his contemporaneous intellectual climate, argue that the brain was the organ of the mind. The brain in turn gets subdivided into smaller organs each being responsible for particular cognitive functions. Given that other anatomical functions have particular organs of their own—seeing, hearing, salivating etc.—Gall asks “But, if she [Nature] has constructed a particular apparatus for each function, why should she have made an exception of the brain? Why should she not have destined this part, so curiously contrived, for particular functions?” (Young 1971 p. 19).

3.2 Localisation

Young defines cerebral localisation as “the doctrine that various parts of the brain have relatively distinct mental, behavioural and/or physiological functions” (Young 1971 p. 10). Localisation of function is hardly new; it is as old as anatomy and physiology. Early forms of localisation were inadequate; they were ventricular, speculative and based on the old faculty psychology. They localised function not in the solid regions of the brain but in the ventricles. They were non-empirical. The faculties postulated were derived from the Platonic division of the soul and mind. So, sensation and imagination were localised in the anterior, reason or thought in the middle and memory in the anterior ventricles. Later the centre of localisation was to shift from the ventricles to the solid cerebral structures. Gall obviously rejected the ventricular nature of the old faculties; but his most important contribution was to reject their speculative nature and in so doing reshape

---

4Young (1971 p. 19). Other faculties postulated often included the will, judgement and language.
the kinds of faculties needed to be postulated. Much more will be said about localisation in chapter 9.

4 Empirical Methodology

One reason for Gall's rejection of the old faculties was that psychology thus far had been non-empirical. The faculties had grown largely out of philosophical psychology. It is Gall's chief contribution to contemporary functional psychology that he made the shift to an empirical methodology. Gall intended to replace the speculatively derived faculties postulated by the philosophers with empirically derived faculties which reflected the mental capacities of individual organisms and were the determinant variables in individual behaviour (Young 1971 p. 16). As Young says of Gall:

His conception of the domain of psychology makes their [the philosopher's—JF] categories quite useless. Gall's faculties are designed to serve a purpose quite different from those of the philosophers. He sees the goal of psychology as a differential one with its domain as the behaviour, roles, talents and differences of men and animals. (Young 1971 p. 18)

Gall thus argues:

None of the faculties mentioned, describes either an instinct, a propensity, a talent, nor any other determinate faculty, moral or intellectual. How are we to explain by sensation in general, by attention, by comparison, by reasoning, by desire, by preference, and by freedom, the origin and exercise of the principle of propagation; that of the love of offspring, of the instinct of attachment? How explain, by all these generalities, the talents for music, for mechanics, for a sense of the relations of space, for painting, poetry, etc.? (Young 1971 p. 18)

The existence of the philosopher's categories is not denied by Gall. He claims they are merely abstractions and generalities:

they are not applicable to the detailed study of a species, or an individual. Every man, except an idiot, enjoys all these faculties. Yet all men have not the same intellectual or moral character. We need faculties, the different distribution of which shall determine the different species of animals, and their different proportions of which
explain the difference in individuals. All bodies have weight, all have extension, all are impenetrable in a philosophical sense; but all bodies are not gold or copper, such a plant or such an animal. Of what use a naturalist the abstract and general notions of weight, extent, impenetrability? By confining ourselves to these abstractions, we should always remain in ignorance of all branches of physics, and natural history. ...From most ancient to most modern, they have not made a step further, one than another, in the exact knowledge of the true nature of man...(Young 1971 p. 18. Emphasis added)

The argument in this passage is not convincing. An abstract notion such as weight qua atomic weight would seem to provide the naturalist with a good criterion for taxonomising natural kinds, for instance. The important point for our purposes here, though, is the italicised sentence. By ‘empirical’ Gall means not that functional psychology should be interested in studying and accounting for brain localisation to behaviour correlations central to modern neuropsychology. Rather, he means that one should be studying and accounting for the intra- and inter-species differences of organisms. It is because of his erroneous assumptions (4) and (5) that Gall was to look for these differences in cranial prominences and anecdotal evidence involving striking behaviour. By and large, his neuro-anatomical studies played no part in his data collection. Where neuroanatomical details were included they were restricted to natural mutilations where the localisation had been established on other grounds (ie. cranial and anecdotal grounds). Gall rejected experimental ablation to either humans or animals, not for ideological reasons involving, say, vivisection, but various technical and theoretical reasons. Examples being the imprecision of his contemporary surgical techniques and the fact that the structure of the brain requires that “a part being wounded or irritated, wounds or irritates all the rest” (Young 1971 p. 49). However, no matter how imprecise are surgical techniques or how pervasive the non-isolatability of brain lesions, accidental mutilations of brain tissue are going to be less isolatable than experimental lesions. Again, Gall's reasoning seems to lead him methodological problems.

