1

Cartesianism, behaviourism, and the philosophical context

> I had even intended to ask your attention a little while on trust, and (as sometimes one contrives, in taking a friend to see a favourite piece of scenery) to hide what I wanted most to show with such imperfect cunning as I might, until we had unexpectedly reached the best point of view by winding paths. But . . . since I have heard it said by men practised in public address, that hearers are never so much fatigued as by the endeavour to follow a speaker who gives them no clue to his purposes – I will take off the slight mask at once.
>
> (John Ruskin, 1865)

1.1 AFTER BEHAVIOURISM

In the sixth of the *Meditations on First Philosophy*, which he published in 1641, Descartes expresses the core of the dominant philosophy of mind of the last three centuries:

> from the mere fact that I know for certain that I exist and that I cannot see anything else that belongs necessarily to my nature or essence, except that I am a thinking thing, I rightly conclude that my essence consists in this alone: that I am a thinking thing, a substance whose whole nature or essence is to think . . . it is certain that this *I*, that is to say my soul, which makes me what I am, is entirely and truly distinct from my body, and can be or exist without it. (1968: 156, translation slightly altered)

For a Cartesian, therefore, the mind is the private domain of a single consciousness, and it is possible, at least in principle, that there should be disembodied minds, unable, however hard they tried, to become aware of each other. Descartes knew, of course, that the way we do in fact come to know what is happening in other minds is by way of observing the speech and actions of 'other bodies'. But for Descartes it was always a serious conceptual possibility[1] that the evidence we normally take as adequate for supposing that other bodies are inhabited by minds should be produced by automata. Minds and bodies are quite distinct sorts of things – 'substances' – whose causal relations are obscure; and there is a serious epistemological worry about how one mind should know anything about another.

I have said this was the dominant view: and so it was. So dominant, in fact,

1. Ruled out, in fact, only by the guarantee of a God, who is 'no deceiver'.

that by the post-war era the central problems of the philosophy of mind were reduced, in effect to two. First, the 'problem of other minds':[2] 'What justifies our belief that other minds exist at all?' And secondly, 'the mind–body question':[3] 'How are we to explain the relations of a mind and its body?'

The very evident fact that we *do* know that other people have minds has lead some philosophers and psychologists in our own century to behaviourism. Faced with these problems for the Cartesian view of mind as different in substance from the body, they have identified the mind with certain bodily dispositions. In particular, they have sought to characterise belief, which for Descartes (and for the British empiricists who followed him) was a paradigmatically private matter, in terms of dispositions to produce and respond to language. On such a view, which we can find, for example, in Quine, believing that, say, grass is green, is simply being disposed, in certain circumstances, to assent to the sentence 'Grass is green.'[4] Since bodily dispositions seem to be less epistemologically puzzling than the states of a mental substance, this solution to the second question also solves the first.

This behaviourist view may solve some problems, but it leaves others. It makes it impossible, for example, to give a straightforward account of the beliefs of non-speaking creatures (including infants), and has led some philosophers to deny that such creatures can have beliefs at all.[5] The behaviourist also has to deny what is, I think, the natural view of language. In a Cartesian framework, of the sort adopted by Descartes' contemporaries Arnauld and Hobbes,[6] language can be seen as simply the expression of our thought; or, as Hobbes puts it, with characteristic directness:

> Words so connected as that they become signs of our thoughts, are called SPEECH, of which every part is a name. (Hobbes, 1839)

The behaviourist objection to this account is rooted in a scepticism as to the existence of the private states – Hobbes' 'thoughts' – which Cartesianism regards as the one sort of thing 'that I know for certain.' Blaming the defects of the Cartesian view on its commitment to the existence of private mental states, they have placed their confidence in the certain existence of the public sphere of utterance.

2. See, for example, John Wisdom's well-known *Other Minds* (1952).
3. See, for example, H. Feigl's *The Mental and the Physical* (1967).
4. Quine's reflective view is more subtle than this; but there are passages where he commits himself to what looks like just this position; see (Quine, 1960: 217; 1975). This view, which seems to me one that common sense should find faintly bizarre, is nevertheless, held also by, for example, Donald Davidson and Michael Dummett; see Davidson's 'Thought and Talk' (1975), and Drummett's *Frege, The Philosophy of Language* (1973), *passim*. This unanimity is especially striking as they agree about almost nothing else.
5. Davidson, Dummett, Quine, see previous footnote.
6. See Hacking (1975*a*), which influences this whole (brisk!) account.

And, in my view, it has been a significant part of the appeal which language has had for many recent philosophers, as an object of philosophical study, that it is public. Spoken and written language, unlike the intentions and beliefs of its speakers, is open to the inspection of all. In short, a critique of the Cartesian view of the mind as a pure consciousness, privately available to itself, led many philosophers in the analytic tradition to embrace language as something empirically accessible, and to reject an account of language, therefore, as the expression of interior states.

