2

Minds, bodies and people

A perennial issue in the philosophy of mind has been the so-called *mind–body problem*: the problem of how the mind is related to the body. However, as I indicated in the previous chapter, this way of putting the problem is contentious, since it suggests that 'the mind' is some sort of *thing* which is somehow related to the body or some part of the body, such as the brain. We are invited to consider, thus, whether the mind is *identical* with the brain, say, or merely causally related to it. Neither proposal seems very attractive – the reason being, I suggest, that there is really no such thing as 'the mind'. Rather, there are *minded* beings – subjects of experience – which feel, perceive, think and perform intentional actions. Such beings include human persons, such as ourselves, who have bodies possessing various physical characteristics, such as height, weight and shape. The mind–body problem, properly understood, is the problem of how subjects of experience are related to their physical bodies.

Several possibilities suggest themselves. In describing them, I shall restrict myself to the case of *human persons*, while recognising that the class of subjects of experience may be wider than this (because, for instance, it may include certain non-human animals). One possibility is that a person just *is* – that is, is identical with – his or her body, or some distinguished part of it, such as its brain. Another is that a person is something altogether distinct from his or her body. Yet another is that a person is a composite entity, one part of which is his or her body and another part of which is something else, such as an immaterial spirit or soul. The latter

two views are traditionally called forms of 'substance dualism'. A 'substance', in this context, is to be understood, quite simply, as any sort of persisting *object* or *thing* which is capable of undergoing changes in its properties over time. It is important not to confuse 'substance' in this sense with 'substance' understood as denoting some kind of *stuff*, such as water or iron. We shall begin this chapter by looking at some arguments for substance dualism.

CARTESIAN DUALISM

Perhaps the best-known substance dualist, historically, was René Descartes – though it is not entirely clear which of the two forms of substance dualism mentioned above he adhered to.[1] Often he writes as if he thinks that a human person, such as you or I, is something altogether distinct from that person's body – indeed, something altogether non-physical, lacking all physical characteristics whatever. On this interpretation, a human person is an immaterial substance – a spirit or soul – which stands in some special relation to a certain physical body, *its* body. But at other times he speaks more as if he thinks that a human person is some sort of combination of an immaterial soul and a physical body, which stand to one another in a rather mysterious relation of 'substantial union'. I shall set aside this second interpretation, interesting though it is, largely because when philosophers today talk about 'Cartesian dualism' they usually mean the former view, according to which a person is a wholly immaterial substance

[1] Descartes's views about the relationship between self and body receive their best-known formulation in his *Meditations on First Philosophy* (1641), to be found in *The Philosophical Writings of Descartes*, ed. J. Cottingham, R. Stoothoof and D. Murdoch (Cambridge: Cambridge University Press, 1984). In recent times, one of Descartes's best-known and severest critics has been Gilbert Ryle: see his *The Concept of Mind* (London: Hutchinson, 1949), ch. 1. For a controversial critique of the received view that Descartes was a 'Cartesian dualist', see Gordon Baker and Katherine J. Morris, *Descartes' Dualism* (London: Routledge, 1996). It is unfortunate that many modern philosophers of mind tend to distort or oversimplify the historical Descartes's views, but this is not the place for me to engage with them over that issue.

possessing mental but no physical characteristics. But it is important, when considering this view, not to confuse the term 'substance' in the sense in which we have just been using it with the sense in which it denotes a kind of stuff. Cartesian dualism does not maintain that a person is, or is made of, some sort of ghostly, immaterial stuff, such as the 'ectoplasm' beloved of nineteenth-century spiritualists. On the contrary, it maintains that a person, or self, is an altogether simple, indivisible thing which is not 'made' of anything at all and has no parts. It contends that you and I are such simple things and that we, rather than our bodies or brains, are subjects of experience – that is, that we rather than our bodies or brains have thoughts and feelings. In fact, it contends that we and our bodies are utterly unlike one another in respect of the sorts of properties that we possess. Our bodies have spatial extension, mass, and a location in physical space, whereas we have none of these. On the other hand, we have thoughts and feelings – states of consciousness – whereas our bodies and brains lack these altogether.

What reasons did Descartes have for holding this seemingly strange view of ourselves – and how good were his reasons? He had several. For one thing, he considered that our bodies were simply incapable of engaging in intelligent activity on their own account – incapable of *thinking*. This is because he believed that the behaviour of bodies, left to themselves, was entirely governed by mechanical laws, determining their movements as the effects of the movements of other bodies coming into contact with them. And he couldn't see how mechanically determined behaviour of this sort could be the basis of such manifestly intelligent activity as the human use of speech to communicate thoughts from one person to another. With the benefit of hindsight, we who live the age of the electronic computer may find this consideration less than compelling, because we are familiar with the possibility of machines behaving in an apparently intelligent fashion and even using language in a way which seems to resemble our own use of it. Whether it is right to think of

computers as really being capable of intelligent behaviour on their own account, or merely as cleverly constructed devices which can *simulate* or *model* intelligent behaviour, is an open question, to which we shall return in chapter 8. But, certainly, there is no simple and obvious argument from our own capacity for intelligent behaviour to the conclusion that we are not to be identified with our bodies or brains.

THE CONCEIVABILITY ARGUMENT

The argument that we have just considered and found wanting is an empirical argument, at least to the extent that it appeals in part to the laws supposedly governing the behaviour of bodies. (Descartes himself thought that those laws had an *a priori* basis, but in this he was almost certainly mistaken.) However, Descartes also had, more importantly, certain *a priori* arguments for his belief that there is, as he puts it, a 'real distinction' between oneself and one's body. One of these is that he claims that he can 'clearly and distinctly perceive' – that is, coherently conceive – the possibility of himself existing without a body of any kind, that is, in a completely disembodied state. Now, if it is *possible* for me to exist without any body, it seems to follow that I cannot be *identical* with any body. For suppose that I were identical with a certain body, *B*. Given that it is possible for me to exist without any body, it seems to follow that it is possible for me to exist without *B* existing. But, clearly, it is *not* possible for me to exist without *me* existing. Consequently, it seems that I cannot, after all, be identical with *B*, because what is true of *B*, namely, that I could exist without it existing, is not true of me.