The upshot of this is that Gall is left with the largely physiognomical empirical methods of craniology and anecdote collection. And these forms of data compilation led him into serious error. For instance, because of a protuberance over the ear was found on a medical student so fond of testing animals that he became a surgeon, the organ of destructiveness was placed there. And as Gross says: “the organ of amativeness was placed in the cerebellum because Gall had noticed that a passionate widow’s neck was hot to his touch!” (Gross 1985 p. 17). This uncritical methodology also allowed the
easy explaining away of counter-examples. The most amusing being that revolving around Descartes’ skull. The skull was found to have remarkably small anterior and superior regions of the forehead where the rational faculties were located. Spurzheim replied that Descartes was not as great a thinker as we had previously thought! (Young 1971 p. 43).

So despite his claims to be interested in the brain and not cranioscopy Gall’s methodology was empirical but it was at the same time physiognomical. His theory failed to be consistent with his practice.

5 Early Objections to Faculties

The idea that mental capacities depend upon the brain was around a long time before Gall. The principle of localisation that Gall was arguing for went beyond this mere dependence; on his account fundamental faculties could be located in specific areas of the brain. Gall’s chief critic in the nineteenth century was Pierre Flourens. Flourens also believed that the brain was the organ of the mind, and gave credit to Gall for the establishment of the point. He says:

the proposition that the brain is the exclusive seat of the soul is not a new proposition, and hence does not originate with Gall. It belonged to science before it appeared in his Doctrine. The merit of Gall, and it is by no means a slender merit, consists in having understood better than any of his predecessors the whole of its importance, and in having devoted himself to its demonstration. It existed in science before Gall appeared—it may be said to reign there ever since his appearance. (Quoted in Young 1971 pp. 20-1)

Despite this agreement and Flourens’ recognition of Gall’s contribution—in fact, Flourens held Gall in utter contempt—there is little else upon which they agreed. They had methodological disagreements. Flourens, for instance, was committed to experimental ablation (surgical removal of various components of the nervous system) in his data collection; his precise use of ablation becoming the standard practice in cerebral research. However, the psychological aspects of his work is not strong. He was particularly weak when it came to the behavioural consequences of his physiological experiments. He had no serious criteria for loss of function (Young 1971 pp. 30-1). Indeed, lacking from Flourens work is a set of psychological categories that his ablation experiments were designed to test. Gall recognised this deficiency when he said:
It would have been requisite to know what could be found, and what ought to be sought for, in the brain. It would also have been necessary, that the mutilators should be divested of every metaphysical prejudice; that they should have a detailed knowledge of the fundamental powers. Where is the physiologist, where, the anatomist, who has been able to follow this direction, and who has not wished to find generalities and abstractions. (Young 1971 p. 70)

Young claims that Flourens granted this point in principle, although there is no evidence of it affecting his research. (Young 1971 p. 71)

From the point of view of methodology, Gall and Flourens are interesting contrasts. Flourens was especially sound when it came to the physiology of the brain although his methods regarding the psychological affects of that physiological work are very poor. On the other hand, Gall's acceptance of the cranium-brain correspondence meant that the physiological underpinnings of his programme were unsound. But he, at least, rejected an introspective appeal to establishing psychological generalisations and opted for an investigation of overt psychological phenomena (Young 1971 p. 73).

Part of the reason for these disagreements was the different heritages each saw himself working within. Gall saw himself reacting against the old introspectionist philosophical psychology in favour of a modern empirical psychology. Flourens saw himself as attempting to justify the philosopher's categories of the mind from an empirical perspective. He says:

I frequently quote Descartes: I even go further; for I dedicate my work to his memory. I am writing on opposition to a bad philosophy, while I am endeavouring to recall a sound one .... "I remark here, in the first place," says Descartes, "that there is a great difference between the mind and the body, in that body is, by its nature, always divisible, and the mind wholly indivisible. For, in fact, when I contemplate it—that is, when I contemplate my own self—and consider myself as a thing that thinks, I cannot discover in myself any parts, but I clearly know and conceive that I am a thing absolutely one and complete".