Now though there is something rather unsatisfactory about the privacy of the Cartesian mind, there is something absurd about the publicity of the behaviourist one. 'Hello; you're fine. How am I?' says the behaviourist in the cartoon; and the cartoonist has a point. What was needed was somewhere in a gap which most philosophers could not discover between the Cartesian Scylla and the behaviourist Charybdis: and this book is an attempt to see what we can do in the philosophy of language, now that what we call 'functionalism' has shown us how to chart a course between them.

1.2 THE BACKGROUND

Functionalism, therefore, is the key thought behind the account I offer in this book of language and its relation to the mind and to reality; and a large portion of Part I of this book is devoted to explaining and defending a functionalist account of the mental. But I can best explain the detailed structure of the book by saying how it came to be written. The story I shall tell is, like most accounts of the origin of an intellectual project, a kind of friction. But it is, I hope, an explanatory fiction, in the sense that it helps to make intelligible the project whose origins it purports to describe.

I had been concerned, when I started research for my doctoral dissertation, with what then (in 1975) seemed to me the most important question in philosophical semantics: the issue between realism and antirealism, which Dummett had made central to his discussion of Frege; see Dummett (1973). Dummett's view was that semantic realism, the thesis that the meaning of declarative sentences was given by their truth conditions, was irremediably flawed; and that we should do better if we tried to explain meaning in terms of conditions which provided epistemic warrant for assertion; in terms, in other words, of what have been called 'assertibility conditions'. My initial feeling was that Dummett's arguments against realist semantics were challenging, perhaps even unanswerable; but that their conclusion – that our grasp of the meaning of most sentences could not issue from knowledge of their truth conditions – was very hard to accept. If Dummett was wrong in his arguments against realism, there was no compulsion to do assertibility condition

semantics. I felt he *was* wrong. But in philosophy we are supposed to follow reason, even where it conflicts with our hunches; so I thought it was necessary to examine the question: 'What would a semantics in terms of assertibility conditions look like?'

In casting about for ways to approach this question, I came across some work on the logic of conditionals which offered a promising starting place. For, just in this case, it seemed, there was overwhelming evidence that a truth-conditional account – which Dummett claimed was the essence of realism – could not be provided. This, I thought, was the message of David Lewis' triviality proofs, which were finally published in Lewis (1976). Ernest Adams had suggested, in two papers (1965, 1966) and, finally, in his book *The Logic of Conditionals* (1975), that the indicative conditional's semantics should be given not by way of truth conditions, but by way of a rule of assertibility. That rule was that conditionals were assertible when the conditional probability of the consequent, given the antecedent, was high. I call this Adams' Hypothesis.

In examining this view I came to two conclusions; first, that if Adams was right, there was a good sense in which indicative conditionals do not have truth conditions; and secondly, that the relevant notion of assertibility was one that needed further examination.

There was an obvious place to look for an account, both of the sense in which conditionals do not have truth conditions and of the notion of assertibility. And that was in an examination of the nature of the subjective probabilities in terms of which the assertibility rule was couched: and of their relation to the speech-act of assertion. So I began the work which takes up the first two parts of this book.

I set out, then, to explain subjective probabilities in a way which could plausibly ground the kind of assertibility rule that Adams had proposed. In doing so I became dissatisfied with the standard accounts of the status of subjective probabilities. The source of my dissatisfaction was the simple fact that nothing adequate had been said about the way subjective probability theory could form part of a descriptive account of agents and the sources of their behaviour. Plainly, people's degrees of belief are not coherent in the way that standard decision theory requires, and the question why that was seemed to lack an answer I could believe.

In finding the answer I was lucky in two things: first, I had read some time in the mid-1970s Ian Hacking's paper on 'slightly more realistic personal probability' (1967*b*); second, while spending a year at work in the Yale Graduate School, I had attended a seminar of Jerry Fodor's on the philosophy of psychology. What Fodor provided me with was the notion of computational structure, which I shall use to explain the relation between decision

theory and the actual behaviour of agents. What Hacking gave me was the only paper I know of in which what is, in essence, a computational approach had been applied to subjective probability. Classical decision theory cannot explain how people can come to give assent to sentences which ought, because they are logically impossible, to have zero probability in every probability function; or how they can fail to give assent to sentences which follow from sentences they believe. Given the theory of computational structure, which is outlined in Chapter 4, this can now be explained.

In coming to see how to deal with these questions I became more convinced that decision theory, in its descriptive guise, is best seen as part of the functionalist theory of representational states.

