Dualism

First published Tue Aug 19, 2003; substantive revision Mon Feb 29, 2016

This entry concerns dualism in the philosophy of mind. The term ‘dualism’ has a variety of uses in the history of thought. In general, the idea is that, for some particular domain, there are two fundamental kinds or categories of things or principles. In theology, for example a ‘dualist’ is someone who believes that Good and Evil—or God and the Devil—are independent and more or less equal forces in the world. Dualism contrasts with monism, which is the theory that there is only one fundamental kind, category of thing or principle; and, rather less commonly, with pluralism, which is the view that there are many kinds or categories. In the philosophy of mind, dualism is the theory that the mental and the physical—or mind and body or mind and brain—are, in some sense, radically different kinds of thing. Because common sense tells us that there are physical bodies, and because there is intellectual pressure towards producing a unified view of the world, one could say that materialist monism is the ‘default option’. Discussion about dualism, therefore, tends to start from the assumption of the reality of the physical world, and then to consider arguments for why the mind cannot be treated as simply part of that world.

- Other Internet Resources
- Related Entries

---

1. The Mind-Body Problem and the History of Dualism

1.1 The Mind-Body Problem

The mind-body problem is the problem: what is the relationship between mind and body? Or alternatively: what is the relationship between mental properties and physical properties?

Humans have (or seem to have) both physical properties and mental properties. People have (or seem to have)the sort of properties attributed in the physical sciences. These physical properties include size, weight, shape, colour, motion through space and time, etc. But they also have (or seem to have) mental properties, which we do not attribute to typical physical objects These properties involve consciousness (including perceptual experience, emotional experience, and much else), intentionality (including beliefs, desires, and much else), and they are possessed by a subject or a self.

Physical properties are public, in the sense that they are, in principle, equally observable by anyone. Some physical properties—like those of an electron—are not directly observable at all, but they are equally available to all, to the same degree, with scientific equipment and techniques. The same is not true of mental properties. I may be able to tell that you are in pain by your behaviour, but only you can feel it directly. Similarly, you just know how something looks to you, and I can only surmise. Conscious mental events are private to the subject, who has a privileged access to them of a kind no-one has to the physical.

The mind-body problem concerns the relationship between these two sets of properties. The mind-body problem breaks down into a number of components.

1. The ontological question: what are mental states and what are physical states? Is one class a subclass of the other, so that all mental states are physical, or vice versa? Or are mental states and physical states entirely distinct?
2. The causal question: do physical states influence mental states? Do mental states influence physical states? If so, how?

Different aspects of the mind-body problem arise for different aspects of the mental, such as consciousness, intentionality, the self.

3. The problem of consciousness: what is consciousness? How is it related to the brain and the body?
4. The problem of intentionality: what is intentionality? How is it related to the brain and the body?
5. The problem of the self: what is the self? How is it related to the brain and the body?

Other aspects of the mind-body problem arise for aspects of the physical. For example:

6. The problem of embodiment: what is it for the mind to be housed in a body? What is it for a body to belong to a particular subject?

The seemingly intractable nature of these problems have given rise to many different philosophical views.

Materialist views say that, despite appearances to the contrary, mental states are just physical states. Behaviourism, functionalism, mind-brain identity theory and the computational theory of mind are examples of how materialists attempt to explain how this can be so. The most common factor in such theories is the attempt to explicate the nature of mind and consciousness in terms of their ability to directly or indirectly modify behaviour, but there are versions of materialism that try to tie the mental to the physical without explicitly explaining the mental in terms of its behaviour-modifying role. The latter are often grouped together under the label 'non-reductive physicalism', though this label is itself rendered elusive because of the controversial nature of the term 'reduction'.

Idealist views say that physical states are really mental. This is because the physical world is an *empirical* world and, as such, it is the intersubjective product of our collective experience.

Dualist views (the subject of this entry) say that the mental and the physical are both real and neither can be assimilated to the other. For the various forms that dualism can take and the associated problems, see below.

In sum, we can say that there is a mind-body problem because both consciousness and thought, broadly construed, seem very different from anything physical and there is no convincing consensus on how to build a satisfactorily unified picture of creatures possessed of both a mind and a body.

Other entries which concern aspects of the mind-body problem include (among many others): behaviorism, consciousness, eliminative materialism, epiphenomenalism, functionalism, identity theory, intentionality, mental causation, neutral monism, and physicalism.

1.2 History of dualism

In dualism, 'mind' is contrasted with 'body', but at different times, different aspects of the mind have been the centre of attention. In the classical and mediaeval periods, it was the intellect that was thought to be most obviously resistant to a materialistic account: from Descartes on, the main stumbling block to materialist monism was supposed to be 'consciousness', of which phenomenal consciousness or sensation came to be considered as the paradigm instance.

The classical emphasis originates in Plato's *Phaedo*. Plato believed that the true substances are not physical bodies, which are ephemeral, but the eternal Forms of which bodies are imperfect copies. These Forms not only make the world possible, they also make it intelligible, because they perform the role of universals, or what Frege called 'concepts'. It is

their connection with intelligibility that is relevant to the philosophy of mind. Because Forms are the grounds of intelligibility, they are what the intellect must grasp in the process of understanding. In *Phaedo* Plato presents a variety of arguments for the immortality of the soul, but the one that is relevant for our purposes is that the intellect is immaterial because Forms are immaterial and intellect must have an affinity with the Forms it apprehends (78b4–84b8). This affinity is so strong that the soul strives to leave the body in which it is imprisoned and to dwell in the realm of Forms. It may take many reincarnations before this is achieved. Plato's dualism is not, therefore, simply a doctrine in the philosophy of mind, but an integral part of his whole metaphysics.

One problem with Plato's dualism was that, though he speaks of the soul as imprisoned in the body, there is no clear account of what binds a particular soul to a particular body. Their difference in nature makes the union a mystery.

Aristotle did not believe in Platonic Forms, existing independently of their instances. Aristotelian forms (the capital ‘F’ has disappeared with their standing as autonomous entities) are the natures and properties of things and exist embodied in those things. This enabled Aristotle to explain the union of body and soul by saying that the soul is the form of the body. This means that a particular person's soul is no more than his nature as a human being. Because this seems to make the soul into a property of the body, it led many interpreters, both ancient and modern, to interpret his theory as materialistic. The interpretation of Aristotle's philosophy of mind—and, indeed, of his whole doctrine of form—remains as live an issue today as it was immediately after his death (Robinson 1983 and 1991; Nussbaum 1984; Rorty and Nussbaum, eds, 1992). Nevertheless, the text makes it clear that Aristotle believed that the intellect, though part of the soul, differs from other faculties in not having a bodily organ. His argument for this constitutes a more tightly argued case than Plato's for the immateriality of thought and, hence, for a kind of dualism. He argued that the intellect must be immaterial because if it were material it could not receive all forms. Just as the eye, because of its particular physical nature, is sensitive to light but not to sound, and the ear to sound and not to light, so, if the intellect were in a physical organ it could be sensitive only to a restricted range of physical things; but this is not the case, for we can think about any kind of material object (*De Anima* III,4; 429a10–b9). As it does not have a material organ, its activity must be essentially immaterial.