Now here is the sum of Gall's psychology. For the understanding, essentially a unit faculty, he substitutes a multitude of little understandings or faculties, distinct and isolate.

Gall reverses the common philosophy ... According to common philosophy, there is one general understanding—a unit; and there are faculties which are but modes of this understanding. Gall asserts that
there as many kinds of peculiar intelligences as there are faculties, and that the understanding in general is nothing more than a mode or attribute of each faculty. (Young 1971 p. 72)

It is seeing Flourens in this light, viz. as a Cartesian, that one can see his greatest disagreement with Gall. Of the plurality of faculties he says:

There are as many faculties as there are truths to be known ... But I do not think that any useful application can be made of this way of thinking; and it seems to me rather more likely to be mischievous, by giving to the ignorant occasion for imagining an equal number of little entities in the soul. (Young 1971 p. 71)

Gall’s philosophy consists wholly in the substitution of multiplicity for unity. In place of one general and single brain, he substitutes a number of small brains; instead of one general sole understanding, he substitutes several individual understandings. (Young 1971 p. 71 & pp. 71-3)

He is willing to grant that men and animals show very different mental capacities (and again the Molierian objection):

No doubt of it. But what sort of philosophy is that, that thinks to explain a fact by a word? You observe such or such a penchant in an animal, such or such a taste or talent in a man; presto, a particular faculty is produced for each one of these peculiarities, and you suppose the whole matter to be settled. You deceive yourself; your faculty is only a word—it is the name of the fact—and all the difficulty remains just where it was before. (Young 1971 p.71)

Flourens also claims empirical support for the thesis that “the cerebral hemispheres concur, by their whole mass, in the full and entire exercise of the intelligence” (Young 1971 p. 73). Flourens sees any qualification to the unity of the soul or its organs (ie. the unity of the brain) to be a denial of the existence of the mind.5

We can see in these passages the beginnings of CSR inspired objections. Flourens believes that there is one “general sole understanding”, one single process underlying the mental capacities of individuals. Such a system has to be horizontal in nature since a unitary mechanism is

5Young (1971 p. 73). Unfortunately, I cannot assess Flourens' actual arguments here since my source does not list them. At the time of this draft Flourens' original work was unavailable.
necessarily undifferentiated with respect to cognitive domains. If, as he says, the whole mass of the brain is exercised in the deployment of intelligence, then, in principle, we cannot strictly localise the capacity in any specific region of the brain. The 'strictly' is important here since Flourens allowed for localisation of function—his proper action—combined with a form of equipotentiality—his common action (Herrnstein and Boring 1965 pp. 221-3). His final assumption is that only a unitary cognitive system can yield a unified consciousness, mind, soul or whatever.

This brief summary of Flourens work has set the stage for the contemporary VFT-CSR debate. It also shows us that what is generally regarded as the tension between philosophical psychology and empirical psychology is far from a recent development.
Appendix B

Neuropsychology

The paradigm within which we have been exploring, contains a background assumption to the effect that the mind is, at least, supervenient upon, or stronger, identical with the brain. Given this assumption, it is not surprising that psychologists and neuroscientists alike have taken an interest in those unfortunate persons who have suffered some sort of brain damage, in the hope of learning something about the mind. Enter the cognitive neuropsychologist who attempts to develop theories about normal cognition from those suffering impaired cognitive abilities due to acquired brain lesions. What I want to do in this appendix is to peruse some pages of the neuropsychologist's clinical notebook to see what light can be cast in the direction of the current dispute between VFT and CSR. Although the evidence available is somewhat indirect, I think it supports the principles underlying VFT.

1 Isolation by Brain Damage

The most common causes of brain damage induced cognitive disfunction are strokes and cerebral intrusions such as automobile accidents. It is a neuropsychological commonplace that brain lesions produce very fine grained cognitive disorders. Let's look at some in some detail.