What I also came to see was that the notion of assertibility was not needed only for structures like the conditional which do not determine truth conditions; rather it was a central explanatory concept in the philosophy of language. For truth-conditional sentences, the account of assertibility is rather straightforward: what is assertible is what you believe to be very probably true. But though it is straightforward it is not unimportant. For without this much theory, we cannot connect work in logic with the actual linguistic behaviour of speakers. It is this connection between probability and assertion that is at the core of what has come to be called 'probabilistic semantics'. And this book is an attempt to set out (one way of looking at) the foundations of this new, and, in my view, important departure in semantic theory. I say 'one way' because the particular way in which I develop probabilistic semantics within a functionalist framework is not one that will appeal to all philosophical tastes: however, I think it is at present the best way.

Though assertibility now seems to me a crucial notion, the issue of antirealism has faded somewhat into the background; I now think it is a diversion. For, for the large class of truth-conditional sentences, the assertibility rule makes explicit reference to the truth conditions. We do not need to chose between truth and assertibility: we have to have them both. For this and other reasons, which are set out in my forthcoming book *For Truth In Semantics*, I think that Dummett's global antirealism is ill advised: but I do argue in this book for a modest local antirealism about indicative conditionals.

1.3 THE STRUCTURE OF THE BOOK

This book is intended, therefore, as an introduction to this new approach to semantic theorising, whose outlines I have discerned in the recent writings of a number of philosophers. What is new in it is not just the technical apparatus of probability theory; for it also offers the possibility of reconciling the view that language is the expression of interior states, on the one hand, with the

legitimate rejection both of the Cartesian view of the inner and of the behaviourist over-reaction to Cartesianism, on the other. Much of the argument in the literature in this area is extremely technically demanding; and, like most new work, its philosophical underpinnings are often glossed over in the urgent desire to communicate the latest insight. In consequence many assumptions are not spelled out in sufficient detail to be accessible to the generally interested reader; indeed, so far as I know, some of these assumptions have never been examined in a detailed way before the work of my own dissertation. They require detailed philosophical exploration and defence.

So Part I begins with a discussion of the nature of beliefs. I have already sketched the central features of the Cartesian tradition – 'our' tradition, one is bound to say – and criticised its notion of privacy; and that critique is a crucial element in the motivation for the first chapter, where I argue for functionalism in the philosophy of mind. The new few chapters elaborate a picture of beliefs as a kind of functional representation, in a theory with two interlocking parts. Chapter 3 discusses the first part, which focusses on the role of beliefs, in concert with desires, in producing action; Chapter 4, the second part, which looks to the role of belief in thought. It is argued that we can capture this latter role by way of a notion of *computational structure*, which is a way of saying how beliefs interact with each other, and with desires, to produce other desires and beliefs. 'Computational' because I think that there is a good analogy here with the way the information-bearing states of computers interact; 'structure', because the patterns of interaction depend on features analogous to the syntactic structure of strings in a natural or artificial language. Chapter 5 shows how, once the central role of belief in thought, prior to action, is understood, truth conditions can be assigned to beliefs along with their computational structures. And truth conditions and computational structure are major features, I claim, of beliefs (and, as it happens, of desires). Each of these kinds of state has other features also: notably degrees or strengths, which, for beliefs are called *subjective probabilities* and for desires are called *desirabilities*.

Part II, which begins at Chapter 6, is about the theory of meaning. I think that the core of the theory of meaning can be given by the simple thought that assertion expresses belief. But Chapter 6 looks first at a way of thinking about meaning in terms of truth not in terms of the expression of inner states, but in terms of the outer states sentences set out to describe. That theory is that the meanings of assertions are their truth conditions – conditions which hold when and only when they are true; and I discuss briefly Donald Davidson's way of setting such a theory up. My interest here is not in Davidson's theory in itself, but in two things that it allows us to see clearly: first, how we can use the structure of a truth-theory to set up a picture of the

relations between the truth conditions of sentences; second, how important is the subjunctive dependency of a sentence's meaning on the way the world is. The first of these things is successfully captured in the structure of Davidson's theory; the second is not. But both Davidson's successes and his failures are instructive.

These lessons learnt, I can say, in Chapter 7, how I think a theory of meaning should work. Since I take it that sentences express beliefs, the job is to say for each sentence what belief it expresses in a given context; and since Part I gives a canonical representation for beliefs, I claim that the right mapping is from sentences into those canonical representations. It will emerge that Davidson is half right. Sentences do have truth conditions and they partly fix their meaning. But the reason they have truth conditions is that beliefs do; and truth conditions won't do on their own to individuate beliefs.

Part III then takes up the problem I began with a decade ago: how do we fix the meanings of indicative conditionals? And I try to show how the theory of meaning of Part II allows us to solve this old question. I think the semantics of conditionals is an intrinsically interesting question; but you do not have to believe that to read Part III. For the fact that we can solve some previously intractable problems about conditionals is part of the evidence for the view that the theory of meaning is right. And if you are not interested in the theory of meaning – in the relation of language, mind and world – you should not be reading a philosophy book.