However, the force of this argument (even accepting its validity, which might be questioned) depends upon the cogency of its premise: that it is indeed possible for me to exist without any body.[2] In support of this premise, Descartes

[2] One possible reason for questioning the argument is that it assumes that it is an *essential* property of any body, *B*, that it is a body, that is, that *B* would not have existed if it had not been a body. I myself find this assumption plausible, but it

claims that he can at least conceive of himself existing in a disembodied state. And, to be fair, this seems quite plausible. After all, many people report having had so-called 'out of body' experiences, in which they seem to float away from their bodies and hover above them, seeing them from an external point view in the way in which another person might do so. These experiences may not be veridical: in all probability, they are hallucinatory experiences brought on by stress or anxiety. But they do at least indicate that we can *imagine* existing in a disembodied state. However, the fact that we can *imagine* some state of affairs is not enough to demonstrate that that state of affairs is even logically possible. Many of us find little difficulty in imagining travelling back in time and participating in historical events, even to the extent of changing what happened in the past. But on closer examination we see that it is logically impossible to change the past, that is, to bring it about that what has happened has not happened. So too, then, we cannot conclude that it really is possible to exist without a body from the fact that one can imagine doing so.

Of course, Descartes doesn't claim merely that he can *imagine* existing without a body: he claims that he can 'clearly and distinctly perceive' that this is possible. But then, it seems, his claim simply amounts to an assertion that it really *is* possible for him to exist without a body and doesn't provide any independent *grounds* for this assertion. On the other hand, is it fair always to insist that a claim that something is possible must be susceptible of proof in order to be rationally acceptable? After all, any such proof will have to make appeal, at some stage, to a further claim that something or other is possible. So, unless *some* claims about what is possible are acceptable without proof, no such claims will be acceptable at all, which would seem to be absurd. Even so, it may be felt that Descartes's particular claim, that it is possible for him to exist without a body, is *not* one of those possibility

has been challenged by Trenton Merricks: see his 'A New Objection to A Priori Arguments for Dualism', *American Philosophical Quarterly* 31 (1994), pp. 80–5.

claims which is acceptable without proof. The upshot is that this argument of Descartes's for the 'real distinction' between himself and his body, even though it could conceivably be sound, lacks persuasive force: it is not the sort of argument that could convert a non-dualist to dualism.

THE DIVISIBILITY ARGUMENT

Descartes has another important argument for the 'real distinction' between himself and his body. This is that he, as a subject of experience, is a simple and indivisible substance, whereas his body, being spatially extended, is divisible and composed of different parts. Differing in these ways, he and his body certainly cannot be one and the same thing. But again, the crucial premise of this argument – that he is a simple and indivisible substance – is open to challenge. Why should Descartes suppose this to be true? There are two ways in which his claim might be attacked, one more radical than the other. The more radical way is to challenge Descartes's assumption that he is a *substance* at all, whether or not a simple one. By a 'substance', in this context, recall that we mean a persisting object or thing which can undergo changes in its properties over time while remaining one and the same thing. To challenge Descartes's assumption that he is a substance, then, is to question whether, when Descartes uses the first-person pronoun, 'I', he succeeds in referring to some single thing which persists identically through time – indeed, more radically still, it is to question whether he succeeds in referring to some *thing* at all. Perhaps, after all, 'I' is not a referring expression but has some other linguistic function.[3] Perhaps the 'I' in 'I think' no more serves to pick out a certain object than does the 'it' in 'It is raining'. Although some philosophers have maintained precisely this, it seems an

[3] For an example of a philosopher who holds that 'I' is not a referring expression at all, see G. E. M. Anscombe, 'The First Person', in S. Guttenplan (ed.), *Mind and Language* (Oxford: Clarendon Press, 1975), reprinted in G. E. M. Anscombe, *Metaphysics and the Philosophy of Mind: Collected Philosophical Papers, Volume II* (Oxford: Blackwell, 1981). I discuss this view more fully in chapter 10.

implausible suggestion. It seems reasonable to suppose that what I have been calling 'subjects of experience', including human persons, do indeed exist and that the first-person pronoun is a linguistic device whose function it is to refer to the subject who is using it. And it also seems reasonable to suppose that subjects of experience persist through time and undergo change without loss of identity. Anyway, I shall assume for present purposes that this is so, though we shall return to the issue when we come to discuss personal identity in chapter 10. In short, I shall consider no further, here, the more radical of the two ways in which Descartes's claim that he is a simple substance might be challenged.

The other way in which this claim might be challenged is to accept that Descartes, and every subject of experience, is a 'substance', in the sense of the term that we have adopted, but to question whether he is a *simple* and *indivisible* substance. Why should Descartes have supposed that he himself is simple and indivisible? After all, if he were to lose an arm or a leg, would he not have lost a part of himself? Descartes's answer, no doubt, is that this would only be to lose a part of his *body*, not a part of *himself*. But this presupposes that he is not identical with his body, which is the very point now in question. What is required is an independent reason to suppose that Descartes's loss of his arm or leg is no loss of a part of *himself*. However, there is perhaps some reason to suppose that this is true, namely, that the loss of an arm or a leg makes no essential difference to oneself as a subject of experience. There are, after all, people who are born without arms or legs, but this makes them no less *people* and subjects of experience. However, even if we accept this line of argument, it doesn't serve to show that *no* part of one's body is part of oneself. For one cannot so easily contend that a loss of part of one's *brain* would make no essential difference to oneself as a subject of experience. Nor do we know of any people who have been born without brains. Of course, if Descartes were right in his earlier claim that he could exist in a completely disembodied state, then this would lend support to his view that even parts of his brain are not parts of

himself. But we have yet to be persuaded that that earlier claim is true. So it seems that, at this stage, Descartes's claim that he himself is a simple and indivisible substance is insufficiently compelling. This is not say that the claim may not be true, however, and I shall give it more consideration shortly.

NON-CARTESIAN DUALISM

So far we have failed to identify any compelling argument for the truth of Cartesian dualism, so perhaps we should give up dualism as a lost cause – especially if there are in addition some compelling arguments *against* it. But before looking at such counterarguments, we need to sound a note of caution. We shouldn't imagine that in rejecting Cartesian dualism we must automatically reject *every* form of 'substance dualism'. There is, in particular, one form of substance dualism which is untouched by any consideration so far raised, because it doesn't appeal to the kind of arguments which Descartes used in support of his position. According to this version of substance dualism, a person or subject of experience is, indeed, not to be identified with his or her body or any part of it, but nor is a person to be thought of as being an immaterial spirit or soul, nor even a combination of body and soul. On this view, indeed, there need exist no such things as immaterial souls. Rather, a person or subject of experience is to be thought of as a thing which possesses *both* mental *and* physical characteristics: a thing which feels and thinks but which also has shape, mass and a location in physical space. But why, it may be asked, should such a thing not simply be identified with a certain physical body or part of it, such as a brain?