1.1 Aphasia

Following Gardner, we may define aphasia as a disturbance in language function following injury to the brain. Characteristic of aphasia is not an across the board reduction in a person's verbal capacity. Instead, capacities such as talking, understanding, reading and writing can be spared or
destroyed in relative isolation from each other. Some combinations of these capacities can be destroyed or spared at once or never at all. It is not even the case that more complex linguistic functions such as comprehension of long sentences are impaired before more simple functions such as repetition or following commands (Gardner 1977 pp. 52-3). Take, for example, what has become known as Broca's aphasia. In this form of aphasia verbal comprehension usually seems more or less spared whereas speech production is vastly impaired. The Broca's aphasic is reduced to the use of content words; a huge disparity exists between his or her ability to name objects and the ability to use grammatical terms. 'If', 'and' and 'but' are almost totally inaccessible. Thanks to relatively intact comprehension the Broca's aphasic can answer questions such as "Does a stone float on water?" and "If I say "The lion has killed the tiger" which animal is dead?".

Contrast this affliction with a second form of aphasia: Wernicke's aphasia, named after Carl Wernicke who in 1874 hypothesised that the posterior portion (front and second temporal convolutions) of the left temporal lobe controlled language comprehension, as opposed to Broca's area of the same lobe which controlled language production. A person suffering this affliction is likely to utter the following collection of sentences:

Oh sure, go ahead, any old think you want. If I could I would. Oh, I'm taking the word the wrong way to say, all of the barbers here whenever they stop you it's going around and around, if you know what I mean, that is tying and tying to repucer, repuceration, well, we are trying the best that we could while another time it was with the beds over there the same thing ... (Gardner 1977 p. 68)

Such word salads are common with this type of aphasia. They articulate extremely well with excellent intonation and syntax. Accompanying this lack of coherent speech is an eerie nonawareness on the part of the sufferer that they have any language deficit. They also perform at a level no better than chance when answering the type of questions posed to Broca's aphasics. As to following commands, some preservation of comprehension is evident. One of Gardner's patients, for instance, could consistently carry out "whole body" commands such as "stand up", "turn around three times" and "assume the position of a boxer" while being unable to follow commands regarding specific limbs. "Raise your hand", "touch the chair" and "make a fist" all went unexecuted (Gardner 1977 p. 72). In the case of naming objects difficulty increased with diminishing familiarity. 'Book' and 'ear' would be recalled...
whereas only attempts could be made at less familiar objects such as ‘paper’
‘fork’ and ‘ankle’.

According to Gardner, in general, Broca’s aphasia and Wernicke’s
aphasia represent complementary clinical patterns. The former’s
strengths—comprehension and general intellectual capacity—are the weak
points of the latter. While the latter’s preserved capacities—the ability to
articulate easily, the inclusion of grammatical particles in speech—are
lacking in the former (Gardner 1977 p. 75).

A third form of aphasia called anomia presents a very different clinical
picture from those forms just mentioned. According to classical aphasiological
theory anomia results from damage to the angular gyrus—that portion of the
brain in which outputs from the various sensory systems are associated. The
anomic displays neither set of deficits exhibited by the Broca’s or Wernicke’s
aphasic. To the casual observer the anomic might appear to possess normal
language capacities. On closer inspection though, a striking impairment is
evident. If asked to name some common objects in a room the anomic is at a
loss. When Gardner pointer to a clock, for example, his patient responded: “Of
course, I know that. It’s the thing you use for counting, for telling the time,
you know, one of those, it’s a ...” (Gardner 1977 p. 76). In cases of severe
anomia the sufferer has difficulty in naming objects and if told the name the
sufferer exhibits considerable uncertainty about whether it was in fact the
correct name. As it turns out, anomias can often choose the correct name for
an object from a group of words—although he or she would remain uncertain
about the choice.

Perhaps the most fascinating aspect of anomia is that the inabilities
just described are evident only when the name of an object is required to be
produced in isolation from its usual context. So the anomic might well say
“Would you mind passing me the chair?” when at the same time he or she is
totally unable to produce the word ‘chair’ if requested. The classic case is
reported by Gardner. Due to anomia a person is unable to produce ‘no’ on
request. After many futile attempts the patient throws up his hands in
disgust exclaiming “No. No! I told you I can’t say ‘no”. Even after such an
outburst the patient remains unable to produce that word on request
(Gardner 1977 p. 78).

