Chapter 4

Dualism

Howard Robinson

4.1 Introduction

Dualism in the philosophy of mind is the doctrine that mind and body (or mental states and physical states) are of radically different natures. How exactly to express this difference is a matter of controversy, but it is generally taken to center on two properties possessed by the mental that are alien to the physical. One of these is the privacy or subjectivity of states of consciousness, as contrasted to the public availability of physical states. The other is the possession of intentionality or "aboutness" by mental states: physical states stand in spatio-temporal and causal relations to each other, but are not intrinsically *about* anything. The principle task for the physicalist is to give an account of these properties in physical or physical-compatible terms. A dualist is someone who thinks that this cannot be done.[1]

There are normally thought to be two forms of dualism, namely substance dualism and bundle dualism. The former is primarily associated with Descartes and the latter with Hume.[2] An important distinction must be made amongst bundle dualists, however. Some, like Hume, do not believe in either mental or physical substance, treating both as just collections of states, properties, or events (depending on how the theory is stated). For others, it is only the mind that is given this treatment: bodies are substantial entities, but minds only collections of states, properties, or events. This constitutes a relative downgrading of the mind and a move toward the *attribute* theory. According to this theory, mental states are non-physical attributes of a physical substance – the human body or brain. This theory can be regarded as the softest or least reductive form of materialism. It is materialistic because it says that the only substances are material substances. It is also a form of dualism, because it allows the irreducibility of mental states and properties.

Both substance and bundle dualisms face the same three problems. The first problem is to show why we need to be dualists at all – why a materialist account

of the mind will not work. The second is to explain the nature of the unity of the immaterial mind. For the Cartesian, that means explaining how he understands the notion of immaterial substance. For the Humean, the issue is to explain the nature of the relationship between the different elements in the bundle that binds them into one thing.[3] Neither tradition has been notably successful in this latter task: indeed, Hume declared himself wholly mystified by the problem, rejecting his own initial solution (though quite why is not clear from the text).[4] The third problem is to give a satisfactory account of the relationship between the immaterial mind and the material body. Which means, for preference, to explain how they can interact, and, failing this, to render plausible either epiphenomenalism (the view that the mental is produced by the physical, but has no influence back on the physical) or parallelism (the view that mental and physical realms "march in step," but without either causally interacting with the other).

I shall use the excuse of limited space for not dealing with all these issues. Rather, I shall attempt, in Cartesian spirit, to show, first, that the thinking subject has to transcend the physical world; and, secondly, that such subjects must be essentially simple. They (that is, we) are more like the immaterial substance in which Descartes believed, than like a Humean bundle of mental events or states. So I shall be concerned with why we should be dualists, and why dualists of a Cartesian stripe. How to explain the unity of the mind – except by showing it to be essentially simple – and how to explain our relations to our bodies, are not issues I can discuss here.[5]

In order to accomplish the first of the tasks I have set myself (that is, to show that the thinking subject must transcend the physical world), I shall introduce a form of dualism not so far mentioned, and which is generally neglected in discussions of dualism, namely *predicate dualism*. That is the theory that psychological or mentalistic predicates are not reducible to physicalistic predicates. (What this means I shall discuss in the next section.) Few philosophers nowadays either believe in such reduction or think that it is necessary for physicalism. Predicate dualism is only dualism at the level of meaning, and this is generally thought to have no ontological consequences. I shall be arguing that this is a mistake, and that predicate dualism – the failure of reduction – is a threat to physicalism because the irreducibility of the special sciences in general implies that the mind is not an integral part of the physical realm with which those sciences deal.

This conclusion does not alone force us to adopt any particular form of dualism. Perhaps the mind, though it transcends the physical world about which it constructs the sciences, is no more than a bundle of mental states or properties, as Hume thought. Perhaps, that is, predicate dualism forces us to nothing more than property dualism, which may not drive one further away from physicalism than the attribute theory. I shall then attempt to show that this is not so, for property dualism is not adequate to cope with certain respects in which *personal* identity is demonstrably different from the identity conditions for physical bodies and other complex entities: these constraints on personal identity can be met only by substance dualism of a roughly Cartesian kind.

4.2 The Argument for Predicate Dualism

If physicalism is true, then it should be possible, in principle, to give what is, in some sense, a total description of the world in the vocabulary of a completed physics. To put it in the material, not the formal, mode, all the properties that there *ultimately* are should be those of the basic physical entities. But there are many ways of talking truly about the world other than that couched in the vocabulary of physics; and there are, in some obvious sense, many properties that the world possesses that are not contained in that physics. These higher-order predicates and properties are expressed in the other – or *special* – sciences, such as chemistry, biology, citology, epidemiology, geology, metereology, psychology, and the supposed social sciences; not to mention our ordinary discourse, which often expresses truths that find no place in anything we would naturally call a science. How does the fundamental level of ontology – which we are presupposing to be captured ideally in physics – sustain all these other ontologies and make true these other levels of discourse?