1.4 NEGLECTED TOPICS

There are two conspicuous absences in this book, absences which will worry those familiar with recent work in the philosophy of language. First, I have not discussed the general issue of psychologism in semantics; which Dummett has on one occasion defined as the 'invasion of the theory of meaning by notions concerned with mental processes' (1975: 240). What may appear worse is that I have actually assumed the opposite of what is now widely believed, since my position is unabashedly psychologistic. I had originally written a chapter in defence of my psychologism. But it seems to me, in the end, that most of the arguments against the kind of psychologism that I espouse were just bad arguments; and that the best defence of my position was just to show that I could proceed on psychologistic assumptions without getting into trouble. My psychologism amounts to this: I hold that we can give an account of the contents of the beliefs of agents, independent of any account of their languages, and then go on to explain linguistic acts as the expression of those beliefs. Dummett (1973, *passim*) and Davidson (1975) both

deny this. I hope that as my argument proceeds my position will at least remain plausible. I have argued against Dummett in the last part of *For Truth In Semantics* (forthcoming); Jonathan Bennett has ably pursued the same conclusion in his *Linguistic Behaviour* (1976); and Jerry Fodor has also done so, in a different way, in *The Language Of Thought* (1976). Between them, I think these defences alone – and there are, of course, others – make the absence of an elaborate discussion of psychologism excusable. All the work of the first five chapters is meant to apply to agents whether or not they have languages: and granted that the account works, I see no reason to accept Dummett's position.

What is also conspicuously absent is any detailed discussion of possible world approaches to meaning and to the conditional. I have one general and one particular reason for avoiding possible worlds. The general reason is this: my account of the mind is functionalist and thus fundamentally causal. Possible worlds do not, of course, enter into causal relations with each other, and so, in particular, there are no causal interactions 'across' worlds. The question of how possible worlds relate to causation is not yet at all clear. Until an account of the causal facts entailed by an ascription to a belief of a class of worlds in which it is true is available, it is difficult to see how possible worlds can be of use to functionalism. When it is available, I conjecture, it will be clear that, *pace* David Lewis (1973), they play no essential explanatory role. That accounts for my general suspicion of possible world accounts.

But my specific objection is, in the context of what I am doing, quite decisive. Possible world semantics assigns to sentences truth conditions: it tells us which sentences are true in the actual world in virtue of truths about other worlds. But I believe, and I argue, that the conditionals I am concerned with do not have truth conditions; and, if I am right, there cannot therefore be a possible world semantics for them.

This argument is, I have no doubt, too brisk to satisfy everyone. There are non-realist ways of construing possible world semantics: seeing possible worlds as maximal consistent sets of sentences, for example, is one, provided the relevant notion of consistency is not defined in terms of truth. But such interpretations require a restructuring of the way possible world semantics is usually done, certainly for conditionals: and I do not feel obliged to do that restructuring. If we can do conditional semantics without possible worlds, I see nothing to be gained by doing it with them.

Part I *Belief*

2

A theory of the mind

> The concept of a mental state is primarily the concept of a state of the person apt for bringing about a certain sort of behaviour.
>
> (Armstrong, 1968: 82)

2.1 OVERVIEW

The central claim of this book is that the meaning of an asserted sentence is, in a certain sense, the content of the belief that it expresses. So I must begin with a story about beliefs and their contents. In fact, as I have said, my story is functionalism, the story encapsulated in the remark of Armstrong's which provides my epigraph. I need to give the outlines of functionalism's picture of the mental because the general account of belief and of the contents of beliefs is needed if I am to make good my central claim. Since my interest is mainly in assertion, and in beliefs as what assertions express, I concentrate mainly on giving an account of the mental state of believing; but because, as I shall argue, that account can only be given in the context of a theory that also takes account of desires, I am obliged also to say something about desire.

The theory I hold is functionalist: for it claims that we can state sufficient *a priori* truths about the causal roles of mental states in interaction with each other and with events outside the mind, to be able to individuate those mental states by their causal roles. For the states of belief and desire the relevant events outside the mind fall into two classes: those events which lead to changes in the sensory and perceptual states of agents, on the one hand; and those events which are caused in the agent's actions, on the other. The next section elaborates some of the features of functionalism, but the general idea is easily grasped, by way of a simpler case. Take, then, by way of example, an account of the workings of a thermostat.

Consider some particular thermostat. We can give a complete physical account of its workings, by saying what its parts are made of, and what physical laws govern those parts. It will follow from such an account that if a certain part of the thermostat – the heat-sensor – is heated, the result will be to open a switch; and that if it is cooled, the result will be to close that switch. It will also follow that the opening of the switch turns off a heater; and that the closing of the switch turns it on. Thus, from the full physical