In order for a word to be produced out of context, it is thought that a
person must be able to abstract away from the present concrete situation.
This flexibility is what is thought to be lacking in the anomic. In examination
of anomia speech one can detect a certain concreteness, to get lost in minute

\footnote{See Gardner 1977 p. 70 for a description of the attempts made. There seems to be
some evidence here for the role of phonemic representation in verbal recall tasks.}
detail sometimes at the expense of the general point. In interpreting proverbs, for instance, there is a distinct tendency toward “concrete thinking”. In interpreting the saying “Too many cooks spoil the broth” one of Gardner’s patients was unable to rephrase it in order to bring out its meaning. His attempts were: “Too many cooks, you know, cooks standing around the broth, they are talking and cooking” and “If you want things to turn out the right way, you’ve got to be careful about that sort of thing” (Gardner 1977 p. 79).

1.2 Alexia

Alexia comes in at least two forms: pure alexia and alexia with agraphia. Pure alexia is the condition in which one loses the ability to read writing. A person with alexia with agraphia can speak and understand perfectly, but cannot read, write or spell. Gardner cites the case of pure alexia from the end of the last century, of a French businessman, call him C., who had a rare stroke which made it impossible to see objects in the right half of his visual field and was responsible for his alexia. Gardner recounts:

Surprisingly, however, C. was able to express himself without difficulty, to recognise and name instantaneously obscure technical and scientific instruments, to understand everything said to him, to recall the most minute detail details of past events. Even more astonishing, he could still write without difficulty, both expressing his thoughts spontaneously and transcribing what was dictated to him; yet he was quite unable to decipher his own handwriting, unless he could independently remember what he had written or had been dictated to him. In fact, he preferred to write with his eyes shut, for he got “tangled” when he monitored his own writing. When letters were etched on his hand, or when his fingers were guided through the air in the form of a word, he could instantly identify these verbal materials. In short, all his language functions, including writing, were preserved with the exception of the decoding of words and letters presented to his eyes. (Gardner 1977 pp. 115-16)

... He continued to play cards, for he recognised the different numbers and suits; furthermore, he was able to read numbers, to perform elaborate calculations, to follow his business and stock market investments. Why the particular lesion in his brain should destroy recognition of verbal and musical symbols, while sparing numerical ones, remains nearly as mysterious to neurologists today as it was to
C., but at any rate his capacities in the numerical realm were not in the least impaired. (Gardner 1977 pp. 116-17)

Certain written materials were understood by C. However, these materials were not understood as a result of reading as such. As Gardner says:

To be sure, when shown the French newspaper L'intransigeant he was unable to read its title in a fifteen minute trial; but when shown the newspaper Le Matin, he immediately called out its name. He could not identify the individual letters ‘R’ and ‘T’, but when a circle was drawn around ‘RF’, he instantly reported “Republique Francaise”. C. himself explained that these recognitions were due to an appreciation of the “form” or “picture” which these symbols assumed. Identifying Le Matin was not a matter of deciphering seven letters or three syllables, but rather the recognition of a familiar shape; ... C. was reading these signs in the same way that a Westerner might learn to recognise a Russian signature despite ignorance of the Cyrillic alphabet, or a child the name of a restaurant by the appearance of its billboard or marquee. (Gardner 1977 p. 116)

Contrast this with the alexic with agraphia. He or she loses this ability to understand letters or words as “graphic entities”.

An interesting variation to alexia occurs in the case of polyglots. In one case the patient was totally unable to read his first language, English. He could, however, read some Latin and read Greek perfectly.

Before the case of C. there had been no anatomical evidence for the causes of alexia. Luckily, a clear clinical profile of C. was available with which to compare post mortem evidence. It appears that the left half of C.'s visual field was intact as was the centre for language normally located in the left hemisphere of right handed individuals. Because of the stroke, C. was, supposedly, unable to transmit information from the preserved part of his visual system in the right hemisphere to the areas of the left hemisphere where lexical concepts are stored in the language centre. Thus, C. could name letters aloud, write, and recognise letters by touch since the pathways connecting the tactile as opposed to visual system to the language centre were intact.