The logical positivists had a simple answer to this question. Any respectable level of discourse was reducible to some level below it and ultimately to physics itself. The kind of reduction of which we are talking has a *strong* form and a *very strong* form. According to the *very strong* form, all respectable statements in the special sciences and in ordinary discourse could, in principle, be translatable into statements in the language of physics. In the end, therefore, all truths could be expressed using the language of physics.[6] According to the merely *strong* form – which was the form in which reductionism was generally discussed – there had only to be scientific laws (called "bridging laws") connecting the concepts and laws in a higher-order science with those in the next lower, and ultimately to physics.[7] So the concepts and laws of psychology would be nomically connected to those of some biological science, and these, in turn, with chemistry, and chemistry would be nomically reducible to physics. So "reducible to," in this sense, meant that the entities and properties invoked in the non-basic discourse were *type identical with* certain basic structures. For example, our ordinary concept *water* is reducible to the chemical type H_2O, and this chemical molecule always consists of the same atomic arrangements. This pattern makes it easy to understand intuitively how the existence of water and the truths of sentences referring to water need involve nothing more than the existence of things in the ontology of physics.

But not all concepts in the special sciences, let alone ordinary discourse and the social sciences, can be fitted into this pattern. Not every hurricane that might be invoked in metereology, or every tectonic shift that might be mentioned in geology, will have the same chemical or physical constitution. Indeed, it is barely conceivable that any two would be similar in this way. Nor will every infectious disease, or every cancerous growth, not to mention every devaluation of the currency or every *coup d'état* share similar structures in depth. Jerry Fodor, in his

important article "Special Sciences" (1974), correctly claims that the doctrine of reductionism requires that all our scientifically legitimate concepts be *natural kind* concepts and – like water – carry their similarities down to the foundations, and that this is not plausible for most of our useful explanatory concepts. It is particularly not plausible for the concepts of psychological science, understood in functionalist terms, nor for the concepts in our lay mentalistic vocabulary. All these concepts are *multiply realizable*, which means that different instances of the same kind of thing can be quite different at lower levels – in their "hard wear" – and that it is only by applying the concepts from the special science that the different cases can be seen as saliently similar at all. Whereas you could eliminate the word "water" and speak always of "H_2O" with no loss of communicative power, you could not do this for "living animal," "thought of the Eiffel Tower," "continental drift," etc.

Fodor (1974) thinks that this is no threat to physicalism, because each instance of a higher-order concept will be identical with some structure describable in terms of basic physics, and nothing more. Token reductionism is all that physicalism and the unity of the sciences require: type reduction is unnecessary. I shall now try to explain why, contrary to appearances, this is wrong.

4.3 Why Predicate Dualism leads to Dualism Proper

Fodor is quite right to think that the very same subject matter can be described in irreducibly different ways and still be just that subject matter. What, in my view, he fails to notice is that such different explanatory frameworks presuppose a perspective on that subject matter which is, prima facie, from outside of it. The outline of my position is as follows. On a realist construal, the completed physics cuts physical reality up at its ultimate joints: any special science which is nomically strictly reducible to physics also, in virtue of this reduction, it could be argued, cuts reality at its joints, but not at its minutest ones. By contrast, a science which is not nomically reducible to physics does not take its legitimation from the underlying reality in this direct way; rather, it is formed from the collaboration between, on the one hand, objective similarities in the world and, on the other, perspectives and interests of those that devise the science. If scientific realism is true, a completed physics will tell one how the world is, independently of any special interest or concern: it is just *how the world is.* Plate tectonics, however, tell you how it is from the perspective of an interest in the development of continents, and talk about hurricanes and cold fronts from the perspective of an interest in the weather. A *selection* of phenomena with a certain *teleology* in mind is required before these structures or patterns are reified. The point is that these sciences and the entities that they postulate exist from certain intellectual perspectives, and a perspective, whether perceptual or intellectual, is external to that on which it is a perspective.[8] The problem for the physicalist is to say what it is for a perspective

on the physical world to be something *within* it. A unified naturalistic view of the world would require that the observer's perspective required by these sciences be integrated into the reality he observes. The integration of perspectives and interests into the one world requires the integration of psychological states of both perceptual and intentional kinds into the physical world. These, however, are paradigms of the kinds of state that seem to resist nomic – type – reduction to physics.

There are, of course, famous arguments that appeal to the phenomenology of consciousness for thinking that token reductions fail: but no appeal to these is involved in the current argument.[9] Even if token reductionism of the mind could meet the phenomenological problems, the fact that it is token, not type, means that it presupposes the existence of a perspective from which the physical world is seen in order to bring out these facts. The perspective that makes possible the nomically irreducible sciences, being itself irreducible, could itself exist (if it were physical) only from a perspective on physical reality. As this second perspective is essentially of the same kind as the one we are trying to explain, namely a psychological or intellectual perspective, there is no prospect of a non-vicious regress here.