At least two sorts of reason might be adduced for denying any such identity. The first is that mental states, such as thoughts and feelings, seem not to be properly attributable to something like a person's brain, nor even to a person's body as a whole, but only to a person himself or herself. One is inclined to urge that it is *I* who think and feel, *not* my brain or body, even if I need to *have* a brain and body in order to

be able to think and feel. (I shall say more in defence of this view in chapter 10.) The second and, I think, more immediately compelling reason is that the *persistence-conditions* of persons appear to be quite unlike those of anything such as a human body or brain. By the 'persistence-conditions' of objects (or 'substances') of a certain kind, I mean the conditions under which an object of that kind continues to survive as an object of that kind. A human body will continue to survive just so long as it consists of living cells which are suitably organised so as to sustain the normal biological functions of the body, such as respiration and digestion; and much the same is true of any individual bodily organ, such as the brain. However, it is not at all evident that I, as a person, could not survive the demise of my body and brain. One needn't appeal here, as Descartes does, to the supposed possibility that I could survive in an altogether disembodied state. That possibility is indeed very hard to establish. All that one need appeal to is the possibility that I might *exchange* my body or brain for another one, perhaps even one not composed of organic tissue at all but of quite different materials. For example, one might envisage the possibility of my brain cells being gradually and systematically replaced by electronic circuits, in such way as to sustain whatever function it is that those cells serve in enabling me to feel and think. If, at the end of such a process of replacement, I were still to exist as the same subject of experience or person as before, then I would have survived the demise of my present organic brain and so could not be identical with it. (Again, I shall discuss this sort of argument more fully in chapter 10.)

If this reasoning is persuasive, it supports a version of substance dualism according to which a person is *distinct* from his or her body, but is nonetheless something which, like the body, possesses physical characteristics, such as shape and mass. An analogy which may be helpful here is that provided by the relationship between a bronze statue and the lump of bronze of which it is composed. The statue, it seems, cannot be identical with the lump of bronze, because the statue may well have come into existence later than the lump did and

has persistence-conditions which are different from those of the lump: for instance, the statue would cease to survive if the lump were squashed flat, but the lump would continue to survive in these circumstances. However, the statue, although distinct from the lump, is none the less like it in having physical characteristics such as shape and mass: indeed, while it is composed of that lump, the statue has, of course, exactly the same shape and mass as the lump does. So too, it may be suggested, a person can have exactly the same shape and mass as his or her body does, without being identical with that body. However, the analogy may not be perfect. The statue is *composed* by the lump. Do we want to say that a person is, similarly, *composed* by his or her body? Perhaps not, for the following reason.

First, let us observe that, so long as the lump composes the statue, every part of the lump is a part of the statue: for example, every particle of bronze in the lump is a part of the statue. However, the reverse seems not to be the case: it doesn't seem correct to say that every part of the statue is a part of the lump of bronze. Thus, for instance, if the statue is a statue of a man, then the statue's arm will be one of its parts and yet it doesn't seem correct to say that the statue's arm is a part of the lump of bronze, even though it is correct to say that a part of the lump of bronze *composes* the arm. For the part of the lump of bronze which composes the statue's arm is not *identical* with the statue's arm, any more than the whole lump of bronze is identical with the statue. So the statue and the lump do not have exactly the same parts – which, of course, is an additional reason for saying that they are not identical with one another. Indeed, if they *did* have exactly the same parts, this would be a good reason for saying that they *were* identical with one another, because it is a widely accepted principle of mereology – the logic of part–whole relations – that things which have exactly the same parts are identical with one another.[4] Suppose that this prin-

[4] For a comprehensive modern treatment of mereology, see Peter Simons, *Parts: A Study in Ontology* (Oxford: Clarendon Press, 1987). I discuss part–whole relations more fully in my *Kinds of Being: A Study of Individuation, Identity and the Logic of*

ciple is correct, then, and turn to the case of a person and his or her body. If a person is *composed* by his or her body but not identical with it, then, it seems, by analogy with the statue and the lump of bronze, every part of the body must be a part of the person but not every part of the person can be part of the body: that is to say, the person must have certain parts *in addition to* parts of his or her body. However, it is very far from evident what these supplementary parts of the person could be, given that we have abandoned any suggestion that a person has an immaterial soul. It will not do to cite such items as a person's *arm*, for this *is*, of course, a part of the person's body. In this respect, the analogy with the statue and the lump of bronze breaks down, because the statue's arm plausibly is *not* a part of the lump. So, on the plausible assumption that a person has no parts which are not parts of his or her body – and yet is not identical with his or her body – it seems that we must deny that a person is *composed* by his or her body.

ARE PERSONS SIMPLE SUBSTANCES?

Now, if the preceding line of reasoning is correct, then we can reach a more remarkable conclusion, namely, that Descartes was right, after all, in thinking that he is a simple substance, altogether lacking any parts. The argument is simply this. First, we have argued that a person is not *identical* with his or her body nor with any part of it, on the grounds that persons and bodily items have different persistence-conditions. Secondly, we have argued that a person is not *composed* by his or her body nor – we may add – by any part of it. Our reason for saying this is that there appear to be no parts that a person could have *other than* parts of his or her body. However, if a person were to have as parts *only* parts of his or her body, then, according to the mereological principle

Sortal Terms (Oxford: Blackwell, 1989), ch. 6. Of course, we should not assume that principles of mereology, even if they are widely accepted ones, are immune to criticism.

mentioned earlier, it would follow, after all, that that person *would* be identical either with his or her body as a whole or with some part of it (depending on whether the parts in question were *all* the parts of the body or just *some* of them). And we have already ruled out any such identity. Consequently, a person can have *no parts at all* of which he or she is composed: a person must be a simple substance. But notice that this argument proceeds in the opposite direction to that in which Descartes argues. He argues from the premise that a person is a simple substance (together with certain other premises) to the conclusion that a person is not identical with his or her body, whereas we have just argued from the premise that a person is not identical with his or her body (together with certain other premises) to the conclusion that a person is a simple substance.

Of course, some philosophers might see the foregoing argument as a *reductio ad absurdum* of one or more of its premises, most likely the premise that a person is not identical with his or her body nor with any part of it. They will urge that it is just obvious that a human person has parts and that the only parts of a person are bodily parts, arguing thence to the conclusion that a person is identical with his or her body or some distinguished part of it. However, I don't think it really is obvious that a person has parts. That, perhaps, is why it is not easy for us to make clear sense of the notion of 'dividing a person in two'. If we remove any part of a person's body, it seems that either we are left with one person who is the same whole person as before or else we are left with no person at all. There are, of course, various science-fiction scenarios in which a single person is envisaged as dividing into two distinct persons, perhaps as a consequence of brain-bisection and transplantation. But whether we can really make sense of such stories is a matter for debate, to which we shall return in chapter 10. Again, there are actual cases of so-called 'multiple personality' syndrome in which, apparently, several different persons or subjects of experience manifest themselves within a single human body – and these different subjects are sometimes described as having resulted

from the fragmentation of what was originally a single subject or person. But how literally one can interpret such descriptions of these cases is also a matter for debate. Uncontentious examples of the division of one person into two or more different persons are simply not available. When a human mother gives birth to a child, it is indeed uncontentious that we begin with one person and end up with two: but it is certainly not uncontentious that this happens as a result of one person, the mother, dividing into two.