It is common for modern Aristotelians, who otherwise have a high view of Aristotle's relevance to modern philosophy, to treat this argument as being of purely historical interest, and not essential to Aristotle's system as a whole. They emphasize that he was not a ‘Cartesian’ dualist, because the intellect is an aspect of the soul and the soul is the form of the body, not a separate substance. Kenny (1989) argues that Aristotle's theory of mind as form gives him an account similar to Ryle (1949), for it makes the soul equivalent to the dispositions possessed by a living body. This ‘anti-Cartesian’ approach to Aristotle arguably ignores the fact that, for Aristotle, the form *is* the substance.

These issues might seem to be of purely historical interest. But we shall see in below, in section 4.5, that this is not so.

The identification of form and substance is a feature of Aristotle's system that Aquinas effectively exploits in this context, identifying soul, intellect and form, and treating them as a substance. (See, for example, Aquinas (1912), Part I, questions 75 and 76.) But though the form (and, hence, the intellect with which it is identical) are the substance of the human person, they are not the person itself. Aquinas says that when one addresses prayers to a saint—other than the Blessed Virgin Mary, who is believed to retain her body in heaven and is, therefore, always a complete person—one should say, not, for example, ‘Saint Peter pray for us', but ‘soul of Saint Peter pray for us'. The soul, though an immaterial substance, is the person only when united with its body. Without the body, those aspects of its personal memory that depend on images (which are held to be corporeal) will be lost.(See Aquinas (1912), Part I, question 89.)

The more modern versions of dualism have their origin in Descartes' *Meditations*, and in the debate that was consequent upon Descartes' theory. Descartes was a *substance dualist*. He believed that there were two kinds of substance: matter, of which the essential property is that it is spatially extended; and mind, of which the essential property is that it thinks. Descartes' conception of the relation between mind and body was quite different from that held in the Aristotelian tradition. For Aristotle, there is no exact science of matter. How matter behaves is essentially affected by the form that is in it. You cannot combine just any matter with any form—you cannot make a knife out of butter, nor a human being out of paper—so the nature of the matter is a necessary condition for the nature of the substance. But the nature of the substance does not follow from the nature of its matter alone: there is no ‘bottom up’ account of substances. Matter is a determinable made determinate by form. This was how Aristotle thought that he was able to explain the connection of soul to body: a particular soul exists as the organizing principle in a particular parcel of matter.

The belief in the relative indeterminacy of matter is one reason for Aristotle's rejection of atomism. If matter is atomic, then it is already a collection of determinate objects in its own right, and it becomes natural to regard the properties of macroscopic substances as mere summations of the natures of the atoms.

Although, unlike most of his fashionable contemporaries and immediate successors, Descartes was not an atomist, he was, like the others, a mechanist about the properties of matter. Bodies are machines that work according to their own laws. Except where there are minds interfering with it, matter proceeds deterministically, in its own right. Where there are minds requiring to influence bodies, they must work by ‘pulling levers' in a piece of machinery that already has its own laws of operation. This raises the question of where those ‘levers' are in the body. Descartes opted for the pineal gland, mainly because it is not duplicated on both sides of the brain, so it is a candidate for having a unique, unifying function.

The main uncertainty that faced Descartes and his contemporaries, however, was not *where* interaction took place, but *how* two things so different as thought and extension could interact at all. This would be particularly mysterious if one had an *impact* view of causal interaction, as would anyone influenced by atomism, for whom the paradigm of causation is like two billiard balls cannoning off one another.

Various of Descartes' disciples, such as Arnold Geulincx and Nicholas Malebranche, concluded that all mind-body interactions required the direct intervention of God. The appropriate states of mind and body were only the *occasions* for such intervention, not real causes. Now it would be convenient to think that occasionalists held that all causation was natural *except* for that between mind and body. In fact they generalized their conclusion and treated all causation as directly dependent on God. Why this was so, we cannot discuss here.

Descartes' conception of a dualism of *substances* came under attack from the more radical empiricists, who found it difficult to attach sense to the concept of substance at all. Locke, as a moderate empiricist, accepted that there were both material and immaterial substances. Berkeley famously rejected material substance, because he rejected all existence outside the mind. In his early *Notebooks*, he toyed with the idea of rejecting immaterial substance, because we could have no idea of it, and reducing the self to a collection of the 'ideas' that constituted its contents. Finally, he decided that the self, conceived as something over and above the ideas of which it was aware, was essential for an adequate understanding of the human person. Although the self and its acts are not presented to consciousness as *objects* of awareness, we are obliquely aware of them simply by dint of being active subjects. Hume rejected such claims, and proclaimed the self to be nothing more than a concatenation of its ephemeral contents.

In fact, Hume criticised the whole conception of substance for lacking in empirical content: when you search for the owner of the properties that make up a substance, you find nothing but further properties. Consequently, the mind is, he claimed, nothing but a 'bundle' or 'heap' of impressions and ideas—that is, of particular mental states or events, without an owner. This position has been labelled *bundle dualism*, and it is a special case of a general *bundle theory of substance*, according to which objects in general are just organised collections of properties. The problem for the Humean is to explain what binds the elements in the bundle together. This is an issue for any kind of substance, but for material bodies the solution seems fairly straightforward: the unity of a physical bundle is constituted by some form of causal interaction between the elements in the bundle. For the mind, mere causal connection is not enough; some further relation of co-consciousness is required. We shall see in 5.2.1 that it is problematic whether one can treat such a relation as more primitive than the notion of belonging to a subject.

One should note the following about Hume's theory. His bundle theory is a theory about the nature of the unity of the mind. As a theory about this unity, it is not necessarily dualist. Parfit (1970, 1984) and Shoemaker (1984, ch. 2), for example, accept it as physicalists. In general, physicalists will accept it unless they wish to ascribe the unity to the brain or the organism as a whole. Before the bundle theory can be dualist one must accept *property* dualism, for more about which, see the next section.