We can now understand the motivation for full-blown reduction. A true basic physics represents the world as it is in itself, and if the special sciences were reducible, then the existence of their ontologies would make sense as expressions of the physical, not just as ways of seeing or interpreting it – they could be understood "from the bottom up," not from above down. The irreducibility of the special sciences creates no problem for the dualist, who sees the explanatory endeavor of the physical sciences as something carried on from a perspective conceptually outside of the physical world. Nor need it worry a physicalist, *if* he can reduce psychology, for then he could understand "from the bottom up" the acts (with their internal, intentional contents) which created the irreducible ontologies of the other sciences. But psychology is one of the least likely of sciences to be reduced.

4.4 Is the Talk of "Perspectives" Legitimate?

Someone who wished to resist this line of argument might deny the claim that the nomically irreducible sciences cannot be given a fully realist interpretation, but are a perspective on the reality. He might argue that the foundations of the special sciences are what Dennett (1991) calls "real patterns" in reality, and that these are as objective as the structures of the ultimate and reducible sciences.

This misses the point. My position is not to deny that the "real patterns" on which the special sciences are based are objective and genuine, but that, as well as this *fundamentum in re*, those sciences require an interpretative component which takes these similarities and picks them out as interesting for certain purposes.

The relation between an ideal physics and the nomically irreducible special sciences is like that between straightforward phenomena and *Gestalt* phenomena.

Entities in physics are analogous to a perfectly circular object, which needs no interpretation to be taken as a circle: those in irreducible special sciences are like a series of discontinuous dots or marks arranged roughly in a circle which one sees as circular. Two hurricanes, for example, are not perfectly similar and would present themselves as a kind only to someone with an interest in weather: plate tectonics exist only given an interest in the habitability of the earth. From a wholly detached viewpoint, both these phenomena could, perfectly correctly, be regarded simply as by-products of more fundamental processes, and not as constituting natural kinds at all. The world in itself is a continuous flow of events – which is not to say that its texture is everywhere the same. Taking some point as the start or end of some process is only non-arbitrary when seen in the light of some interest or concern.

4.5 A Surprising Ally

Support for my treatment of (most of) the special sciences can be drawn from Armstrong's account of universals (1980: vol. 2; 1989). Armstrong is a realist, but not for all properties, only for those required by basic science. Now it might be thought that this includes those in the special sciences, but I think that it does not. A real universal is one that makes a distinctive causal contribution, but non-micro entities, case by case, add nothing to the causal contribution of the micro base. Whatever reservations I may have about Armstrong's close tying of the identity of universals and properties to their causal powers, I think it is not unreasonable, in this context, to take the matter of whether a universal "does work" in its particular instances as criterial of whether a real universal is there needed. This can perhaps be reinforced by appeal to Armstrong's claim that there are no disjunctive universals (1980: 19–23; 1989: 82–4).[10] The properties of any special science not related by simple bridging laws to physics will be disjuncts – perhaps open-ended disjuncts – of more atomic universals. This reinforces the sense in which irreducible universals are not strictly necessary: the corresponding *predicates* are necessary for the schemes of explanation that constitute the special sciences, but predicates, as opposed to universals, are creatures of human thought and talk, and so presuppose the mental perspective on the subject matter.[11]

4.6 The Optionality of Non-basic Levels and the Unavoidability of Psychology

I want to take the matter further by discussing the suggestion that, if a being could understand the world in all its physical (meaning, on the level of physics) detail, but ignored the grosser levels, it would be missing out on nothing. The

purpose of the discussion is to show that amongst the special sciences only psychology could not be omitted without loss, and that this shows the essential difference of the mental from the physical.

Imagine a semi-divine being who follows everything at the level of physics, but takes no notice of any of the more macroscopic patterns of events. Because of his intelligence, he can predict the position of everything with as much accuracy as is in principle possible. Are we to say that his failure to concern himself with grosser patterns is a form of substantive ignorance, or that he merely ignores certain macro patterns that are essential to us for understanding because we cannot grasp the detail: they are, for us, a necessary shorthand and for him, not necessary at all? Someone who thought such a being was substantially ignorant might start by claiming that failure to notice patterns and operative laws constitutes ignorance. But suppose that our semi-divinity were capable of noticing these things, but found them of no interest, given his ability to do everything in terms of physics. It would be necessary to argue that the non-basic levels were, in some way, *significant in their own right*, ends in themselves. The issue is closely parallel to that of the irreducibility of teleological explanation. Supposing the truth of mechanism, do teleological explanations do extra, non-heuristic work?