However, there is another objection to the claim that persons are simple substances, at least if this is combined with the claim that persons share with their bodies such physical characteristics as shape and height. For if persons are spatially extended, must they not be divisible into distinct parts – for instance, must I not have a left half and a right half? If that is so, does it not follow that anyone who maintains that a person is a simple substance must agree with Descartes that persons lack physical characteristics and thus are *immaterial* substances? No, it doesn't follow. For to accept that I have a left half and a right half is not to accept that these are parts of me into which I am divisible and which together compose me, in the way in which my body is composed of cells into which it is divisible. My 'left half' and 'right half' are not items which could, even in principle, exist independently of me, in the way in which individual cells of my body could exist independently of it: they are not, as we might put it, independent substances in their own right and so not items of which I am *composed*. Rather, they are mere abstractions, whose identity depends essentially upon their relation to me as the single person whose 'halves' they are.

I don't expect anyone to be completely convinced, on the basis of what I have said so far, that the non-Cartesian version of substance dualism sketched above is correct.[5] But I

[5] For a fuller exposition of the kind of non-Cartesian substance dualism talked about here, see my *Subjects of Experience* (Cambridge: Cambridge University Press, 1996), ch. 2. This position is similar in some ways to the view of persons defended by P. F. Strawson in his book *Individuals: An Essay in Descriptive Metaphysics* (London: Methuen, 1959), ch. 3, although Strawson would not happily describe himself as a 'dualist'.

hope at least to have demonstrated that questions concerning the ontological status of subjects of experience and their relations to their bodies are complex ones which require careful thought. Off-hand dismissals of substance dualism, treating 'Cartesian' dualism as the only version of it that is available, are not helpful. (I shall return to some of the issues raised here in chapter 10, when I discuss problems of personal identity.)

CONCEPTUAL OBJECTIONS TO DUALISTIC INTERACTION

Cartesian substance dualism is a form of *interactionist* dualism: that is, it maintains that mental states of a subject or person may and often do interact causally with physical states of that person's body, both causing such states and being caused by them. And in this respect the theory is fully in agreement with common sense. Unless we are philosophers, we unquestioningly believe that, for instance, damage to one's foot can cause one to feel pain and that a desire to raise one's arm can have the effect of that arm's going up. But for many critics of Cartesian dualism, its interactionism is its Achilles' heel. These critics hold that, because Cartesian dualism regards mental states as states of a wholly non-physical substance, it faces grave difficulties in maintaining that such states are causes and effects of physical states. What, exactly, are these supposed difficulties? They are of two types, one conceptual and the other empirical.

The alleged conceptual difficulties centre upon the contention that we cannot really make sense of there being causal transactions between items as radically different in nature from one another as the dualist conceives mental and physical states to be. The Cartesian dualist treats the two kinds of states as having virtually nothing in common – apart, perhaps, from their existence in time and their alleged capacity to enter into causal relationships. According to the Cartesian dualist, thoughts and feelings, being states of an immaterial substance altogether lacking location in physical space, must themselves lack location in physical space. So how, it may be

asked, can such mental states act upon or be acted upon by the physical states of a particular body? Perhaps the assumption here is that causation must always be *local*: that there can be no 'action at a distance', much less action between something located in physical space and something which lacks physical location altogether. Another assumption may be that whenever a causal transaction takes place, some property of the cause must be transmitted to the effect – as, for example, when motion in one billiard ball gives rise to motion in another upon impact, or when heat in a poker gives rise to heat in some water into which the poker is plunged. And then the objection to Cartesian dualism would be that, since it treats mental and physical states as having virtually no properties in common and hence no properties which could be transmitted between them, it leaves itself no scope for saying that there can be causal transactions between mental and physical states.

These objections are not particularly convincing. The idea that causation must be 'local' was effectively abandoned by the Newtonian theory of gravitation, some 300 years ago. And although the theory was criticised by contemporaries as being 'occult' on this account, these criticisms were rightly soon laid aside. It is true that some modern physicists propose that gravitational force is carried by particles known as 'gravitons', which would imply that gravitational effects are, after all, 'local' rather than being the results of 'action at a distance'. But the point is that this proposal is just part of an empirical theory – one which has yet to be strongly confirmed – not a consequence of a *conceptual* constraint on the intelligibility of the notion of gravitational attraction. Similarly, the claim that in any causal transaction some property must be transmitted from cause to effect does not express a conceptual truth. Indeed, it even appears to have many scientific counterexamples. For instance, motion in a body may be produced by a cause which does not itself involve motion, as when an electrically charged object moves under the influence of an electromagnetic field. Of course, it may be said that even here something *is* transmitted from cause

to effect, namely, *energy*, which may be converted from one form to another (for instance, from potential energy into kinetic energy). But the point, once again, is that this is not a consequence of a *conceptual* constraint on the intelligibility of the notion of such a causal transaction, but at most a consequence of an empirically well-confirmed theory concerning transactions of this kind.

David Hume long ago gave the decisive answer to all such conceptual objections to the possibility of dualistic mental–physical causation. This is that there are simply no *a priori* constraints on what *kinds* of states or events can enter into causal relationships with one another. As Hume himself puts it at one point: 'to consider the matter *a priori*, any thing may produce any thing'.[6] We discover what *does* produce what by *empirical* means, not least by observing that certain kinds of states or events are 'constantly conjoined' with other kinds of states or events. We can agree with Hume about this, even if we do not agree with him about how we should *define* causation (if, indeed, we think that it can be defined at all). Commentators dispute amongst themselves over precisely how Hume himself thought that causation should be defined, but the 'Humean' definition is widely taken to be something like this: to say that state S_1 caused (or was a cause of) state S_2 is to say that S_1 was followed by S_2 and that every state of the same kind as S_1 is followed by a state of the same kind as S_2.[7] We may concur with critics of this 'Humean' definition of causation in terms of 'constant conjunction' that it fails to capture important features of our concept of causation – for instance, that it fails to capture our conviction that, if state S_1 caused state S_2, then, other things being equal, if S_1 had

[6] See David Hume, *A Treatise of Human Nature* (1739–40), ed. L. A. Selby-Bigge and P. H. Nidditch (Oxford: Clarendon Press, 1978), Book I, Part IV, Section V: the sentence quoted in the text is taken from p. 247 of this edition.