But why could C. retain the ability to recognise and name people, objects and especially numbers? An answer is offered by Norman Geschwind. In studying a case of pure alexia without agraphia in the early 1960's, Geschwind and Fusillo discovered that pure alexia is accompanied by “an
inability to match seen colours to their spoken names" (Geschwind and Fusillo 1966 p. 146). They say:

The patient failed in all tasks in which he was required to match the seen colour with its spoken name. Thus, the patient failed to give the names of colours and failed to choose a color in response to its name. By contrast, he succeeded on all tasks where the matching was either purely verbal or purely nonverbal. Thus, he could give verbally the names of colours corresponding to named objects and vice versa. He could match seen colours to each other and to pictures of objects and could sort colours without error. (Geschwind and Fusillo 1966 p. 144)

Such deficiencies are not perceptually based, despite the patient's insistence that objects looked grey when they were in fact some other colour. The patient passed pseudo-isochromatic tests of colour vision, outscoring one of the examiners who was moderately red-green colour blind. The patient's confabulatory responses regarding colour names (ie. insisting that a red object was in fact grey) results from the speech centre's being disconnected from that area of the right hemisphere of the brain undergoing the visual experiences because of damage to the corpus callosum. The colour name offered by the examiner which is heard and hence processed in the left hemisphere cannot be compared with the colour seen which is processed by the right hemisphere.

Nor can this inability result from loss of verbal memory for colour names since the patient correctly named objects corresponding to named colours; the patient admits that bananas are yellow and paper is white but insists that these particular instances are a different colour.

The reason why reading and colour naming are lost whereas object naming and number recognition are not seems to be that letters and colours are accessed through the visual modality only; written words and colours have visual but not tactile associations, for instance. So when the informational links between the visual centre of the right hemisphere and the language centre in the left hemisphere are severed in the damaging of the splenium of the corpus callosum, the naming centre cannot access the visual centre. In contrast to this, objects in the world and people may have associations involving other sensory modalities including tactile, auditory and kinesthetic. In the classical pathology of pure alexia without agraphia those portions of the corpus callosum (viz. those portions anterior to the splenium) which carry nonvisual information and in particular tactile information from the right to left hemisphere are intact. So the access of the centre involved in
naming to the tactile centre is unaffected. Hence, naming of objects with tactile associations is possible.

Numbers, of course, do not exist in the world like objects and people. But unlike letters, words and colours, numbers have strong tactile associations—at least according to some neuropsychological hypotheses. Consider (and this can be no more than anecdotal evidence) the role of the fingers in the learning of numbers and how to count. So the retention of the ability to read numbers as well as naming of objects is explained by the continued access of the left hemisphere to right hemispheric nonvisual information.

1.3 Prosopagnosia

Disorders of recognition are called agnosias. As with aphasia, the existence of a general disorder here is open to dispute. More interesting perhaps, since the data is more clear cut in this case, is a particular form of agnosia—prosopagnosia. Prosopagnosics can identify objects, are intellectually intact, and possess normal language functions. They remember the individuals they have known, and can recognise them from their voices, descriptions of their appearance or from some idiosyncratic feature such as moustache, glasses or hat. Where the prosopagnosic fails is in the recognition of the person's face. Even close relatives or the sufferer's own face will seem totally unfamiliar. They will often express disbelief at being told the identity of what appears to be a stranger. They may also comment that the individual, even the patient himself, has drastically changed since they last met.

Our chief question with prosopagnosia is: is there a special mechanism which underlies the recognition of faces? Some support for a confirmatory answer comes from studies of the recognition of inverted faces (Gardner 1977 p. 155). Modern tensor network theory also provides a means for such a capacity, in which if the particular state-space sandwich in which facial features are represented is destroyed, there will be a selective impairment of this capacity alone. However, in order for there to be a unique facial recognition mechanism, lesions to this area of the brain would have to impair this capacity alone. But this does not seem to be the case. When it comes to fine discriminations such as distinguishing bird species or kinds of leaves, the prosopagnosic also runs into difficulty.

In defense of the existence of prosopagnosia it might be argued that the ability to make such fine discriminations results from the ability to make

---

2See Gardner (1977) ch. 4 for a full discussion.
3See P.S. Churchland (1986) ch.10 for details.
fine discriminations amongst facial features. So the advantage from an evolutionary perspective in the positing of a distinct facial recognition capacity in terms of mother-child relations and conspecific recognition might be generalised to operate in other cognitive tasks. A multipurpose state-space sandwich might well perform such generalised tasks.

2 Split Brain Studies

The human brain is divided into two hemispheres. It is possible through natural malformation and surgery for the connections between these hemispheres to be severed. By severing the corpus callosum and/or the anterior and posterior commissures one can cut the informational access of one hemisphere to the other. When such surgery, called commissurotomy, is carried out, usually in an attempt to reduce intractable epilepsy, there have been some interesting behavioural side effects. Much of the data resulting from experiments performed on these “split brain” patients has produced a popular split brain mythology, I will not discuss here. For present purposes, though, I restrict my attention to aspects of split brain studies that have a direct bearing upon the VFT-CSR debate.