The situation is at its most crucial for psychology, as is brought out by Dennett's discussion in "True Believers" (in 1987: 13–42). Dennett argues that even an omnicompetent observer who was able to predict the behavior of humans by predicting the behavior of the individual atoms that make them up would need folk psychology. He would need it if he wished to understand the utterances of humans when he talked to them, and, more fundamentally, he would need it to understand what he himself was doing. So the folk psychological level of description is ineliminable, though it carries no fundamental ontological clout. The problem with Dennett's position is that there can be no explanation of why we must adopt the folk psychological perspective. If we are all just clouds of atoms, why are we obliged to see ourselves in this particular ontologically non-basic way? It is true that we cannot see ourselves as *people* or *understand* our *actions* unless we adopt this perspective, but why see ourselves in these ways? An eliminativist would argue that it is just conceptual conservatism. But if one rejects the idea that we just happen to be hooked on this way of seeing ourselves and agree that the applicability of these categories is truly fundamental, then there is the problem of explaining why this should be so. A reductionist believes that statements on this level can be true, because they are reducible. But this fact does not explain why, amongst all possible non-basic levels of discourse, this one should be unavoidable, rather than merely available if required. It is possible to argue that the question "why should we see ourselves as persons?" answers itself, because the use of "we" already presupposes the personal perspective. But this misses the point. The behavior of the physical structures that we call "people" cannot be understood in a way that seems complete or remotely adequate without the personal perspective. Physicalistically speaking, there should be no "we" that exists at some particular level. But, even if one tries to think in a physicalistic

manner, one cannot avoid thinking that, at a certain level of complexity, there emerges something which is neither a matter of seeing or interpreting the organism in a certain way from outside (on pain of regress) nor is it just one of those levels of complexity which one might notice or ignore. There is present there, in a manner wholly different from other forms of emergent complexity (because others are either or both interpretative and ignorable) something of which it makes no sense to say one might ignore it. This is at least the seed of what Descartes expresses in the *cogito*.

The truth is that even if we were able to do all the predicting that physical omniscience would make possible, it would be impossible to restrict one's understanding of oneself to the physical terms. The Cartesian certainty that I *think* is absolute, not relative to adopting one possible but, like all the rest, optional level of discourse. Our existence on the personal level is a fundamental, not a pragmatic, fact. There is no way it can be thought of as a function of a certain way of thinking or conceptualizing: it is a basic fact, in the sense of being unavoidable in a more than pragmatic way, and it could not be thus basic if the physicalist ontology were correct.

4.7 Why Bundle Dualism Will Not Do

If what is said above is correct, the mind transcends the physical world and is, *ipso facto*, non-physical. But this does not indicate whether it is a substance or only a collection of states. I shall argue that bundle dualism will not suffice because this would make it into a complex entity and only by supposing it to be simple can we accommodate certain irresistible intuitions concerning personal identity.

There is a long tradition, dating at least from Reid, for arguing that the identity of persons over time is not a matter of convention or degree in the way that the identity of other (complex) substances is. Criticism of these arguments and of the intuitions on which they rest, running from Hume to Parfit, has left us with an inconclusive clash of intuitions. My argument does not concern identity through time, but the consequences for identity of certain counterfactuals concerning origin. It can, I hope, therefore, break the stalemate which faces the debate over diachronic identity. My claim will be that the broadly conventionalist ways, which are used to deal with problem cases through time for both persons and material objects, and which can also be employed in cases of counterfactuals concerning origin for bodies, cannot be used for similar counterfactuals concerning persons or minds.

It is nowadays respectable to maintain that individuals have essential properties, though it is somewhat less generally agreed that they have essences. Kripke's claim that a particular wooden table could not have been made of ice seems to be widely accepted, so there is at least one necessary condition for the existence of that individual table: but whether there are necessary and sufficient conditions –

i.e. an essence – as well as merely necessary conditions for it being the object it is, is more controversial (Kripke 1980: 39–53). Even granted that the table has some essential properties, it is doubtful whether it has an essence. We can scale sentences as follows:

1 This table might have been made of ice.
2 This table might have been made of a different sort of wood.
3 This table might have been made of 95 per cent of the wood it was made of and 5 per cent of some other wood.

There will come a point along the spectrum illustrated by (1) and (2) and towards (3) where the question of whether the hypothesized table would be the same as the one that actually exists has no obvious answer. It seems that the question of whether it "really" is the same one has no clear meaning: it is of, say, 75 per cent the same matter and of 25 per cent different matter. These are the only genuine facts in the case; the question of numerical identity can be decided in any convenient fashion, or left unresolved. There will thus be a penumbra of counterfactual cases where the question of whether two things would be the same is not a matter of fact.

Suppose that a given human individual had had origins different from those which he in fact had, such that whether that difference affected *who he was* was not obvious to intuition. What would count as such a case might be a matter of controversy, but there must be one. Perhaps it is unclear whether, if there had been a counterpart to Jones's body from the same egg but a different though genetically identical sperm from the same father, the person there embodied would have been Jones. Some philosophers might regard it as obvious that sameness of sperm is essential to the identity of a human body and to personal identity. In that case imagine a counterpart sperm in which some of the molecules in the sperm are different; would that be the same sperm? If one pursues the matter far enough there will be indeterminacy which will infect that of the resulting body. There must therefore be some difference such that neither natural language nor intuition tells us whether the difference alters the identity of the human body; a point, that is, where the question of whether we have the same body is not a matter of fact.