[7] For a thought-provoking examination of Hume's views about causation in general, denying that Hume himself accepted a 'constant conjunction' theory of causation of the sort described in the text, see Galen Strawson, *The Secret Connexion: Causation, Realism, and David Hume* (Oxford: Clarendon Press, 1989). For an alternative account of Hume's position, see Tom L. Beauchamp and Alexander Rosenberg, *Hume and the Problem of Causation* (New York: Oxford University Press, 1981).

not existed, S_2 would not have existed either. But I think that we should, none the less, agree with Hume that, as I put it a moment ago, there are no *a priori* constraints on what *kinds* of states can be causally related to one another.

EMPIRICAL OBJECTIONS TO DUALISTIC INTERACTION

So let us turn to the empirical objections to Cartesian interactionism. These are best approached by seeing first why some of Descartes's contemporaries raised objections of this kind against his account of psychophysical causation. Descartes supposed that interaction between the non-physical self and its body takes effect in a specific organ situated in the middle of the brain, the pineal gland. This seemed to him the most probable seat of mind–brain interaction, not only because of the gland's central location, but also because it is unique, whereas many other brain-structures are duplicated in the brain's two hemispheres. Hence, he thought, the pineal gland could readily serve as a unitary control-centre for the whole brain and nervous system. (With the benefit of hindsight, we now know that the pineal gland serves no such purpose, but Descartes's hypothesis was not unreasonable in his own day.) Descartes regarded the nervous system as something like a network of pipes and valves operating in accordance with hydraulic principles, with the nerve filaments conducting quantities of so-called 'animal spirits' to and fro throughout the body. These animal spirits were thought to be a rarefied and highly motile fluid, capable of flowing freely and rapidly through tiny pores in the nerve filaments. (Thus, the term 'spirits', in this context, was certainly not intended to denote something *immaterial* in nature.) Descartes believed that, when these animal spirits flowed through the nerve filaments in the region of the pineal gland, the non-physical self could subtly alter their direction of flow, thus giving rise to variations in their movements. These variations, after being conveyed by nerve filaments to the body's extremities, could ultimately bring about concomitant variations in the movements of a person's limbs. In the reverse

direction, movements in the animal spirits at the body's extremities, brought about by impact with external objects, could be conveyed to the central region of the brain and there – so Descartes supposed – cause a person to undergo appropriate experiences, such as those of pain or pleasure.

An important feature of Descartes's theory was that it held that the self acted upon its body only by altering the *direction* of motion of the animal spirits, not by imparting new motion to them. Indeed, Descartes believed that the total 'quantity of motion' in the physical universe never changes, but is only redistributed amongst material bodies as they interact with each other upon impact. Thus, in a collision between two material bodies, one which was formerly at rest might begin to move, but only at the expense of the other body losing some of its 'quantity of motion'. The principle that Descartes was advocating here was an early form of *conservation law*. Modern physics recognises several such laws, the most important being the law of the conservation of momentum and the law of the conservation of energy. Unfortunately – as Descartes's near-contemporary Leibniz appreciated – Descartes's conservation law is incompatible with these modern laws.[8] In particular, Descartes did not seem to realise that one cannot alter the *direction* of motion of a material body without altering its *momentum*, which is a conserved quantity according to the modern laws. A body's momentum is the product of its mass and its velocity. And velocity is a *vector* rather than a scalar quantity. What this means is that if a body is undergoing a change in its direction of motion, then it is undergoing a change of velocity and hence a change of momentum, even if it is moving at a constant speed. Thus, for example, a body which is moving in a circular path at a

[8] Leibniz's criticisms of Cartesian interactionism are interestingly discussed by R. S. Woolhouse in his 'Leibniz's Reaction to Cartesian Interactionism', *Proceedings of the Aristotelian Society* 86 (1985/86), pp. 69–82 and by Daniel Garber in his 'Mind, Body, and the Laws of Nature in Descartes and Leibniz', *Midwest Studies in Philosophy* 8 (1983), pp. 105–33. I myself go more fully into some of the issues raised in the text concerning the conservation laws of physics in my *Subjects of Experience*, pp. 56–63.

constant rate is none the less constantly undergoing a change in its velocity. The modern laws of motion, first postulated by Newton, state that a massive body can only undergo a change of velocity if it is acted upon by a corresponding force. Hence, it appears that these laws imply that any change of direction in the motion of the animal spirits – which Descartes thought could be brought about by the non-physical self – would in fact have to be a consequence of some *force* acting upon the animal spirits: and it is hard to see how the Cartesian self, being non-physical, could be the source of any such force. In the light of this apparent difficulty, Leibniz himself rejected interactionism in favour of *parallelism* – the doctrine that mental and physical states never interact causally, but merely 'keep in step' with one another in accordance with a divinely preordained plan. However, an alternative and perhaps more plausible response is to accept that mental and physical states interact causally, but to reject the dualist claim that mental states are quite distinct from physical states. Instead, one could espouse the *physicalist* claim that mental states just *are* – that is, are identical with – certain physical states, most plausibly certain neural states of the brain.

THE CAUSAL CLOSURE ARGUMENT

Of course, the preceding line of argument against Cartesian interactionism not only appeals to quite specific empirical premises, in the form of the conservation laws of modern physics, but also has as its target the quite specific mechanism of mind–brain interaction proposed by Descartes himself. Thus, it leaves at least two possible escape-routes for the would-be dualist advocate of mind–brain interaction. One is to challenge the presumed correctness of the laws in question, though this might seem a foolhardy venture in view of their well-entrenched status in modern physics. The other is to propose some quite different system of causal interaction between mental and physical states. This then raises the question of whether the physicalist might not be able to present some much more general line of empirical argument against dualist interactionism, which would rule out any conceivable system of mind–brain interaction of a dualist character. In the view of many present-day physicalists, just such a line of argument is indeed available and it is one which is, moreover, fairly simple in form.[9] The argument has three premises, as follows.

(1) At every time at which a physical state has a cause, it has a fully sufficient physical cause. (Call this premise *the principle of the causal closure of the physical*.)

(2) Some physical states have mental states amongst their causes. (Call this premise *the principle of psychophysical causation*.)

(3) When a physical state has a mental state amongst its causes, it is rarely if ever causally overdetermined by that mental state and some other physical state. (Call this premise *the principle of causal non-overdetermination*.)