A crisis in the history of dualism came, however, with the growing popularity of *mechanism* in science in the nineteenth century. According to the mechanist, the world is, as it would now be expressed, 'closed under physics'. This means that everything that happens follows from and is in accord with the laws of physics. There is, therefore, no scope for interference in the physical world by the mind in the way that interactionism seems to require. According to the

mechanist, the conscious mind is an *epiphenomenon* (a notion given general currency by T. H. Huxley 1893): that is, it is a by-product of the physical system which has no influence back on it. In this way, the facts of consciousness are acknowledged but the integrity of physical science is preserved. However, many philosophers found it implausible to claim such things as the following; the pain that I have when you hit me, the visual sensations I have when I see the ferocious lion bearing down on me or the conscious sense of understanding I have when I hear your argument—all have nothing directly to do with the way I respond. It is very largely due to the need to avoid this counterintuitiveness that we owe the concern of twentieth century philosophy to devise a plausible form of materialist monism. But, although dualism has been out of fashion in psychology since the advent of behaviourism (Watson 1913) and in philosophy since Ryle (1949), the argument is by no means over. Some distinguished neurologists, such as Sherrington (1940) and Eccles (Popper and Eccles 1977) have continued to defend dualism as the only theory that can preserve the data of consciousness. Amongst mainstream philosophers, discontent with physicalism led to a modest revival of property dualism in the last decade of the twentieth century. At least some of the reasons for this should become clear below.

2. Varieties of Dualism: Ontology

There are various ways of dividing up kinds of dualism. One natural way is in terms of what sorts of things one chooses to be dualistic about. The most common categories lighted upon for these purposes are *substance* and *property*, giving one *substance dualism* and *property dualism*. There is, however, an important third category, namely *predicate dualism*. As this last is the weakest theory, in the sense that it claims least, I shall begin by characterizing it.

2.1 Predicate dualism

Predicate dualism is the theory that psychological or mentalistic predicates are (a) essential for a full description of the world and (b) are not reducible to physicalistic predicates. For a mental predicate to be reducible, there would be bridging laws connecting types of psychological states to types of physical ones in such a way that the use of the mental predicate carried no information that could not be expressed without it. An example of what we believe to be a true type reduction outside psychology is the case of water, where water is always H_2O: something is water if and only if it is H_2O. If one were to replace the word ‘water’ by ‘H_2O’, it is plausible to say that one could convey all the same information. But the terms in many of the special sciences (that is, any science except physics itself) are not reducible in this way. Not every *hurricane* or every *infectious disease*, let alone every *devaluation of the currency* or every *coup d'etat* has the same constitutive structure. These states are defined more by *what they do* than by their *composition or structure.* Their names are classified as *functional terms* rather than *natural kind terms*. It goes with this that such kinds of state are *multiply realizable*; that is, they may be constituted by different kinds of physical structures under different circumstances. Because of this, unlike in the case of water and H_2O, one could not replace these terms by some more basic physical description and still convey the same information. There is no particular description, using the language of physics or chemistry, that would do the work of the word ‘hurricane’, in the way that ‘H_2O’

would do the work of ‘water’. It is widely agreed that many, if not all, psychological states are similarly irreducible, and so psychological predicates are not reducible to physical descriptions and one has predicate dualism. (The classic source for irreducibility in the special sciences in general is Fodor (1974), and for irreducibility in the philosophy of mind, Davidson (1971).)

2.2 Property Dualism

Whereas predicate dualism says that there are two essentially different kinds of *predicates* in our *language*, property dualism says that there are two essentially different kinds of *property* out in the world. Property dualism can be seen as a step stronger than predicate dualism. Although the predicate ‘hurricane’ is not equivalent to any single description using the language of physics, we believe that each individual hurricane is *nothing but* a collection of physical atoms behaving in a certain way: one need have no more than the physical atoms, with their normal physical properties, following normal physical laws, for there to be a hurricane. One might say that we need more than the *language* of physics to describe and explain the weather, but we do not need more than its *ontology*. There is *token identity* between each individual hurricane and a mass of atoms, even if there is no *type identity* between hurricanes as kinds and some particular structure of atoms as a kind. Genuine property dualism occurs when, even at the individual level, the ontology of physics is not sufficient to constitute what is there. The irreducible language is not just another way of describing what there is, it requires that there be something more there than was allowed for in the initial ontology. Until the early part of the twentieth century, it was common to think that biological phenomena (‘life’) required property dualism (an irreducible ‘vital force’), but nowadays the special physical sciences other than psychology are generally thought to involve only predicate dualism. In the case of mind, property dualism is defended by those who argue that the qualitative nature of consciousness is not merely another way of categorizing states of the brain or of behaviour, but a genuinely emergent phenomenon.

2.3 Substance Dualism

There are two important concepts deployed in this notion. One is that of *substance*, the other is the *dualism* of these substances. A substance is characterized by its properties, but, according to those who believe in substances, it is more than the collection of the properties it possesses, it is *the thing which* possesses them. So the mind is not just a collection of thoughts, but is *that which* thinks, an immaterial substance over and above its immaterial states. Properties are the properties of *objects*. If one is a property dualist, one may wonder what kinds of objects possess the irreducible or immaterial properties in which one believes. One can use a neutral expression and attribute them to *persons*, but, until one has an account of *person*, this is not explanatory. One might attribute them to human beings *qua* animals, or to the brains of these animals. Then one will be holding that these immaterial properties are possessed by what is otherwise a purely material thing. But one may also think that not only mental states are immaterial, but that the subject that possesses them must also be immaterial. Then one will be a dualist about *that to which mental states and properties belong* as well about the properties themselves. Now one might try to think of these subjects as just bundles of the immaterial states. This is Hume's view. But if one thinks

that the owner of these states is something quite over and above the states themselves, and is immaterial, as they are, one will be a *substance dualist*.

Substance dualism is also often dubbed ‘Cartesian dualism’, but some substance dualists are keen to distinguish their theories from Descartes's. E. J. Lowe, for example, is a substance dualist, in the following sense. He holds that a normal human being involves two substances, one a body and the other a person. The latter is not, however, a purely mental substance that can be defined in terms of thought or consciousness alone, as Descartes claimed. But persons and their bodies have different identity conditions and are both substances, so there are two substances essentially involved in a human being, hence this is a form of substance dualism. Lowe (2006) claims that his theory is close to P. F. Strawson's (1959), whilst admitting that Strawson would not have called it substance dualism.

3. Varieties of Dualism: Interaction

If mind and body are different realms, in the way required by either property or substance dualism, then there arises the question of how they are related. Common sense tells us that they interact: thoughts and feelings are at least sometimes caused by bodily events and at least sometimes themselves give rise to bodily responses. I shall now consider briefly the problems for interactionism, and its main rivals, epiphenomenalism and parallelism.

3.1 Interactionism

Interactionism is the view that mind and body—or mental events and physical events—causally influence each other. That this is so is one of our common-sense beliefs, because it appears to be a feature of everyday experience. The physical world influences my experience through my senses, and I often react behaviourally to those experiences. My thinking, too, influences my speech and my actions. There is, therefore, a massive natural prejudice in favour of interactionism. It has been claimed, however, that it faces serious problems (some of which were anticipated in section 1).