How one is to describe these cases is, in some respects, a matter of controversy. Some philosophers think one can talk of *vague identity* or *partial identity*, others think that such expressions are nonsensical. I do not have the space to discuss this issue. I am assuming, however, that questions of how one is allowed to use the concept of identity affect only the care with which one should characterize these cases, not any substantive matter of fact. There are cases of substantial overlap of constitution in which *that* fact is the only bedrock fact in the case: there is no further fact about whether they are "really" the same object. If there were then there would have to be a *haecceitas* or *thisness* belonging to and individuating each complex physical object, and this I am assuming to be implausible if not

unintelligible. (More about the conditions under which *haecceitas* can make sense will be found below.) My claim is that no similar overlap of constitution can be applied to the counterfactual identity of minds. In Geoffrey Madell's words: "But while my present body can thus have its partial counterpart in some possible world, my present consciousness cannot. Any present state of consciousness that I can imagine either is or is not mine. There is no question of degree here" (1981: xx).[12]

Why is this so? Imagine the case where we are not sure whether it would have been Jones's body – and, hence, Jones – that would have been created by the slightly modified sperm and the same egg. Can we say, as we would for an object with no consciousness, that the story *something the same, something different* is the whole story, that overlap of constitution is all there is to it? For the Jones body as such, this approach would do as well as for any other physical object. But suppose Jones, in reflective mood, asks himself "if that had happened, would I have existed?" There are at least three answers he might give to himself: (1) "I either would or would not, but I cannot tell;" (2) "There is no fact of the matter whether I would or would not have existed: it is just a mis-posed question;" (3) "In some ways, or to some degree, I would have, and in some ways, or to some degree, I would not. The creature who would have existed would have had a kind of overlap of psychic constitution with me."

The third answer parallels the response we would give in the case of bodies. But as an account of the subjective situation, it makes no sense. Call the creature that would have emerged from the slightly modified sperm "Jones2." Is the overlap suggestion that, just as, say, 85 per cent of Jones2's original body would have been identical with Jones's, about 85 per cent of his psychic life would have been Jones's? That it would have been *like* Jones's – indeed, that Jones2 might have had a psychic life 100 per cent like Jones's – makes perfect sense, but that he might have been to that degree, the same psyche – that Jones "85 per cent existed" – makes no sense. Take the case in which Jones and Jones2 have *exactly* similar lives throughout: which 85 per cent of the 100 per cent similar mental events do they share? Nor does it make sense to suggest that Jones might have participated in the whole of Jones2's psychic life, but in a rather ghostly *only 85 per cent there* manner. Clearly, the notion of overlap of numerically identical psychic parts cannot be applied in the way that overlap of actual bodily part constitution quite unproblematically can.

This might make one try the second answer. We can apply the "overlap" answer to the Jones body, but the question of whether the minds or subjects would have been the same has no clear sense. It is difficult to see why it does not. Suppose Jones found out that he had originally been one of twins, in the sense that the zygote from which he developed had divided, but that the other half had died soon afterwards. He can entertain the thought that if it had been his half that had died, he would never have existed as a conscious being, though someone would, whose life, both inner and outer, might have been very similar to his. He might feel rather guiltily grateful that it was the other half that died. It would be

strange to think that Jones is wrong to think that there is a matter of fact about this. And how is one to "manage" the transition from the case where there is a matter of fact to the case where there is not?

This only leaves us with the first option. There has to be an absolute matter of fact from the subjective point of view. But the physical examples we have considered show that when something is essentially complex, this cannot be the case. When there is constitution, degree and overlap of constitution are inevitably possible. So the mind must be simple, and this is possible only if it is something like a Cartesian substance.

4.8 Two Reflections on this Conclusion

The first reflection concerns the difference between Jones's failure to imagine his relation to the existence of Jones2, and other more traditional problems in personal identity. Unlike the other cases, Jones's is not a matter of what one might call *empathetic distance*.