These premises call for a few explanatory comments. Premise (1) means that if P is a physical state which has a cause existing at a certain time t, then there is a non-empty set of physical states, all of them existing at t, such that (a) each of these states is a cause of P and (b) collectively these states are causally sufficient for P. (To say that a number of physical states are *collectively causally sufficient* for another physical state, P, is to say that, given that all of those states exist, it follows of causal necessity that P also exists.) Premise (2) is self-explanatory. Premise (3) rules out the possibility that, whenever a mental state M is a cause of a physical state P, there is another physical state Q such that (a) Q is a cause of P and yet (b) even if one of the two states M and Q had not existed, the other would still have sufficed, in the circumstances, to cause P to exist. Causal overdetermination of the kind generally ruled out by premise (3) can be illustrated (using a non-psychophysical example) as follows. Suppose

[9] A good example of a version of this argument may be found in David Papineau's recent book, *Philosophical Naturalism* (Oxford: Blackwell, 1993), ch. 1.

that two assassins independently shoot at the same time and both bullets inflict fatal wounds upon their victim, who promptly dies: in this case, the victim's death is causally overdetermined by the two acts of shooting, since (a) each act is a cause of the death and yet (b) even if one of the acts had not existed, the other would still have sufficed, in the circumstances, to cause the death. Premise (3) rules out the possibility that, as a general rule, mental states cause physical states rather in the way in which one of these acts of shooting causes the death, that is, in such a way that the physical effects of those mental states simultaneously have independent but fully sufficient physical causes.

I shall consider in a moment the question of how plausible these three premises are, but first we must see what they are supposed to entail. Physicalists take them to entail the following conclusion:

(4) At least some mental states are identical with certain physical states.

How does (4) follow from premises (1), (2) and (3)? In the following way, it would seem. Suppose, in accordance with premise (2), that M is a mental state, existing at a time t, and that M is a cause of a certain physical state, P. From premise (1), we can infer that there is a non-empty set of physical states, all of them existing at t, which are collectively causally sufficient for P. Call these physical states $P_1, P_2, \ldots P_n$. Finally, suppose, in accordance with premise (3), that P is not causally overdetermined by M and any one of these physical states. That is to say, suppose it is not the case that there is one of these physical states, say P_i, such that if either one of the states M and P_i had not existed, the other would still have sufficed, in conjunction with the remaining physical states in the set, to cause P to exist. Then it would seem that we have no option but to *identify* the mental state M with one or other of the physical states $P_1, P_2, \ldots P_n$. For suppose that M is *not* identical with any of these states. We have assumed that M is a cause of P but we have also assumed that the physical states $P_1, P_2, \ldots P_n$ are collectively causally sufficient for P. Hence we can apparently infer that even if M had not existed but all of $P_1, P_2, \ldots P_n$ had still existed – which must be a possibility given that M is not identical with any of $P_1, P_2, \ldots P_n$ – then $P_1, P_2, \ldots P_n$ would still have sufficed to cause P to exist. But this is to imply that P is causally overdetermined by M and one or more of $P_1, P_2, \ldots P_n$, contrary to what we have hitherto assumed. Hence we must reject the supposition that M is not identical with any one of $P_1, P_2, \ldots P_n$ and conclude that (4) is true.

OBJECTIONS TO THE CAUSAL CLOSURE ARGUMENT

I shall not question the validity of the foregoing argument, but we need to examine its premises carefully before accepting it. Obviously, advocates of dualist interactionism, who are the targets of the argument, will not want to query premise (2), since they accept it themselves. But they might well be wary of accepting premise (3), once they see what it leads to. Perhaps they would be well advised to maintain that systematic causal overdetermination is a pervasive feature of psychophysical causation.[10] It is not obvious that this is an untenable position to adopt, although to adopt it merely in order to evade the unwelcome conclusion of the physicalist's argument would be blatantly *ad hoc*. To make this position credible, the dualist needs to find some independent reason for supposing that such causal overdetermination is commonplace – and it is far from clear what that reason might be. The dualist's best hope, then, would seem to lie in challenging premise (1), the principle of the causal closure of the physical. As I have stated the principle here, it is an extremely strong one which – the dualist might urge – cannot plausibly be claimed to be an indisputable implication of currently accepted empirical science. Premise (1) must be carefully distinguished from a weaker and correspondingly more plausible principle, namely,

[10] For a defence of such a position, see Eugene Mills, 'Interactionism and Overdetermination', *American Philosophical Quarterly* 33 (1996), pp. 105–17.

(1*) Every physical state has a fully sufficient physical cause.

To distinguish them, we could call (1) and (1*), respectively, the *strong* and the *weak* principle of the causal closure of the physical. I should perhaps point out that, despite these titles, there is one respect in which (1*) is, in fact, stronger than (1). This is that (1*) implies, as (1) does not, that there are no *uncaused* physical states. That assumption is not altogether uncontroversial, as we shall see when we come to discuss the issue of free will and determinism in chapter 9. It also raises cosmological and theological questions of whether there was a 'first' cause. However, I shall set aside any such concerns here. (In particular, I shall ignore the fact that modern quantum physics suggests that at least some physical states do *not* have fully sufficient physical causes. Perhaps, however, it is safe to ignore this fact in the present context, given that quantum physics is chiefly concerned with physical phenomena on the atomic scale, rather than at the level of neural structure and function in the brain.) We shall see in a moment precisely in what sense (1) *is* stronger than (1*).

If we replace (1) by (1*) in the physicalist argument examined above, we can no longer draw the physicalist conclusion, (4), that at least some mental states are identical with physical states. The explanation of this involves the fact that causation is a transitive relation. Causation is transitive inasmuch as if a state S_1 is a cause of a state S_2, and S_2 is a cause of a state S_3, then it follows that S_1 is a cause of S_3. Moreover, if S_1 is a *fully sufficient* cause of S_2, and S_2 is a *fully sufficient* cause of S_3, then S_1 is a *fully sufficient* cause of S_3. However – and this is the crucial point – if S_1 and S_2 are thus *both* fully sufficient causes of S_3, because S_1 is a fully sufficient cause of S_2 and S_2 is a fully sufficient cause of S_3, this does not imply that S_3 is *causally overdetermined* by S_1 and S_2. Consequently, the following situation is perfectly consistent with the truth of principles (1*), (2) and (3): M may be a mental state which is *not* identical with any physical state and yet which is also a cause of a certain physical state, P. In this case, principle (2) is obviously satisfied. But principle (1*), the weak principle of the causal closure of the physical, may also be satisfied in respect of the physical state P. All that is required is that the mental state M should itself have a fully sufficient physical cause, say Q: if it does, then, by the transitivity of causation, Q is a cause of P and, moreover, may evidently either itself be, or else be part of, a *fully sufficient* physical cause of P. Furthermore, for the reason just explained, nothing here implies that P is *causally overdetermined* by M and any other cause of P, so that principle (3) is also satisfied. Thus, even an advocate of dualist interactionism could happily accept the weak principle of the causal closure of the physical, (1*). Indeed, a dualist who espoused a doctrine of *emergentism* concerning the mental would have a positive reason for endorsing (1*). Emergentists hold that mental states have not always existed in the spacetime universe and that at one time all states of the universe were purely physical – for instance, at the time of the so-called 'Big Bang' when the universe was created. They further hold that mental states have come into existence as a result of the natural evolution of highly complex biological entities, rather than through any kind of divine intervention by a being who exists 'outside' the spacetime universe, such as God. Consequently, they hold that all mental states ultimately have fully sufficient causes which are purely physical in character. What we can now see is that this emergentist doctrine is consistent not only with dualist interactionism, but also with the weak principle of the causal closure of the physical.[11]