The simplest objection to interaction is that, in so far as mental properties, states or substances are of radically different kinds from each other, they lack that communality necessary for interaction. It is generally agreed that, in its most naive form, this objection to interactionism rests on a ‘billiard ball’ picture of causation: if all causation is by impact, how can the material and the immaterial impact upon each other? But if causation is either by a more ethereal force or energy or only a matter of constant conjunction, there would appear to be no problem in principle with the idea of interaction of mind and body.

Even if there is no objection in principle, there appears to be a conflict between interactionism and some basic principles of physical science. For example, if causal power was flowing in and out of the physical system, energy would not be conserved, and the conservation of energy is a fundamental scientific law. Various responses have been made to this. One suggestion is that it might be possible for mind to influence the *distribution* of energy, without altering its quantity. (See Averill and Keating 1981). Another response is to

challenge the relevance of the conservation principle in this context. The conservation principle states that ‘in a causally isolated system the total amount of energy will remain constant’. Whereas ‘[t]he interactionist denies…that the human body is an isolated system’, so the principle is irrelevant (Larmer (1986), 282: this article presents a good brief survey of the options).

Robins Collins (2011) has claimed that the appeal to conservation by opponents of interactionism is something of a red herring because conservation principles are not ubiquitous in physics. He argues that energy is not conserved in general relativity, in quantum theory, or in the universe taken as a whole. Why then, should we insist on it in mind-brain interaction?

Most discussion of interactionism takes place in the context of the assumption that it is incompatible with the world's being ‘closed under physics’. This is a very natural assumption, but it is not justified if causal overdetermination of behaviour is possible. There could then be a complete physical cause of behaviour, and a mental one. The strongest intuitive objection against overdetermination is clearly stated by Mills (1996: 112), who is himself a defender of overdetermination.

For *X* to be a cause of *Y*, *X* must contribute something to *Y*. The only way a purely mental event could contribute to a purely physical one would be to contribute some feature not already determined by a purely physical event. But if physical closure is true, there is no feature of the purely physical effect that is not contributed by the purely physical cause. Hence interactionism violates physical closure after all.

Mills says that this argument is invalid, because a physical event can have features not explained by the event which is its sufficient cause. For example, “the rock's hitting the window is causally sufficient for the window's breaking, and the window's breaking has the feature of being the third window-breaking in the house this year; but the facts about prior window-breakings, rather than the rock's hitting the window, are what cause this window-breaking to have this feature.”

The opponent of overdetermination could perhaps reply that his principle applies, not to every feature of events, but to a subgroup—say, intrinsic features, not merely relational or comparative ones. It is this kind of feature that the mental event would have to cause, but physical closure leaves no room for this. These matters are still controversial.

The problem with closure of physics may be radically altered if physical laws are indeterministic, as quantum theory seems to assert. If physical laws are deterministic, then any interference from outside would lead to a breach of those laws. But if they are indeterministic, might not interference produce a result that has a probability greater than zero, and so be consistent with the laws? This way, one might have interaction yet preserve a kind of nomological closure, in the sense that no laws are infringed. Because it involves assessing the significance and consequences of quantum theory, this is a difficult matter for the non-physicist to assess. Some argue that indeterminacy manifests itself only on the subatomic level, being cancelled out by the time one reaches even very tiny macroscopic objects: and human behaviour is a macroscopic phenomenon. Others argue that the

structure of the brain is so finely tuned that minute variations could have macroscopic effects, rather in the way that, according to ‘chaos theory’, the flapping of a butterfly's wings in China might affect the weather in New York. (For discussion of this, see Eccles (1980), (1987), and Popper and Eccles (1977).) Still others argue that quantum indeterminacy manifests itself directly at a high level, when acts of observation collapse the wave function, suggesting that the mind may play a direct role in affecting the state of the world (Hodgson 1988; Stapp 1993).

3.2 Epiphenomenalism

If the reality of property dualism is not to be denied, but the problem of how the immaterial is to affect the material is to be avoided, then epiphenomenalism may seem to be the answer. According to this theory, mental events are caused by physical events, but have no causal influence on the physical. I have introduced this theory as if its point were to avoid the problem of how two different categories of thing might interact. In fact, it is, at best, an incomplete solution to this problem. If it is mysterious how the non-physical can have it in its nature to influence the physical, it ought to be equally mysterious how the physical can have it in its nature to produce something non-physical. But that this latter is what occurs is an essential claim of epiphenomenalism. (For development of this point, see Green (2003), 149–51). In fact, epiphenomenalism is more effective as a way of saving the autonomy of the physical (the world as ‘closed under physics') than as a contribution to avoiding the need for the physical and non-physical to have causal commerce.

There are at least three serious problems for epiphenomenalism. First, as I indicated in *section 1*, it is profoundly counterintuitive. What could be more apparent than that it is the pain that I feel that makes me cry, or the visual experience of the boulder rolling towards me that makes me run away? At least one can say that epiphenomenalism is a fall-back position: it tends to be adopted because other options are held to be unacceptable.

The second problem is that, if mental states do nothing, there is no reason why they should have evolved. This objection ties in with the first: the intuition there was that conscious states clearly modify our behaviour in certain ways, such as avoiding danger, and it is plain that they are very useful from an evolutionary perspective.

Frank Jackson (1982) replies to this objection by saying that it is the brain state associated with pain that evolves for this reason: the sensation is a by-product. Evolution is full of useless or even harmful by-products. For example, polar bears have evolved thick coats to keep them warm, even though this has the damaging side effect that they are heavy to carry. Jackson's point is true in general, but does not seem to apply very happily to the case of mind. The heaviness of the polar bear's coat follows directly from those properties and laws which make it warm: one could not, in any simple way, have one without the other. But with mental states, dualistically conceived, the situation is quite the opposite. The laws of physical nature which, the mechanist says, make brain states cause behaviour, in no way explain why brain states should give rise to conscious ones. The laws linking mind and brain are what Feigl (1958) calls *nomological danglers*, that is, brute facts added onto the body of

integrated physical law. Why there should have been by-products of *that* kind seems to have no evolutionary explanation.