Suppose that my parents had emigrated to China whilst my mother was pregnant with me, and that, shortly after my birth, both my parents had died. I was then taken in by Chinese foster parents, lived through the revolution and ended up being brought up in whatever way an alien would have been brought up in Mao's China. None of this person's post-uterine experiences would have been like mine. It seems, on the one hand, that this person would obviously have been me, and, on the other, that it is utterly unclear what kind of empathetic connection I can feel to this other "me." If I ask, like Jones, "would this have been me?," I am divided between the conviction that, as the story is told, it obviously would, and a complete inability to feel myself into the position I would then have occupied. This kind of *failure of empathy* plays an important role in many stories that are meant to throw doubt on the absoluteness of personal identity. It is important to the attempt to throw doubt on whether I am the same person as I would become in fifty years time, or whether brain damage would render me "a different person" in more than a metaphorical sense. It is also obviously something that can be a matter of degree: some differences are more empathetically imaginable than others. In all these cases our intuitions are indecisive about the effect on identity. It is an important fact that problems of empathy play no role in my argument. The twin who might have survived in my stead, or the person who would have existed if the sperm had been slightly different, could have had as exactly similar a psychic life to mine as you care to imagine. This shows the difference between the cases I have discussed and the problematic cases that involve identity through time. In those cases the idea of "similar but not quite the same" gets empirical purchase. My future self feels, in his memory, much, but not all, of what I now feel. In these cases, overlap of conscious constitution is clearly intelligible. But in the counterfactual cases, imaginative or empathetic

distance plays no essential role, and the accompanying relativity of identification gets no grip.

Secondly, I think that the argument is reinforced by the light it throws on the concept of *haecceitas*. In the case of complex physical bodies it is impossible to imagine what a *haecceitas* would consist in or how it relates to the other features of the object, and so the suggestion that there is such a thing seems to be pure mystery-mongering. By contrast, in the case of minds we do have a form of *haecceitas* which, in a sense, we all understand, namely subjectivity. It is because we intuitively understand this that we feel we can give a clear sense to the suggestion that it would, or would not, have been ourselves to which something had happened, if it had happened: and that we feel we can understand very radical counterfactuals – e.g. that I might have been an ancient Greek or even a non-human – whereas such radical counterfactuals when applied to mere bodies – e.g. that this wooden table might have been the other table in the corner or even a pyramid – makes no intuitive sense. It is possible to argue that the suggestion that my mind might have been in another body ultimately makes no sense, but it makes a prima facie sense – it seems to have content – in a way that a similar suggestion for mere bodies does not. The very fact that the counterfactuals for subjects seem to make sense exhibits something not present in the other cases, which is available to function in the role of *haecceitas*. Only with consciousness understood in a Cartesian fashion can *haecceitas* be given an empirical interpretation.

4.9 An Objection

One response sometimes made to this argument is that it is correct as an account of our *concept* of the mind, but not correct about the actual nature of the mind.[13] Reality is, so to speak, deconstructive of the concept that we have. So our conceptual scheme does commit us to something like the Cartesian conception of the mind, but we have other grounds for thinking that this is a mistake. As it stands, this is more an expression of unease than a worked out objection. I shall consider two ways of filling it out.

First, one might argue as follows. If we suppose the mind to be only a collection of mental states related by a co-consciousness relation, the phenomenology would still seem to be to us as it is in fact. The argument does not, therefore, show that the bundle theory is false, for even if the bundle theory were true, it would seem to us as if we were simple substances. It could be compared to what a "hard determinist" might say about free will, namely we cannot help but feel we have it, but the feeling is mistaken.[14] There are two problems with this argument. First, it does not help Jones to answer his question. In order to avoid answer (1) – that he either would or would not be identical with Jones2 – he would have to make sense of one of the other alternatives, and this objection gives him no

help with that. Is the suggestion that when Jones tries to imagine overlap of psychic constitution, our concepts prevent him from doing so, but, in reality, such a thing would be possible? If so, I do not think this very plausible. It seems to me to be a real fact that this makes no sense. My objections above to the other option – that there is no fact of the matter – seem also to be untouched. Secondly, the argument is question-begging. It is a moot point between the bundle theorist and the substantivalist whether there could be a co-consciousness relation that would produce an experientially united mind. My argument supports the view that experiential unity involves a simple substance and so supports the view that there is no such thing as a self-standing co-consciousness relation. So it is not proper simply to claim that it could be the same for us if the bundle theory were true, if that condition is in fact an impossible one. The analogy with free will, though illustrative of what the objector is driving at, does little to show that he is correct. First, the coherence of the hard determinist's position is controversial. Secondly, the determinist can give a rationale for why we *must* feel free in terms of the conceptual impossibility of replacing one's own practice of *deciding* by one of merely *predicting* one's own behavior. There seems to be no parallel explanation of why it seems all or nothing for counterfactual identity. This is especially mysterious given that it can seem be a matter of degree in cases that turn on empathetic distance.[15]

There is a completely different way of filling the objection out. It concerns my use of counterfactuals. Counterfactuals are a controversial matter and I make no attempt to discuss them. I blatantly assume the falsehood of Lewis's counterpart analysis, for if Jones's question whether he would exist only enquired whether there would be a counterpart which possessed states very like his own, then there would be no phenomenological problem. All counterparts are strictly different objects. However, I am quite happy, along with almost all other philosophers, simply to deny Lewis's theory. But it is not from this source that the challenge comes, but from someone who takes a non-realist attitude to counterfactuals. There is an empiricist tradition which denies truth values to counterfactuals and says that they express policies or attitudes. There will be no truth about what would have happened if the relevant sperm had been slightly different.