A word of caution before we start. Not all split brain patients exhibit the same pathological histories. Some, due to brain lesions early in life, developed significant linguistic skills localised in the right hemisphere (which is rather uncommon for right handers). Moreover, the type of commissurotomy performed varies across patients. Some have the corpus callosum sectioned together with the anterior commissure, while others have that commissure spared. There is also the possibility of information exchange between the two hemispheres via other commissures and, indeed, the midbrain and brain stem at which points there are no hemispheric divisions.4

In human vision (as well as higher vertebrates) there is a division of labour in visual processing. Information from a person’s right visual field is processed by the left hemisphere and information from the left field processed by the right hemisphere. In normal individuals this division of processing poses no problems since the visual information being processed in each hemisphere is relayed to the other hemisphere via the commissures. But in the post-operative commissurotomy patient this information exchange is blocked. Thus, the opportunity to study two independent visual processing systems and the resultant affects upon behaviour this independence may bring about.

4For fuller description of some of the caveats associated with this line of data acquisition see P.S. Churchland (1986) p.174.
The main source of experimental data regarding visual processing comes from the use of the tachistoscope. By having a patient focus on a fixed midpoint on a screen with a projector flashing different images to each vertical side of the midpoint, the image to the left of the midpoint will be restricted to the left visual field (hemifield) and hence restricted to right hemispheric processing, while the image to the right of the midpoint is restricted to the right hemifield and hence to left hemispheric processing.

With this set up the word 'key' is flashed to the right hemifield. If asked to identify the visual stimuli, split brain patients have no problem; they would say "key" correctly. But if that word is presented to the left hemifield, and hence restricted to right hemispheric processing, the patient would fail to say "key" and if asked what was seen they would deny having seen anything. This is exactly what we would expect given the account of language production given above. If visual information cannot be transferred from the right to left hemisphere where language functioning in general and speech production in particular takes place then the subject naturally would fail to give a verbal response to the stimuli.

Interestingly, if the subject, having denied the presence of stimuli in the left hemifield, is asked to pick an item from a collection of hidden objects with one's left hand (remembering that the right hemisphere has motor control over the left half of the body) the subject correctly picks out the key.

It is concluded from these results that the reason the subject fails to verbally respond to stimuli in the left hemifield is due to the left hemisphere's being the linguistically competent hemisphere. But notice that the word 'key' was flashed to the right hemisphere for processing and not, say, a pictorial representation of a key. Thus the right hemisphere would seem to have some language comprehension skills. Otherwise, the subject would be unable to pick the key from the collection of objects.

As a variation on this experiment the patient is asked to draw with his or her left hand a picture of a word flashed. Typically, these patients protest that the request is pointless since nothing was shown on the screen. When coaxed to have a go at drawing something the patient accurately draws a pictorial representation of the referent of the word presented. Upon producing the drawing the patient will exclaim that he or she does not know why it was that particular picture drawn, and hypothesises that something must have been presented on the screen. Clearly, in these cases there must be linguistic comprehension evident on the part of the right hemisphere (Patricia Churchland 1986 pp. 176-7).

There has to be verbal comprehension on the part of the right hemisphere since in each experiment the subject understands a task command.
Assuming that the task commands are received by both hemispheres from auditory information presented to both ears, and that the exchange of auditory information processed by a language input system (wherever it is) is blocked due to the commissurotomy, then if a command is carried out under the control of the right hemisphere, then that hemisphere only could have processed that auditory information. It is not possible for the left hemisphere to have processed the command and then passed on control of the output behaviour resulting from the original stimuli to the right hemisphere since, ex hypothesi, the connections between the hemispheres have been cut. Patricia Churchland claims that task commands are not purely verbal, and that non-verbal cues accompany the command (Patricia Churchland 1986 p. 184). In order for this to count against the right hemisphere’s having linguistic comprehension it would be necessary to show that the non-linguistic component of the task command is sufficient for the subject to understand the command. I am unsure that this is the case.