The third problem concerns the rationality of belief in epiphenomenalism, *via* its effect on the problem of other minds. It is natural to say that I know that I have mental states because I experience them directly. But how can I justify my belief that others have them? The simple version of the 'argument from analogy' says that I can extrapolate from my own case. I know that certain of my mental states are correlated with certain pieces of behaviour, and so I infer that similar behaviour in others is also accompanied by similar mental states. Many hold that this is a weak argument because it is induction from one instance, namely, my own. The argument is stronger if it is not a simple induction but an 'argument to the best explanation'. I seem to know from my own case that mental events can be the explanation of behaviour, and I know of no other candidate explanation for typical human behaviour, so I postulate the same explanation for the behaviour of others. But if epiphenomenalism is true, my mental states do not explain my behaviour and there is a physical explanation for the behaviour of others. It is explanatorily redundant to postulate such states for others. I know, by introspection, that I have them, but is it not just as likely that I alone am subject to this quirk of nature, rather than that everyone is?

For more detailed treatment and further reading on this topic, see the entry epiphenomenalism.

3.3 Parallelism

The epiphenomenalist wishes to preserve the integrity of physical science and the physical world, and appends those mental features that he cannot reduce. The parallelist preserves both realms intact, but denies all causal interaction between them. They run in harmony with each other, but not because their mutual influence keeps each other in line. That they should behave *as if* they were interacting would seem to be a bizarre coincidence. This is why parallelism has tended to be adopted only by those—like Leibniz—who believe in a pre-established harmony, set in place by God. The progression of thought can be seen as follows. Descartes believes in a more or less natural form of interaction between immaterial mind and material body. Malebranche thought that this was impossible naturally, and so required God to intervene specifically on each occasion on which interaction was required. Leibniz decided that God might as well set things up so that they always behaved *as if* they were interacting, without particular intervention being required. Outside such a theistic framework, the theory is incredible. Even within such a framework, one might well sympathise with Berkeley's instinct that once genuine interaction is ruled out one is best advised to allow that God creates the physical world directly, within the mental realm itself, as a construct out of experience.

4. Arguments for Dualism

4.1 The Knowledge Argument Against Physicalism

One category of arguments for dualism is constituted by the standard objections against physicalism. Prime examples are those based on the existence of qualia, the most important of which is the so-called ‘knowledge argument’. Because this argument has its own entry (see the entry qualia: the knowledge argument), I shall deal relatively briefly with it here. One should bear in mind, however, that all arguments against physicalism are also arguments for the irreducible and hence immaterial nature of the mind and, given the existence of the material world, are thus arguments for dualism.

The knowledge argument asks us to imagine a future scientist who has lacked a certain sensory modality from birth, but who has acquired a perfect scientific understanding of how this modality operates in others. This scientist—call him Harpo—may have been born stone deaf, but become the world's greatest expert on the machinery of hearing: he knows everything that there is to know within the range of the physical and behavioural sciences about hearing. Suppose that Harpo, thanks to developments in neurosurgery, has an operation which finally enables him to hear. It is suggested that he will then learn something he did not know before, which can be expressed as *what it is like to hear*, or the *qualitative* or *phenomenal nature of sound*. These qualitative features of experience are generally referred to as *qualia*. If Harpo learns something new, he did not know everything before. He knew all the physical facts before. So what he learns on coming to hear—the facts about the nature of experience or the nature of qualia—are non-physical. This establishes at least a state or property dualism. (See Jackson 1982; Robinson 1982.)

There are at least two lines of response to this popular but controversial argument. First is the ‘ability’ response. According to this, Harpo does not acquire any new factual knowledge, only ‘knowledge how’, in the form of the ability to respond directly to sounds, which he could not do before. This essentially behaviouristic account is exactly what the intuition behind the argument is meant to overthrow. Putting ourselves in Harpo's position, it is meant to be obvious that what he acquires is knowledge of what something is *like*, not just *how to do* something. Such appeals to intuition are always, of course, open to denial by those who claim not to share the intuition. Some ability theorists seem to blur the distinction between knowing what something is like and knowing how to do something, by saying that the ability Harpo acquires is to *imagine* or *remember* the nature of sound. In this case, what he acquires the ability to *do* involves the representation to himself of what the thing is *like*. But this conception of representing to oneself, especially in the form of imagination, seems sufficiently close to producing in oneself something very like a sensory experience that it only defers the problem: until one has a physicalist gloss on what constitutes such representations as those involved in conscious memory and imagination, no progress has been made.

The other line of response is to argue that, although Harpo's new knowledge is factual, it is not knowledge of a new fact. Rather, it is new way of grasping something that he already knew. He does not realise this, because the concepts employed to capture experience (such as ‘looks red’ or ‘sounds C-sharp’) are similar to demonstratives, and demonstrative concepts lack the kind of descriptive content that allow one to infer what they express from other pieces of information that one may already possess. A total scientific knowledge of the world would not enable you to say which time was ‘now’ or which place was ‘here’.

Demonstrative concepts pick something out without saying anything extra about it. Similarly, the scientific knowledge that Harpo originally possessed did not enable him to anticipate what it would be like to re-express some parts of that knowledge using the demonstrative concepts that only experience can give one. The knowledge, therefore, appears to be genuinely new, whereas only the mode of conceiving it is novel.

Proponents of the epistemic argument respond that it is problematic to maintain both that the qualitative nature of experience can be genuinely novel, and that the quality itself be the same as some property already grasped scientifically: does not the experience's phenomenal nature, which the demonstrative concepts capture, constitute a property in its own right? Another way to put this is to say that phenomenal concepts are not pure demonstratives, like ‘here’ and ‘now’, or ‘this’ and ‘that’, because they do capture a genuine qualitative content. Furthermore, experiencing does not seem to consist simply in exercising a particular kind of concept, demonstrative or not. When Harpo has his new form of experience, he does not simply exercise a new concept; he also grasps something new—the phenomenal quality—with that concept. How decisive these considerations are, remains controversial.

4.2 The Argument from Predicate Dualism to Property Dualism

I said above that predicate dualism might seem to have no ontological consequences, because it is concerned only with the different way things can be described within the contexts of the different sciences, not with any real difference in the things themselves. This, however, can be disputed.

The argument from predicate to property dualism moves in two steps, both controversial. The first claims that the irreducible special sciences, which are the sources of irreducible predicates, are not wholly objective in the way that physics is, but depend for their subject matter upon interest-relative perspectives on the world. This means that they, and the predicates special to them, depend on the existence of minds and mental states, for only minds have interest-relative perspectives. The second claim is that psychology—the science of the mental—is itself an irreducible special science, and so it, too, presupposes the existence of the mental. Mental predicates therefore presuppose the mentality that creates them: mentality cannot consist simply in the applicability of the predicates themselves.

First, let us consider the claim that the special sciences are not fully objective, but are interest-relative.

No-one would deny, of course, that the very same subject matter or ‘hunk of reality’ can be described in irreducibly different ways and it still be just that subject matter or piece of reality. A mass of matter could be characterized as a hurricane, or as a collection of chemical elements, or as mass of sub-atomic particles, and there be only the one mass of matter. But such different explanatory frameworks seem to presuppose different perspectives on that subject matter.