It is not possible to get deeply engaged in a discussion of counterfactuals here. I would make two points. First, most philosophers do accept a realist account of counterfactuals – the anti-realist view is not very plausible – and the argument would go through for them. Secondly, the anti-realist approach has a weaker and a stronger form. The weaker version simply denies truth value to counterfactuals: there is no fact of the matter about whether it is *a* or *b* that would have happened if *C* had obtained. *C* could have obtained and, if it had, either *a* or *b* (or something else) would have occurred: there is just no truth from the perspective of the actual world about which it would have been. This does not affect my argument at all, which only requires that the only options about what might have happened are all or nothing, not that there is a fact about which. The stronger version says that the whole notion of *might have been otherwise* is a projection of

our mode of thought – of our ability to imagine things – not something that obtains in reality. This is not to say – as it might seem – that the actual world is necessary (because there is nothing else that might have been) but only that all these modal categories are mere projections. Even if we accepted this – which I do not recommend – it would not entirely deflate the argument. It would still show something interesting about the nature of mind, namely that it made no sense to treat it in the same way as bodies within the logical space of possibility that we create by projection. The fact that we create that space does not imply that what we express within it does not reflect real differences between the objects about which we are talking.

4.10 Conclusion

My arguments in this chapter have been in a Cartesian spirit. First, in sections 4.2–4.6 I argued that the thinking subject has to transcend the physical world about which (among other things) it thinks. Only if a strong reductionism were true could its thinking be part of that physical world. Then, in sections 4.7–4.9, I argued that the thinking subject has to be a simple substance, on pain of entertaining incoherent counterfactuals. These arguments complement each other, but they are logically independent and the second can establish its conclusion on its own.

Notes

1 I do not enter further into a fuller discussion of these properties here, for that belongs principally to an examination of the problems for materialism. For a fuller description of these properties and a brief outline of the strategies that modern materialists have employed to cope with them, see Robinson (1999).

2 Descartes' *Sixth Meditation* is the *locus classicus* for substance dualism. Modern defenses of the theory can be found in Popper and Eccles (1977), Swinburne (1986), and Foster (1991). Hume develops his theory in the *Treatise* (Bk I, Part iv, Section 6) and expresses his dissatisfaction with what he has said in the Appendix to the *Treatise*. There are several modern philosophers who account for the unity of the mind in terms of the relations between mental events, and so could be said to have a bundle theory, but they do not tend to be dualists. Parfit (1971; 1984) is a materialist and Dainton (2000) is neutral on ontological questions.

3 It might be thought that the attribute theory already has an account of the unity of the mind, in terms of the dependence of all the elements in a given bundle on the same brain. But, though this may be a causal explanation, it is not an analysis, of unity. Mere dependence on the same brain does not conceptually guarantee unity of consciousness. See Foster (1968) in reply to Ayer (1963).

4 For the doubts, see the Appendix to the *Treatise*.

5 I discuss embodiment – though not specifically the problem of interaction – in Robinson (1989).
6 Examples of translation reductionism are Hempel (1980) and Carnap (1934).
7 The classic source for this is Ernest Nagel (1961).
8 The withdrawal from genuine reductionism in psychology, then, began when Skinner accepted that a stimulus-response model was inadequate, and developed the notion of operant conditioning. Whereas the former required only mechanistic causal concepts, the latter is irreducibly teleological. The behavior of the rat which is learning how to get the food pellet may have a mechanical description on a lower level, but the understanding of it as operant conditioning has to be teleological, for it concerns what the rat is trying to achieve, or the point of its behavior. Furthermore, the behaviorist is prevented, by his own principled disinterest in what happens inside, from having views about the nature of the process in which the learning is realized. This brings out the ambiguity of the concept of reduction when applied to the philosophy of mind. Its central concern is to eliminate "the ghost in the machine" – that is, anything irreducibly private or subjective. This form of reduction is entirely irrelevant to any of the physical sciences. The second element is the elimination or analysis away of concepts of a kind that have no place in a purely physical science. Operant conditioning meets the first objective but not the second. It is the brunt of the argument of this part of my chapter that, contra Fodor, the second objective is as essential to the physicalist as the first.
9 These are the much discussed qualia objections to physicalism. See, for example, Jackson (1982), Robinson (1993).
10 Armstrong's acceptance of conjunctive universals also reinforces the intuition that strong reduction preserves full realism for the special sciences. Water is a conjunction of instances of the universals of hydrogen-ness and oxygen-ness in a certain spatial arrangement. These, in their turn, are conjunctions of more atomic universals.
11 It follows from this, of course, that if psychology (which includes not only the science, but our ordinary mentalistic concepts) is not reducible in a strong sense, its "properties" are only predicates and its subject matter is in part created by an act of the mind – the mind not being present until that act has been performed. Armstrong's theory becomes less different from Dennett's interpretative theory, with the attendant threat of regress, than was the intention.
12 Madell's book is an excellent treatment of the topics I discuss in this section.
13 This objection has been made to me, on different occasions, by Simon Blackburn, Derek Parfit, and Katalin Farkas. It is worth noting that this objection involves a major concession. If the argument I have presented shows that we are committed by the way we think of ourselves to a Cartesian concept of the self, this was not in virtue of some easily revisable definition. The argument was not a derivation of logical consequences from some necessary and sufficient conditions for being a subject, leaving the option of altering those conditions. It proceeded on the basis of what was conceivable for a conscious subject. The associated concept of the self must be unavoidable in a "Kantian" manner. The suggestion that it is mistaken is, therefore, a form of skeptical nihilism, which we can only live through by pretending to ignore.
14 I owe the comparison with free will to Katalin Farkas.
15 There is a more complicated version of the argument presented in sections 4.6–4.9, which would resist the objection. I believe that it can be argued that vague predicates