The question that we must ask ourselves now is whether we have any reason to suppose that not only the weak but also the *strong* principle of the causal closure of the physical is correct. The difference between them is precisely this: the strong principle requires that at *every* time at which it has a cause, any physical state has a fully sufficient physical cause, whereas the weak principle requires only that at *some* time at which it has a cause, any physical state has a fully sufficient

[11] I defend an emergentist doctrine of dualist interactionism in my *Subjects of Experience*, ch. 3.

physical cause. The sort of consideration which supports the weak principle – namely, that, as far as we know, the spacetime universe at its beginning contained only physical states and has not been subject to intervention from 'outside' since then – certainly does not serve to support the strong principle. Moreover, we need to appreciate what a *very* strong claim the strong principle makes. It implies, for instance, that if a physical state has a cause *one nanosecond* before it comes into existence, then it has a fully sufficient physical cause *at that time*. It is not at all clear to me what currently available empirical evidence could be taken to support such a strong claim quite generally. But perhaps we should not be surprised that it is only possible to mount a decisive argument against dualist interactionism if one starts from a premise, like (1), which already embodies strong physicalist presumptions. Knock-down arguments in philosophy often appear, upon closer examination, to be implicitly question-begging. That is why fundamental philosophical disputes, such as that between dualism and physicalism, stubbornly refuse to go away. Many dualist philosophers find it so incredible that a mental state, such as a feeling or a thought, could just *be* – that is, be identical with – a physical state, such as a state of neuronal activity in the brain, that they would find it more compelling to argue from (2), (3) and the *negation* of (4) to the *negation* of (1) than to argue, as the physicalist does, from (1), (2) and (3) to (4). Both arguments are valid, if either is: the only question is which set of premises is the more plausible and the more defensible. But that question, I think, is at present still an open one.

OTHER ARGUMENTS FOR AND AGAINST PHYSICALISM

The causal closure argument is not the only general argument that has been advanced in favour of identifying mental states or events with physical states or events. Another well-known argument in favour of this identity thesis has been

advanced by Donald Davidson.[12] Like the causal closure argument, Davidson's argument has three premises, the first of which is similar to premise (2) of the causal closure argument:

(5) Some mental events interact causally with physical events. (Davidson calls this premise *the principle of causal interaction*.)

Davidson's remaining two premises are these:

(6) Events related as cause and effect fall under strict deterministic laws. (Davidson calls this premise *the principle of the nomological character of causality*.)

(7) There are no strict deterministic psychophysical laws. (Davidson calls this premise *the principle of the anomalism of the mental*.)

Davidson's conclusion, similar to the conclusion (4) of the causal closure argument, is that at least some mental events are identical with physical events. To see how the argument works, suppose that a certain mental event, *M*, causes a certain physical event, *P*, in line with premise (5). It follows, by premise (6), that *M* and *P* must be characterisable in terms which allow them to fall under a strict deterministic law. But according to premise (7), there are no such laws under which *M* falls in virtue of being characterisable in mentalistic terms, so that it would seem that *M* must also be characterisable in *physical* terms and thus qualify as a physical event. Interesting though Davidson's argument is, I shall not examine here his reasons for advancing premises (6) and (7), which are evidently controversial – even more so, indeed, than premises (1) and (3) of the causal closure argument. It

[12] See Donald Davidson, 'Mental Events', in L. Foster and J. W. Swanson (eds.), *Experience and Theory* (London: Duckworth, 1970), reprinted in Davidson's *Essays on Actions and Events* (Oxford: Clarendon Press, 1980). Davidson's version of the identity theory is couched in terms of *events* rather than *states*, because he regards causation as a relation between events: but for our present purposes nothing of significance turns on the distinction between events and states.

is debatable whether, whenever one event causes another, there is always some general causal law under which this particular causal interaction may be subsumed as an instance.[13] And it is debatable whether, even in physics, there are strict – that is, exceptionless – and deterministic laws of causal interaction.[14] In view of these doubts, physicalists might be better advised to pin their hopes on the causal closure argument, contestable though it is.

On the other side, there are general arguments which have been advanced *against* the thesis that mental events or states are identical with physical events or states.[15] Some physicalists have urged that, just as science has revealed to us that water is simply H_2O and that heat in a gas is simply the mean kinetic energy of its constituent molecules, so neuroscience will one day reveal to us that pain, say, is simply a certain neural state, such as the stimulation of C-fibres. These physicalists believe, then, that the discovery that pain just is – that is, is identical with – a certain neural state will be an empirical or *a posteriori* one, resembling discoveries that science has already made in many other fields of enquiry. However, there seems to be an important disanalogy between the proposed identification of pain with something like C-fibre stimulation and the other 'scientific' identifications exemplified a moment ago. Take the identification of water with H_2O. We can readily understand why it took empirical enquiry to establish this identity, because we can easily imagine encountering a substance which to all outward appear-

[13] For scepticism on this point, see G. E. M. Anscombe, 'Causality and Determination', in her *Metaphysics and the Philosophy of Mind*. Of course, the 'Humean' definition of causation in terms of 'constant conjunction', discussed earlier in this chapter, does imply the truth of Davidson's premise (6): but that definition is itself highly controversial.

[14] For doubts on this score, see Nancy Cartwright, *How the Laws of Physics Lie* (Oxford: Clarendon Press, 1983).