This is where basic physics, and perhaps those sciences reducible to basic physics, differ from irreducible special sciences. 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. 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*. It would seem that, 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, such a science is formed from the collaboration between, on the one hand, objective similarities in the world and, on the other, perspectives and interests of those who devise the science. The concept of hurricane is brought to bear from the perspective of creatures concerned about the weather. Creatures totally indifferent to the weather would have no reason to take the real patterns of phenomena that hurricanes share as constituting a single kind of thing. With the irreducible special sciences, there is an issue of *salience* , which involves a subjective component: a *selection* of phenomena with a certain *teleology* in mind is required before their structures or patterns are reified. The entities of metereology or biology are, in this respect, rather like *Gestalt* phenomena.

Even accepting this, why might it be thought that the perspectivality of the special sciences leads to a genuine property dualism in the philosophy of mind? It might seem to do so for the following reason. Having a perspective on the world, perceptual or intellectual, is a psychological state. So the irreducible special sciences presuppose the existence of mind. If one is to avoid an ontological dualism, the mind that has this perspective must be part of the physical reality on which it has its perspective. But psychology, it seems to be almost universally agreed, is one of those special sciences that is not reducible to physics, so if its subject matter is to be physical, it itself presupposes a perspective and, hence, the existence of a mind to *see matter as* psychological. If this mind is physical and irreducible, it presupposes mind to see it as such. We seem to be in a vicious circle or regress.

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 top 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 this 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. If psychology cannot be reduced, this line of reasoning leads to real emergence for mental acts and hence to a real dualism for the properties those acts instantiate (Robinson 2003).

4.3 The Modal Argument

There is an argument, which has roots in Descartes (*Meditation VI*), which is a modal argument for dualism. One might put it as follows:

1. It is imaginable that one's mind might exist without one's body.
   therefore

2. It is conceivable that one's mind might exist without one's body.
   therefore

3. It is possible one's mind might exist without one's body.
   therefore

4. One's mind is a different entity from one's body.

The rationale of the argument is a move from imaginability to real possibility. I include (2) because the notion of conceivability has one foot in the psychological camp, like imaginability, and one in the camp of pure logical possibility and therefore helps in the transition from one to the other.

This argument should be distinguished from a similar 'conceivability' argument, often known as the 'zombie hypothesis', which claims the imaginability and possibility of my body (or, in some forms, a body physically just like it) existing without there being any conscious states associated with it. (See, for example, Chalmers (1996), 94–9.) This latter argument, if sound, would show that conscious states were something over and above physical states. It is a *different* argument because the hypothesis that the unaltered body could exist without the mind is *not* the same as the suggestion that the mind might continue to exist without the body, nor are they trivially equivalent. The zombie argument establishes only property dualism and a property dualist might think disembodied existence inconceivable—for example, if he thought the identity of a mind through time depended on its relation to a body (e.g., Penelhum 1970).

Before Kripke (1972/80), the first challenge to such an argument would have concerned the move from (3) to (4). When philosophers generally believed in contingent identity, that move seemed to them invalid. But nowadays that inference is generally accepted and the issue concerns the relation between imaginability and possibility. No-one would nowadays identify the two (except, perhaps, for certain quasi-realists and anti-realists), but the view that imaginability is a solid test for possibility has been strongly defended. W. D. Hart ((1994), 266), for example, argues that no clear example has been produced such that "one can imagine that p (and tell less imaginative folk a story that enables them to imagine that p) plus a good argument that it is impossible that p. No such counterexamples have been forthcoming…" This claim is at least contentious. There seem to be good arguments that time-travel is incoherent, but every episode of *Star-Trek* or *Doctor Who* shows how one can imagine what it might be like were it possible.

It is worth relating the appeal to possibility in this argument to that involved in the more modest, anti-physicalist, zombie argument. The possibility of this hypothesis is also challenged, but all that is necessary for a zombie to be possible is that all and only the things that the physical sciences say about the body be true of such a creature. As the concepts involved in such sciences—e.g., neuron, cell, muscle—seem to make no reference, explicit or implicit, to their association with consciousness, and are defined in purely physical terms

in the relevant science texts, there is a very powerful prima facie case for thinking that something could meet the condition of being just like them and lack any connection with consciousness. There is no parallel clear, uncontroversial and regimented account of mental concepts as a whole that fails to invoke, explicitly or implicitly, physical (e.g., behavioural) states.

For an analytical behaviourist the appeal to imaginability made in the argument fails, not because imagination is not a reliable guide to possibility, but because we cannot imagine such a thing, as it is a priori impossible. The impossibility of disembodiment is rather like that of time travel, because it is demonstrable a priori, though only by arguments that are controversial. The argument can only get under way for those philosophers who accept that the issue cannot be settled a priori, so the possibility of the disembodiment that we can imagine is still prima facie open.

A major rationale of those who think that imagination is not a safe indication of possibility, even when such possibility is not eliminable a priori, is that we can *imagine* that a posteriori necessities might be false—for example, that Hesperus might not be identical to Phosphorus. But if Kripke is correct, that is not a real possibility. Another way of putting this point is that there are many epistemic possibilities which are imaginable because they are epistemic possibilities, but which are not real possibilities. Richard Swinburne (1997, New Appendix C), whilst accepting this argument in general, has interesting reasons for thinking that it cannot apply in the mind-body case. He argues that in cases that involve a posteriori necessities, such as those identities that need discovering, it is because we identify those entities only by their ‘stereotypes’ (that is, by their superficial features observable by the layman) that we can be wrong about their essences. In the case of our experience of ourselves this is not true.

Now it is true that the essence of Hesperus cannot be discovered by a mere thought experiment. That is because what makes Hesperus Hesperus is not the stereotype, but what underlies it. But it does not follow that no one can ever have access to the essence of a substance, but must always rely for identification on a fallible stereotype. One might think that for the person him or herself, while what makes that person that person underlies what is observable to others, it does not underlie what is experienceable by that person, but is given directly in their own self-awareness.

This is a very appealing Cartesian intuition: my identity as the thinking thing that I am is revealed to me in consciousness, it is not something beyond the veil of consciousness. Now it could be replied to this that though I do access myself as a conscious subject, so classifying myself is rather like considering myself *qua* cyclist. Just as I might never have been a cyclist, I might never have been conscious, if things had gone wrong in my very early life. *I* am the organism, the animal, which might not have developed to the point of consciousness, and that essence as animal is not revealed to me just by introspection.

