1

Introduction: 'A picture held us captive'

1.1 TWO PROJECTS AND A PICTURE

Any attempt to understand how minds work must address, at the very least, two questions. The first is essentially an engineering question. One way of putting the question would be: *how can one build a mind*? This project is an engineering one. And, adopting a neologism first coined by Colin McGinn (1989), I shall refer to it as the project of *psychotectonics*: 'psycho', here, pertaining to minds, and 'tectonics' deriving from the Latin verb for building. Psychotectonics, then, is the science of building minds. In order to begin the project of psychotectonics, one must first have a reasonably adequate grasp of the things a mind can do, a grasp of the various functions of the mind. Then, it is thought, one must proceed to show, firstly, how these functions can be broken down into component sub-functions and these sub-functions broken down into sub–sub-functions, and so on, and, secondly, how these progressively more and more simple functions can be realized in progressively more and more simple mechanisms. To understand how to build a mind, it is claimed, is to be able to effect this sort of functional and mechanistic decomposition. This is a standard account of what is involved in psychotectonics, an account enshrined in David Marr's (1982) famous tripartite distinction between *computational*, *algorithmic*, and *physical* levels of analysis; whose basic idea is reflected in Dennett's (1978b) distinction between *intentional*, *design*, and *physical* stances; whose ethos is captured in the general project, also endorsed by Dennett (1978b) among others, of *homuncular functionalism*. This, then, is a very orthodox picture of how to do psychotectonics, and, while the picture might be a little worn in places, it is, I think, broadly correct. I do not, therefore, propose to challenge it, although I shall try to show that many of its proponents work with an unduly narrow conception of what a computational specification of the mind

should look like, and, consequently, that their conception of the algo-rithmic and physical realizations of this specification is unrealistic. The general idea of psychotectonics as functional and mechanistic decomposition, however, I shall accept, indeed presuppose. Psychotectonics, the project of building minds, occupies Part I of this book.

In order to understand minds, it is also necessary to understand how they can do what they can do. That is, it is necessary to understand how they come to possess those features considered essential to them. In recent discussions of the mind, two of its features loom large: *consciousness* and *intentionality*. Of consciousness, I shall have nothing to say. My suspicion, for what it is worth, is that the problem of con-sciousness is one that needs dissolution rather than a constructive solution. *If* dissolving the problem of consciousness requires dissolv-ing the picture of the mind upon which it is built, and *if* this book plays a role in dissolving that conception of the mind, then the argu-ments of this book might be considered relevant to the problem of consciousness. But that this is so is not something I shall assume. My concern, and the principal concern of Part II of this book, is with intentionality, with the aboutness or directedness of states of mind. It may ultimately turn out to be the case that one cannot understand intentionality independently of understanding consciousness, that the two are conceptually interlinked in such a way that forming an adequate conception of the one requires adequately conceptualizing the other (McGinn 1991). This may turn out to be the case, but I suspect not. And even if it does turn out to be so, there is no guarantee in advance of which will turn out to be conceptually prior. So, I shall assume that it is possible to understand intentionality, at least to some extent, without understanding consciousness. This assumption can be questioned, but it is by no means idiosyncratic. Indeed, the assumption is fairly standard.

Most recent discussions assume what Cummins (1989) calls a *repre-sentational theory of intentionality*; that is, they assume that the inten-tionality of mental states reduces to the representationality of mental representations. I, also, shall assume that this is the case, and, conse-quently, that the project of accounting for intentionality reduces to the project of accounting for how representations represent; how physical states can have semantic properties. This project I shall refer to as that of *psychosemantics*, employing another well-known neolo-

gism, but this time in a slightly different way from that intended by its author Jerry Fodor. Psychosemantics, as I shall understand it, is the project of accounting for the representationality of representations, for how representations represent, or possess semantic properties. Again, in common with most recent discussion, I shall assume that this is a reductionist project. Representation, or semanticity, is to be explained or accounted for in terms that are non-representational, non-semantic. Psychosemantics, in this sense, is the principal concern of Part II.

The projects of psychotectonics and psychosemantics are, of course, connected. The most straightforward connection is that the two accounts must cohere in that the functional/mechanistic decomposition of the mind yielded by the psychotectonic project, when combined with the reductionist, non-semantic account of representation identified by the psychosemantic project, must, together, be sufficient to add up to intentionality. Conversely, our account of representationality identified in the psychosemantic project must be consistent with what we know, or assume, to be the correct functional/mechanistic decomposition of the mind. What unites the two projects in this book, however, are not these fairly mundane connections but, rather, a conviction that both projects are hindered by a common conception of the mind. To use a phrase popularized by Wittgenstein, a picture of the mind holds us captive. And this picture prevents us from properly understanding what is required by the projects of psychotectonics and psychosemantics.