It was initially hoped that split brain studies would provide the basis for distinguishing between particular styles and content domains of cognitive processing at an inter-hemispheric level. In particular it was hoped to show that the left hemisphere contained the centre for language and processed information somehow analytically whereas the right hemisphere was superior in the synthetic processing of spatio-visual tasks. But this hope appears to have been dashed; it might well only be the case, and there even exists contrary evidence for this claim as well, that speech production is localised to one hemisphere only.\(^5\) If that is the case then we have good reason to infer that speech production is a cognitive process that can function relatively independently of much other cognitive processing. This is exactly what the VFT predicts.

Another tachistoscopic experiment uses chimeric stimuli (Patricia Churchland 1986 pp. 177-9). Two pictures of half faces are joined together to form a whole face, with each half face being restricted to each hemifield. Thus, the left hemisphere would process the right half face and the right hemisphere would process the left face. Upon presentation of the stimuli, the subject was required to choose from a set of faces the one originally seen. The results showed that faces presented to the right hemisphere were more often identified than those presented to the left hemisphere. Whether or not the left hemisphere performed near a chance level I am unsure. But even on the model of localisation considered here, some recognitional capacities might be evident in that hemisphere thus increasing the performance beyond mere chance. If these results are genuine then attempts to localise the major

\(^5\)See Gardner (1977) Ch. 9 for some counterevidence.
mechanism for facial recognition in the temporal lobes of the right hemisphere are going to be justified. This further supports the claims of the VFT.

The disconnection effect brought about by commissure sectioning can also be simulated in normal individuals. Because not all individuals have motor control for speech localised in the same place, it is necessary to determine which hemisphere possesses this control before any major brain surgery to proximal areas of the brain is undertaken. By injecting a patient with sodium amytal (a suppressor of neural activity) into either carotid artery (which exclusively supply blood to either right or left hemisphere only) it is possible to anaesthetise one hemisphere while the other remains active. If the left hemisphere is anaesthetised in this way the right side of the body normally becomes unresponsive whereas the left side remains active. The tester places an object in the left hand of the subject and asks her to remember it. After the effects of the drug wear off the subject is asked what was placed in her hand. The patient normally looks puzzled and denies that anything was placed in her hand. This is what we would expect since the left hemisphere where language is represented has not accessed the tactile and visual information present in the right hemisphere. But when the subject is shown a group of objects containing the held object she is instantly able to identify it (Patricia Churchland 1986 pp. 193-5).

The interesting point to notice here is that upon the return of an operative language processing hemisphere, the information in the possession of the right hemisphere is not subsequently accessed by the language system even when the subject consciously attempts to retrieve it. This suggests that there are distinct mechanisms controlling language function and visual and tactile functions. Access to previously stored information is restricted to the mechanism alone and not some other mechanism in the cognitive system. The language system can only get access to such information at the time of the information's arrival in the system and not later.

Where does this body of data leave the VFT? Well, I think that Michael Gazzaniga, a leading split brain researcher, has the correct intuitions, at least, regarding the evidence:

Clearly what is important is not so much where things are located, but that specific brain systems handle specific tasks. We begin to see that the brain has a modular nature, a point that comes out of all the data. It is of only secondary interest that the modules should always be in the same place. A correlate of this is that much of split-brain research should be viewed as a technique to expose modularity. That is, it is not important that the left brain does this or the right brain does that. But
It is highly interesting that by studying patients with their cerebral hemispheres separated, certain skills can be observed in isolation. It is a hugely significant point. (Gazzaniga 1985 pp. 58-9)

These words seem a fitting way to end this work. We can see that the original hope of finding interesting localisation of function at the interhemispheric level has given way to a basically cross-hemispheric localisation of particular functions. Cutting the cognitive pie in half is far too coarse a division of function if we are to explain the data of cognition. We also see the real importance of modularity in cognitive theorising; the importance of the neuropsychological data is not the emphasis on localising cognitive functions, but the ascertaining of what those basic functions might be. Split brain studies provide evidence for modularity generally. The degree of grain of domain upon which these modules operate is unclear though.
Bibliography


—(Forthcoming). “Perceptual Plasticity and Theoretical Neutrality: AReply to Jerry Fodor,” manuscript.


—(IR). “Information and Representation,” manuscript.


—(1988). "Functionalism and Broad Content," Mind XCVII.


Kitcher, Patricia (1980). "How to Reduce a Functional Psychology" Philosophy of Science 47.


—(OWITH). “On What’s In the Head,” manuscript.


