The Architecture of Belief

John Fitzpatrick

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

October 1989
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

\[(1) \text{ A } \phi \text{ that } p\]

where A is replaced by some agent, such as Reagan, \(\phi\) 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. The state of affairs might be actual or merely taken to be actual by the agent. 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.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 *ceteris paribus* clauses inserted in the relevant positions.4

Granted that intentional psychology in some sense works, just how it is *taken* to work and how it *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.

### 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.

#### 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 *true*, we will still be able to use intentional psychology in order to explain intelligent

---

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

4For 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.

\[\text{Footnote: For criticism of these putative shortcomings see Patricia Kitcher (1984). She argues that no current psychological, biological or neuroscientific accounts exist of the phenomena Churchland cites. If intentional psychology is grossly lacking in these areas then so are many other more supposedly respectable disciplines. If there an argument against the usefulness of a theory to be had in Churchland's examples then his beloved neurosciences will also be caught by its net.}\]
Chapter 1

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-
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 state-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.

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
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 monadidty 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 psychological 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
Chapter 1

The Topography of Intentionality

MP means that P, and

O has A iff O bears R 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)

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 any where in the computer's memory, but

---

9This 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 homuncularism 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 1 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 1 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 1. 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

\footnote{For 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).}

\footnote{For 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.12 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-put 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 ion the prinout 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— 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.
are 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 intentional 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, *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 *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.

---

19He 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.
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 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.

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

---

1A 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.
of the system. This is \textit{not} to say that there is some state which one could \textit{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.

\section*{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 \textit{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.

\footnote{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.\footnote{Dispositions are generally not thought to cause their manifestations. For a discussion see Prior (1985).} 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
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

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 process 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.\footnote{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.}

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
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

8 Well, 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 micro-analysis.

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

---

9Even 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 *kinds* 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 discipline 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 Enquist 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, 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 Eiffel 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

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

3That 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
Chapter 3

Autonomy 59

explanations generated by the primary level must range over the domain of the secondary level and capture the same level of generality as the explanations 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 commitment 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-Psychology 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. 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 Nagelian 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 relative 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, Parfit, and Prior (1982). The trouble with, though, 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 differences 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 parsimionous 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. 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.
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.8

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-Psychology 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.9

---

8I thank Karen Neander for suggesting this line of argument to me.
9Unfortunately, Enç (1980) seems to be disposed towards such arguments.
4 Confirmation Autonomy

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 confirmationaly 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.¹⁰

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 Wernicke’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.