There is a view of the mind which began life as a controversial philosophical thesis and then evolved into common sense. The view is both widespread and tenacious, not only as an explicit doctrine but, more significantly, in the clandestine influence it has on explicit doctrines of the mind. The philosophical thesis from which the view is born is spelled out by Descartes, and its association with him is sufficiently robust for it to be called the *Cartesian conception*.

According to the Cartesian conception, minds are to be assimilated to the category of substance. That is, minds are objects which possess properties. Indeed, minds can, to some extent, be conceived of as relevantly similar to other bodily organs. Just as the heart circulates blood, the liver regulates metabolism, and the kidneys process waste products, the mind *thinks*. According to official Cartesian doctrine, the major difference between the mind and these other organs is that

3

The body in mind

the mind is a *non-physical* substance. The mind and brain are distinct entities, and, while the mind may receive input from the brain, and, in turn, send information back to the brain, the two are none the less distinct. The brain is a physical organ operating exclusively on mechanical principles; the mind is a non-physical organ operating according to principles of reason. And there is, Descartes thought, no prospect of deriving the former from the latter.

The Cartesian conception has been famously ridiculed as the myth of the *ghost in the machine*. And it has been Descartes' decision to make the mind ghostly (i.e., non-physical) that has drawn the principal fire from dissenters. The dissenters' case here has largely been successful, and not many philosophers or psychologists today would regard themselves as Cartesians in this sense. Ryle's expression, however, has another facet. Not only is Descartes' mind a ghost, but it is one that is *in* a machine. This was the principal source of Ryle's ire, of course. But, whereas the revolt against ghostly views of the mind has been overwhelmingly successful, criticism of the second aspect of Descartes' view has been comparatively muted. Most theorizing about the mind is now predicated on the assumption that the mind is physical; that is, that some sort of materialism is true. However, such theorizing has been, and largely still is, predicated on the view that the mind is an internal entity, i.e., located inside the skin of any organism that possesses it. The revolt against Cartesian views of the mind has been restricted to the first aspect of Descartes' view. The other aspect, Descartes' internalism, has, until recently, largely been ignored. Most forms of materialism are, thus, also forms of internalism.

Descartes' dualism and his internalism have, arguably, the same root: the rise of mechanism associated with the scientific revolution. This revolution reintroduced the classical concept of the atom in somewhat new attire as an essentially mathematical entity whose primary qualities could be precisely quantified as modes or aspects of Euclidean space. Macroscopic bodies were composed of atoms, and the generation and corruption of the former was explained in terms of the combination and recombination of the latter. Atomism is, then, mechanistic in the sense that it reduces all causal transactions to the translation, from point to point, of elementary particles, and regards the behaviour of any macroscopic body as explicable in terms of motions of the atoms that comprise it.

It is widely recognized that Descartes' dualism stems, at least in part,

4

Introduction: 'A picture held us captive'

from his acceptance of mechanism. The physical world, for Descartes, is governed by purely mechanical principles. He was, however, unable to conceive of how such principles could be extended to the thinking activities constitutive of the human mind. Minds, for Descartes, are essentially thinking things and, as such, governed by principles of reason. But such principles, Descartes thought, are distinct from, and not reducible to, principles of mechanical combination and association. Rationality, for Descartes, cannot be mechanized. Each mind is, thus, a small corner of a foreign field, inherently non-mechanical, hence inherently non-physical. Descartes' dualism, in this way, stemmed quite directly from his mechanism.

Of equal significance, however, is the connection between mechanism and internalism. Mechanistic atomism is, we might say, methodologically individualist. A composite body is ontologically reducible to its simple constituents. And the behaviour of a composite body is reducible to the local motions of its constituents. Thus, if we want to explain the behaviour of a macroscopic body, we need focus only on local occurrences undergone by its parts. This methodological individualism would also have some purchase on the explanation of the behaviour of human beings, since we are also, in part, physical. It is, therefore, no surprise that minds became analogously and derivatively conceived of by Descartes, and his dualist descendants, in atomistic terms. A mind, for Descartes, is essentially a *psychic monad* (Callicott 1989). Each mind is a discrete substance insulated within an alien material cladding. Just like any other atom, the mind could interact with the physical atoms of the body. But, crucially, and again just like any other atom, the essential nature of the mind was not informed by this interaction. The rational nature of the mind is taken as an independent given, and its interaction with other atoms is extrinsic to this nature. The ghosts of this conception of the mind, and the mechanistic and individualistic conception of explanation that underwrites it, are very much with us today.