But there are vital differences between these cases. A cyclist is explicitly presented as a *human being* (or creature of some other animal species) cycling: there is no temptation to think of a cyclist as a basic kind of thing in its own right. Consciousness is not presented as a property of something, but as the subject itself. Swinburne's claim that when we refer to

ourselves we are referring to something we think we are directly aware of and not to ‘something we know not what’ that underlies our experience seemingly ‘of ourselves’ has powerful intuitive appeal and could only be overthrown by very forceful arguments. Yet, even if we are not referring primarily to a substrate, but to what is revealed in consciousness, could it not still be the case that there is a necessity stronger than causal connecting this consciousness to something physical? To consider this further we must investigate what the limits are of the possible analogy between cases of the water-H_2O kind, and the mind-body relation.

We start from the analogy between the water stereotype—how water presents itself—and how consciousness is given first-personally to the subject. It is plausible to claim that something like water could exist without being H_2O, but hardly that it could exist without *some* underlying nature. There is, however, no reason to deny that this underlying nature *could* be homogenous with its manifest nature: that is, it would seem to be possible that there is a world in which the water-like stuff is an element, as the ancients thought, and is water-like all the way down. The claim of the proponents of the dualist argument is that this latter kind of situation can be known to be true a priori in the case of the mind: that is, one can tell by introspection that it is not more-than-causally dependent on something of a radically different nature, such as a brain or body. What grounds might one have for thinking that one could tell that a priori?

The only general argument that seem to be available for this would be the principle that, for any two levels of discourse, *A* and *B*, they are more-than-causally connected only if one entails the other a priori. And the argument for accepting this principle would be that the relatively uncontroversial cases of a posteriori necessary connections are in fact cases in which one can argue a priori from facts about the microstructure to the manifest facts. In the case of water, for example, it would be claimed that it follows a priori that *if* there were something with the properties attributed to H_2O by chemistry on a micro level, then that thing would possess waterish properties on a macro level. What is established a posteriori is that it is in fact H_2O that underlies and explains the waterish properties round here, not something else: the sufficiency of the base—were it to obtain—to explain the phenomena, can be deduced a priori from the supposed nature of the base. This is, in effect, the argument that Chalmers uses to defend the zombie hypothesis. The suggestion is that the whole category of a posteriori more-than-causally necessary connections (often identified as a separate category of *metaphysical necessity*) comes to no more than this. If we accept that this is the correct account of a posteriori necessities, and also deny the analytically reductionist theories that would be necessary for a priori connections between mind and body, as conceived, for example, by the behaviourist or the functionalist, does it follow that we can tell a priori that consciousness is not more-than-causally dependent on the body?

It is helpful in considering this question to employ a distinction like Berkeley's between *ideas* and *notions*. Ideas are the objects of our mental acts, and they capture transparently—‘by way of image or likeness' (*Principles*, sect. 27)—that of which they are the ideas. The self and its faculties are not the *objects* of our mental acts, but are captured only obliquely *in* the performance of its acts, and of these Berkeley says we have *notions*, meaning by this that what we capture of the nature of the dynamic agent does not seem to have the same transparency as what we capture as the normal objects of the agent's mental acts. It is not

necessary to become involved in Berkeley's metaphysics in general to feel the force of the claim that the contents and internal objects of our mental acts are grasped with a lucidity that exceeds that of our grasp of the agent and the acts *per se*. Because of this, notions of the self perhaps have a ‘thickness’ and are permanently contestable: there seems always to be room for more dispute as to what is involved in that concept. (Though we shall see later, in 5.2.2, that there is a ‘non-thick’ way of taking the Berkeleyan concept of a *notion*.)

.

Because ‘thickness’ always leaves room for dispute, this is one of those cases in philosophy in which one is at the mercy of the arguments philosophers happen to think up. The conceivability argument creates a prima facie case for thinking that mind has no more than causal ontological dependence on the body. Let us assume that one rejects analytical (behaviourist or functionalist) accounts of mental predicates. Then the above arguments show that any necessary dependence of mind on body does not follow the model that applies in other scientific cases. This does not show that there may not be other reasons for believing in such dependence, for so many of the concepts in the area are still contested. For example, it might be argued that identity through time requires the kind of spatial existence that only body can give: or that the causal continuity required by a stream of consciousness cannot be a property of mere phenomena. All these might be put forward as ways of filling out those aspects of our understanding of the self that are only obliquely, not transparently, presented in self-awareness. The dualist must respond to any claim as it arises: the conceivability argument does not pre-empt them.

4.4 From property dualism to substance dualism

All the arguments so far in this section have been either arguments for property dualism only, or neutral between property and substance dualism. In this subsection, and in section 4.5 we will consider some arguments that have been proposed in favour of substance dualism.The ones in this section can be regarded as preliminaries to that in 4.5 and they centre on discontent with property dualism in its Humean form.

Hume is generally credited with devising what is known as the 'bundle' theory of the self (*Treatise* Book I, Part IV, section VI), according to which there are mental states, but no further subject or substance which possesses them. He famously expresses his theory as follows.

> ...when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade,love or hatred, pain or pleasure. I can never catch *myself* at any time without a perception, and can never observe any thing but the perception.

Nevertheless, in the Appendix of the same work he expressed dissatisfaction with this account. Somewhat surprizingly, it is not very clear just what his worry was, but it is expressed as follows:

> In short there are two principles, which I cannot render consistent; nor is it in my power to renounce either of them, viz. *that all our distinct perceptions are distinct existences,* and *that the mind never perceives any real connection between distinct existences*.

Berkeley had entertained a similar theory to the one found in Hume's main text in his *Philosophical Commentaries*, (Notebook A, paras 577-81), but later rejected it for the claim that we could have a *notion*, though not an *idea* of the self. This Berkeleian view is expressed in more modern terms by John Foster.

> A natural response to Hume would be to say that, even if we cannot detect ourselves apart from our perceptions (our conscious experiences)we can at least detect ourselves in them...Surely I am aware of [my experience], so to speak, from the inside - not as something presented, but as something which I have or as the experiential state which I am in...and this is equivalent to saying that I detect it by being aware of myself being visually aware. (1991: 215)

There is a clash of intuitions here between which it is difficult to arbitrate. There is an argument that is meant to favour the need for a subject, as claimed by Berkeley and Foster.

1. If the bundle theory were true, then it should be possible to identify mental events independently of, or prior to, identifying the person or mind to which they belong.
2. It is not possible to identify mental events in this way.

Therefore,

3. The bundle theory is false.

E. J. Lowe (1996) defends this argument and argues for (2) as follows.

> What is wrong with the [bundle] theory is that...it presupposes, untenably,that an account of the identity conditions of psychological modes can be provided which need not rely on reference to persons. But it emerges that the identity of any psychological mode turns on the identity of the person that possesses it. What this implies is that psychological modes are essentially modes of persons, and correspondingly that persons can be conceived of as substances.