[15] The thoughts which follow are partly inspired by Saul A. Kripke's *Naming and Necessity* (Oxford: Blackwell, 1980), first published in G. Harman and D. Davidson (eds.), *Semantics of Natural Language* (Dordrecht: D. Reidel, 1972): see especially pp. 144–55. I should remark that, while I am ignoring for present purposes the distinction between 'type-type' and 'token-token' identity theories, this is something that I shall have more to say about in chapter 3.

ance *looks* just like water, but which turns out upon investigation to have a quite different chemical constitution. This pseudo-water, as we might call it, would not *be* water, given that the chemists are right in telling us that water is in fact H_2O. But now suppose, in like manner, that future neuroscientists tell us that pain just *is* the stimulation of C-fibres. And suppose, too, that we were to encounter a mental state, perhaps in some creature belonging to an alien species, which *felt* just like pain but which, none the less, could *not* be identified with the stimulation of C-fibres – because, say, this creature simply does not have C-fibres. Then it follows, if the analogy is sound, that we should have to say that what this creature feels is not, after all, *pain* at all, but only 'pseudo-pain'. But that seems absurd, because what *feels* like pain surely *is* pain. And the same absurdity would threaten to beset us *whatever* physical state the neuroscientists proposed to identify with pain, thus calling into question the intelligibility of *any* such identification. It is very tempting to conclude from this that a mental state such as pain simply *cannot* be identified with any physical state whatever.

Of course, it might be urged on behalf of the physicalists that if neuroscientists *do* one day discover that pain just is the stimulation of C-fibres, then for that very reason we shall *not* be able to encounter a creature having a mental state which feels just like pain even though it lacks C-fibres. The fact that we can *imagine* encountering such a creature is no guarantee that such a creature really *could* exist because – as we saw when we were discussing Descartes's 'conceivability argument' for the distinction between himself and his body – even something which is logically impossible may be 'imaginable'. However, physicalists who respond in this manner at least owe us an explanation of why it should be that imagination is especially prone to mislead us in this sort of case. Indeed, they owe us a good deal more than that, because what is at issue is the very intelligibility of identifying pain with some physical state. Until that issue is resolved, it seems idle to speculate about what we could or could not encounter *if* neuroscientists discover that pain 'just is' the stimulation

of C-fibres, since such speculation appears to presume that we can, after all, make sense of such a discovery being made. More generally, the point is that anyone who wishes to propose a certain kind of theoretical identification owes us a demonstration that such an identification is in principle *possible* – and it is not at all clear that physicalists in the philosophy of mind have yet succeeded in doing this. Here is another analogy: no one would be remotely satisfied if some Pythagorean philosopher were to tell us that he had discovered a wealth of empirical evidence supporting the thesis that the things which we think of as being material objects really 'just are' *numbers*. We should demand, first, that he explain to us how such an identification even *makes sense*. To many of their opponents, physicalists who propose that mental states really 'just are' physical states seem to be in much the same sort of position as this hypothetical Pythagorean.[16]

CONCLUSIONS

In this chapter, we have looked at a number of fundamental metaphysical problems in the philosophy of mind, beginning with the question of how persons, or subjects of experience, are related to their physical bodies. We saw that 'Cartesian' substance dualism, which holds that persons are wholly immaterial things capable of existing independently of the bodies which they may happen to 'inhabit', has no very compelling argument in its favour. But we also saw that there are reasons for doubting that persons can simply be *identified* with their bodies, or with any distinguished part of their bodies, such as their brains. It may be that a person is related to his or her body somewhat in the way in which a bronze statue is related to the lump of bronze which composes it –

[16] For this analogy, see P. T. Geach, *Truth, Love and Immortality: An Introduction to McTaggart's Philosophy* (London: Hutchinson, 1979), p. 134. I say more about the conceptual difficulties involved in trying to identify mental states and events with physical states and events in my *Kinds of Being*, pp. 113–14 and pp. 132–3.

though even this analogy would appear to be imperfect. Perhaps we should see personal 'embodiment' as a unique kind of relationship in its own right, one which can be reduced neither to a mere causal relationship, nor to identity, nor to composition. (Perhaps, indeed, this is what the historical Descartes had in mind in talking of the relation of 'substantial union' between self and body.) It would be facile to suppose that a reductionist account of the relationship of embodiment *must* be available, even in principle. *Some* relationships, after all, must be primitive and irreducible and we may in the end have to accept that embodiment is one of these. On the other hand, we should not be too quick to accept this conclusion, either, while we continue to lack a convincing argument that a reductionist account is unattainable. At present, then, it seems that we should retain an open mind about this question.

The other main question which we examined was whether mental states ought to be identified with physical states, such as states of neuronal activity in the brain, given the plausible assumptions that mental states stand in causal relations to such physical states and yet that systematic causal overdetermination is not a feature of psychophysical causation. Here we saw that, while there is no satisfactory *a priori* argument against interactionist dualism, Descartes's own favoured system of psychophysical causation seems to fall foul of the currently accepted conservation laws of physics. However, we also saw that it is rather more difficult to rule out interactionist dualism in a perfectly general way, by appealing to some version of the principle of the causal closure of the physical, because only a rather strong version of that principle will suffice for this purpose and such a version of it cannot claim to be strongly supported by currently available empirical evidence. Once again, then, it seems that we should keep an open mind about this question – especially in view of the issues raised, in the last section, about the very intelligibility of identifying mental states with physical states. That we have come to no firm conclusion about either of the main questions addressed in this chapter should not disconcert us

unduly, however. These questions are amongst the most difficult in the whole of philosophy and it would be surprising if we had been able to resolve them conclusively when so many other able thinkers have failed to do so.

3

Mental states

In the previous chapter, we focused on two important metaphysical questions in the philosophy of mind. One was the question of whether persons or subjects of experience are identical with their physical bodies, or certain parts of those bodies, such as their brains. The other was the question of whether the mental states of persons, such as thoughts and feelings, are identical with certain physical states of their bodies, such as states of neuronal activity in their brains. Many materialists would endorse positive answers to both of these questions, although later in this chapter we shall encounter a species of materialism which denies that mental states, as we ordinarily conceive of them, really exist at all. But before we examine that position, it is worth remarking that, so long as one is a *realist* about mental states – that is, so long as one considers that states of thinking and feeling really do exist – one can, for many purposes, afford to remain neutral with regard to the question of whether or not mental states are identical with physical states. There are many issues in the philosophy of mind which we can usefully discuss without presuming to be able to resolve that question. And this is just as well, knowing as we now do how thorny a question it is. One of these issues is that of how we can best *characterise* and *classify* the various different kinds of mental state which, if we are realists, we believe to exist. So far we have been talking about mental states quite generally, without differentiating between them in any significant fashion. But in a detailed description of the mental lives of persons we need to be able to distinguish, in principled ways, between