These ghosts occupy a house with many mansions. Ontological theses are entangled with epistemological ones, each giving support and succour to the other. It is genuinely unclear if any particular thesis precedes any of the others. It is more realistic to suppose, perhaps, that ontological and epistemological aspects of Cartesianism grew up, indeed, evolved, together. A close relative of Descartes' ontological internalism is epistemic internalism, a view which has its modern

5

roots in Descartes. The central idea of epistemic internalism is that the difference between true belief and knowledge consists in some form of justification and, crucially, that justification consists in factors that are, in some sense, internal to the subject of the belief. The relevant notion of internality, however, is fundamentally epistemic. The activities of my heart, lungs, and liver are activities internal to me, but clearly these are not candidates for transformers of true belief into knowledge. Whatever transforms true belief into knowledge is, according to internalism, something of which the believing subject can be aware, something to which the subject has epistemic access. I can, however, be aware of many things, including whether or not it is presently raining, and the week's activity on the New York stock exchange. But this is not access of the relevant sort. Rather, epistemic internalism claims that justification consists in some sort of special access. According to Descartes, for example, the special access consists in the fact that the thinking subject can determine with certainty whether a belief has justification, and, consequently, whether a belief qualifies as knowledge. And, according to Chisholm (1966), a recent internalist descendant of Descartes, whether a belief has justification is something that can be determined by reflection alone. So, the central idea of epistemic internalism is that the factors that make a true belief justified, and, consequently, that transform a true belief into knowledge, are properties to which the believer has a special sort of epistemic access.

All these theses would be, if not straightforwardly undermined, then at least significantly threatened by the rejection of the Cartesian conception of the mind. If the mind is not self-contained in the relevant way, if the world enters into the very constitution of the mind so that the very identity of mental states involves something external, then it is not clear that this leaves any room for the special sort of epistemic access required by the internalist tradition. For the identity of the contents of one's mind would now, in part, consist of items to which one had no special epistemic access, and this threatens the epistemic relation one bears to the contents of one's mind as a whole. In this way, Descartes' ontological internalism is bound up with, supports and is supported by, his epistemic internalism.

Indeed, the content of internalism extends even beyond the ontological and epistemological spheres into the moral domain. This is because epistemic internalism is closely connected with the deonto-

Introduction: 'A picture held us captive'

logical notion of epistemic responsibility. The justification, or lack thereof, of my beliefs, is something for which I am responsible, something for which I can be praised or blamed. I may be victimized by a malevolent demon, I may be a brain in a vat, and so all or most of my beliefs may, in fact, be false. The truth of my beliefs depends on external factors and so is something beyond my control. But the justification of my beliefs depends only on internal factors, indeed, internal factors to which I have a special kind of access, and this is something that does lie within my control. Even though I may be hopelessly deceived about the truth of my beliefs, I can still do my epistemic duty with regard to their justification. Accordingly, the justification of my beliefs is something for which I can legitimately be praised or blamed; it is something for which I can be morally assessed. Thus, according to Descartes, epistemic justification is a form of deontological justification. If I do not have certainty but believe anyway, then I do not escape the blame of misusing my freedom. Ontological and epistemic forms of internalism, then, are also closely bound up with a certain view of the moral nature and responsibilities of human beings.

To speak of the Cartesian conception, therefore, is to speak of not just a single view of the mind, but of an array of interwoven views, each lending support to the others, and each being supported by the others. The strength of the Cartesian picture lies not merely in the strength of the individual theses – ontic, epistemic, ethical – that constitute it, but also, and perhaps even more importantly, in its scope. The Cartesian picture provides us with a sweeping and comprehensive vision of the nature of human beings. And the strength of the individual components of this vision derives, in an important sense, from the strength of the vision as a whole.

This book, in one clear sense, seeks to undermine the Cartesian picture of human beings. However, its primary concern is not with the picture as such, but with the influence it has exerted on subsequent theorizing about the mind. Moreover, the principal focus of the book will be restricted to a sub-set of mental phenomena that have proved particularly central to twentieth-century concerns. These phenomena comprise what are known as cognitive processes: processes such as perceiving, remembering, and reasoning whereby an organism gains and uses information about its environment. The Cartesian picture has bequeathed us a conception of such processes

The body in mind

whose essence can be distilled into the following two principles. One is an ontological claim about the nature of cognitive processes, the other an epistemological corollary about how these processes are best studied or understood.

> *The Ontological Claim*: Mental states and processes are located exclusively inside the skin of cognizing organisms.

> *The Epistemological Claim*: It is possible to understand the nature of mental states and processes by focusing exclusively on what is occurring inside the skin of cognizing organisms.

These two assumptions constitute that particular version of the internalist picture of the mind with which this book is concerned. Cognitive processes are *essentially* internal items. They may stand in various relations to events, states, and processes occurring outside the skin of cognizers, and these external items may play an important, even essential, role in the facilitation or satisfaction of the internal processes themselves. Nevertheless, cognitive processes are, in essence, internal items. This internalist picture of cognition, I shall try to show, has greatly distorted our conception of what is required of us by the projects of psychotectonics and psychosemantics. And the principal task of this book is to unseat this conception of cognition, and outline, in broad strokes, the ramifications of this for the two projects.