are never ontologically basic and can, in principle, be eliminated. Amongst these will be the notion of identity under counterfactual circumstances for physical bodies of all kinds. There is no real factual difference between an assertion that some physical body would have existed if such and such had been different, and an assertion that there would have been a "counterpart" body of a similar kind under those circumstances. This applies even if the counterfactual change does not directly involve the object in question. But this treatment is wholly unacceptable for subjects. Suppose that, contrary to fact, someone had coughed on the other side of the world just before you were conceived. On the principle that applies to bodies, there is no factual difference between the proposition that *you* would still have come into existence, and the proposition that someone with the same qualities as you would have. As the twin example shows, this difference is real and not eliminable. The full version of this argument is not in print, but for discussion of some of the relevant issues concerning vagueness, see Robinson (2001).

References

Armstrong, D. M. (1980). *Universals and Scientific Realism* (2 vols). Cambridge: Cambridge University Press.

—— (1989). *Universals: An Opinionated Introduction*. Boulder, CO: Westview Press.

Ayer, A. J. (1963). "The Concept of a Person." In *The Concept of a Person and other Essays*. London: Macmillan: 82–128.

Carnap, R. (1934). *The Unity of Science*. London: Kegan Paul.

Dainton, B. (2000). *Stream of Consciousness*. London: Routledge.

Dennett, D. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.

—— (1991). "Real Patterns." *Journal of Philosophy*, 89: 27–51.

Descartes, R. (1984–5). *The Philosophical Writings of Descartes*, trans. J. Cottingham, R. Stoothof, and D. Murdoch (2 vols). Cambridge: Cambridge University Press.

Fodor, J. (1974). "Special Sciences or the Disunity of Science as a Working Hypothesis." *Synthese*, 28: 77–115.

Foster, J. (1968). "Psychophysical Causal Relations." *American Philosophical Quarterly*, 5.

—— (1991). *The Immaterial Self*. London: Routledge.

Hempel, C. G. (1980). "The Logical Analysis of Psychology." In N. Block (ed.), *Readings in Philosophy of Psychology*, vol. 1. London: Methuen: 14–23. (Originally published in French in 1935.)

Hume, D. (1978). *A Treatise on Human Nature*, ed. P. H. Nidditch. Oxford: Clarendon Press.

Jackson, F. (1982). "Epiphenomenal Qualia." *Philosophical Quarterly*, 32: 127–36.

Kripke, S. (1980). *Naming and Necessity*. Oxford: Blackwell.

Madell, G. (1981). *The Identity of the Self*. Edinburgh: Edinburgh University Press.

Nagel, E. (1961). *The Structure of Science*. London: Routledge and Kegan Paul.

Parfit, D. (1971). "Personal Identity." *Philosophical Review*, 80: 3–27.

—— (1984). *Reasons and Persons*. Oxford: Clarendon Press.

Popper, K. R. and Eccles, J. C. (1977). *The Self and its Brain*. Berlin: Springer International.

Robinson, H. (1989). "A Dualist Account of Embodiment." In Smythies and Beloff (1989): 43–57.

—— (1993). "The Anti-materialist Strategy and the 'Knowledge Argument'." In H. Robinson (ed.), *Objections to Physicalism*. Oxford: Clarendon Press: 159–83.
—— (1999). "Materialism and the Mind–Body Problem." In E. Graig (ed.), *The Routledge Encyclopaedia of Philosophy*. London: Routledge.
—— (2001). "Vagueness, Realism, Language and Thought." In T. Horgan and M. Potric (eds.), *Essays on Vagueness*. Oxford: Oxford University Press.
Smythies, J. R. and Beloff, J. (eds.) (1989). *The Case for Dualism*. Charlottesville: University of Virginia Press.
Swinburne, R. (1986). *The Evolution of the Soul*. Oxford: Clarendon Press.