1.2 PICTURE AS MYTHOLOGY

The task of this book is to unseat not a particular philosophical or psychological theory, but a certain pre-theoretical conception or picture of the mind. This picture is prior to theory in that it is what guides theory construction and thus lends coherence and unity to the experimental practices and procedures judged relevant to the conformation or falsification of particular theories within its domain. The notion of a pre-theoretical picture, then, corresponds largely to what Kuhn (1970) has, famously, labelled a *paradigm*.

In trying to unseat a pre-theoretical picture of a certain domain of inquiry, one's options are usually fairly restricted. One might try to attack the picture on grounds of internal incoherence. Many pictures, perhaps most, however, are not internally incoherent, and, certainly, I

8

Introduction: 'A picture held us captive'

would not want to suggest that the internalist picture of the mind suffers from this sort of defect. One might try to attack the picture on grounds of empirical inadequacy. As Wittgenstein, Kuhn, and others have pointed out, however, this sort of approach is unlikely to work. The problem is not so much that a pre-theoretical picture is at a further remove from the evidence, hence at a further remove from the possibility of experimental confirmation or falsification, than the theories predicated upon it (although this certainly might be a problem). The problem, rather, is that the very descriptions of the evidence employed for the purposes of experimental testing are based on, and thus presuppose, this pre-theoretical picture. It is very difficult to see how a pre-theoretical picture of a domain could be unseated solely on the basis of evidence whose very identification and conceptualization as evidence, presupposes the picture itself.

Kuhn, as is well known, sees the unseating of a pre-theoretical picture or paradigm as a matter of piecemeal accretions of problems unsolved within the framework of the picture gradually inducing a *crisis* within the picture itself. Such a crisis is likely to come to a head only when there is a competitor to the picture, an alternative paradigm that can play the same sort of role as the original. As Kuhn points out, however, the new picture, initially, is likely to be partial, restricted, and susceptible to all sorts of apparent refutations. This is because the evidence thought relevant to its truth or falsity is still conceptualized in terms of the old paradigm. It is only when the new paradigm becomes accepted that the relevant evidence can gradually be reconceptualized, and then the scope and coherence of the new paradigm can be progressively enhanced. Arguably, such a situation may today be occurring in cognitive science. Arguably, it may not. And, in any event, I would not want to predicate any argument upon such an essentially contestable claim.

The inspiration for the method I propose to adopt in this book derives from Wittgenstein rather than Kuhn. A pre-theoretical picture in the sense described above has the status of what Wittgenstein calls a *mythology*. To call something a mythology in this sense is not to cast, at least not directly, any aspersions at its truth or validity, although it may mean that, for it, questions of truth and validity do not arise. What Wittgenstein is getting at when he calls something a mythology is that it plays a certain role in organizing experience. More precisely, a mythology has the role of legislating

that *this is how things must be*. One of the principal tasks of Wittgenstein's later philosophy is to show how certain pre-theoretical conceptions of the mind and its contents are mythologies in precisely this sense. And one of the principal methods employed by Wittgenstein in this context is to undermine a mythology by showing that we do not, in fact, have to think of things in the way the mythology tells us we do. Other ways of thinking about the mind and its contents, for example, are possible.

A good example of Wittgenstein's idea of mythology is to be found in his attitude toward Freud's concept of the unconscious. In his *Cambridge Lectures 1932–35*, Wittgenstein writes:

> What Freud says about the subconscious sounds like science, but in fact is just a *means of representation*. New regions of the soul have not been discovered, as his writings suggest. The display of elements of a dream, for example, a hat (which may mean practically anything) is a display of similes. As in aesthetics, things are placed side by side so as to exhibit certain features. (1979:40)

Wittgenstein is quite willing to allow that Freud has discovered certain psychological reactions of a hitherto unknown sort, but the apparatus he invokes to explain these is not a theory but simply a *means of representation*. That is, he has simply imposed, as Wittgenstein would say, a system of notation which allows him to redescribe these psychological reactions in these terms. Psychoanalysis, while presenting itself as an experimental discipline, does not, in fact, satisfy any of the conditions necessary to a discipline of this kind.

The comparison with aesthetics is indicative of Wittgenstein's attitude towards the unconscious. Consider the difference between the role of analogy ('simile') in fields like aesthetics and its role in the empirical sciences. An analogy of the first type might consist, for example, in comparing architecture with a language and, then, attempting to identify the vocabulary and grammar of this language. This type of analogy, however, does not generate hypotheses that can be tested in experiments, nor does it produce a theory that can be used to predict events. Thus, whatever understanding is occasioned by the use of such analogies is not the result of imparting new information, nor does it lead to new empirical discoveries. Furthermore, such understanding does not lead to the asking of fresh questions that can be answered by further empirical research. The analogy, rather, func-

10