To say that, according to the bundle theory, the identity conditions of individual mental states must be independent of the identity of the person who possesses them, is to say that their identity is independent of the bundle to which they belong. Hume certainly thought something like this, for he thought that an impression might 'float free' from the mind to which it belonged, but it is not obvious that a bundle theorist is forced to adopt this position. Perhaps the identity of a mental event is bound up with the complex to which it belongs. That this is impossible certainly needs further argument.

Hume seems, however, in the main text to unconsciously make a concession to the opposing view, namely the view that there must be something more than the items in the bundle to make up a mind. He says:

> The mind is a kind of theatre where several perceptions successively make their appearance; pass, re-pass, glide away and mingle in an infinite variety of postures and situations.

Talk of the mind as a theatre is, of course, normally associated with the Cartesian picture, and the invocation of any necessary medium, arena or even a field hypostasize some kind of entity which binds the different contents together and without which they would not be a single mind. Modern Humeans - such as Parfit (1971; 1984) or Dainton (2008) - replace the theatre with a co-consciousness relation. So the bundle theorist is perhaps not as restricted as Hume thought. The bundle consists of the objects of awareness *and the co-consciousness relation (or relations) that hold between them*, and I think that the modern bundle theorist would want to say that it is the nexus of co-consciousness relations that constitutes our sense of the subject and of the act of awareness of the object. This involves abandoning the second of Hume's principles. *that the mind can never perceive any connection between distinct existences*, because the co-consciousness relation is something of which we are aware. The Humean point then becomes that we mistake the nexus of relations for a kind of entity, in a way similar to that in which, Hume claims,we mistake the regular succession of similar impressions for an entity called an enduring physical object. Whether this really makes sense in the end is another matter. I think that it is dubious whether it can accommodate the subject as *agent*, but it does mean that simple introspection probably cannot refute a sophisticated bundle theory in the way that Lowe and Foster want. Hume's original position seems to make him deny that we have any 'sense of self' at all, whilst the version that allows for our awareness of the relatedness accommodates it, but explains how it can be an illusion. The rejection of bundle dualism, therefore, requires more than an appeal to our intuitive awareness of ourselves as subjects.We will see in the next section how arguments that defend the simplicity of the self attempt to undercut the bundle theory.

4.5 Arguments from Personal Identity

There is a long tradition, dating at least from Reid (1785), 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 and that this shows that the self is a different kind of entity from any physical body. Criticism of these arguments and of the intuitions on which they rest, running from Hume to Parfit (1970: 1984), have left us with an inconclusive clash of intuitions. The argument under consideration and which, possibly, has its first statement in Madell (1981), does not concern identity through time, but the consequences for identity of certain counterfactuals concerning origin. It can, perhaps, therefore, break the stalemate which faces the debate over diachronic identity. The claim is 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.

Concerning ordinary physical objects, it is easy to imagine counterfactual cases where questions of identity become problematic. Take the example of a particular table. We can scale counterfactual suggestions 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% of the wood it was made of and 5% of some other wood.

The first suggestion would normally be rejected as clearly false, but there will come a point along the spectrum illustrated by (i) and (iii) and towards (iii) where the question of whether the hypothesised table would be the same as the one that actually exists have 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% the same matter and of 25% 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.

Let us now apply this thought to conscious subjects. 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' 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. There is no space to discuss this issue here. It is enough to assume, however, that questions of how one is allowed to use the concept of identity effect 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.)

One might plausibly claim that no similar *overlap of constitution* can be applied to the counterfactual identity of minds. In Geoffrey Madell's (1981) 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. (91)

Why is this so? Imagine the case where we are not sure whether it would have been Jones' 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. (i) I either would or would not, but I cannot tell. (ii) There is no fact of the matter whether I would or would not have existed: it is just a mis-posed question. (iii) 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 is arguable that this 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% of Jones2's original body would have been identical with Jones', about 85% of his psychic life would have been Jones'? That it would have been *like* Jones'—indeed that Jones2 might have had a psychic life 100% like Jones'—makes perfect sense, but that he might have been to that degree, the same psyche—that Jones '85% existed' —arguably makes no sense. Take the case in which Jones and Jones2 have *exactly* similar lives throughout: which 85% of the 100% 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% 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?

If the reasoning above is correct, one is left with only the first option. If so, 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.

4.5 The Aristotelian Argument in a Modern Form

Putting his anti-materialist argument outlined above, in section 1, in very general terms, Aristotle's worry was that a material organ could not have the range and flexibility that are required for human thought. His worries concerned the cramping effect that matter would have on the range of *objects* that intellect could accommodate. Parallel modern concerns centre on the restriction that matter would impose on the range of rational *processes* that we could exhibit. Godel, for example, believed that his famous theorem showed that there are demonstrably rational forms of mathematical thought of which humans are capable which could not be exhibited by a mechanical or formal system of a sort that a physical mind would have to be. Penrose (1990) has argued that Turing's halting problem has similar consequences. In general, the fear is that the materialist monist has to treat the organ of thought as, what Dennett (1987:61) calls, a *syntactic engine*: that is, as something that operates without any fundamental reference to the *propositional content* of what it thinks. It works as a machine that only shadows the pattern of meaning. But it is hard to convince oneself that, as one, for example, reflectively discusses philosophy and struggles to follow what is being said, that it is not the semantic content that is driving one's responses. But if we are truly semantic engines, it is difficult to see how we can avoid at least a property dualism. These issues are, of course, connected with problems raised by Brentano, concerning the irreducibility of intentionality. Despite the interest of the arguments for dualism based on the irreducible flexibility of intellect, most of the modern debate turns on arguments that have a Cartesian origin.

5. Problems for Dualism

We have already discussed the problem of interaction. In this section we shall consider two other facets of dualism that worry critics. First, there is what one might term the *queerness* of the mental if conceived of as non-physical. Second there is the difficulty of giving an account of the unity of the mind. We shall consider this latter as it faces both the bundle theorist and the substance dualist.

5.1 The Queerness of the Mental

Mental states are characterised by two main properties, subjectivity, otherwise known as privileged access, and intentionality. Physical objects and their properties are sometimes observable and sometimes not, but any physical object is equally accessible, in principle, to anyone. From the right location, we could all see the tree in the quad, and, though none of us can observe an electron directly, everyone is equally capable of detecting it in the same ways using instruments. But the possessor of mental states has a privileged access to them that no-one else can share. That is why there is a sceptical ‘problem of other minds’, but no corresponding ‘problem of my own mind’. This suggests to some philosophers that minds are not ordinary occupants of physical space.

Physical objects are spatio-temporal, and bear spatio-temporal and causal relations to each other. Mental states seem to have causal powers, but they also possess the mysterious property of intentionality—being *about* other things—including things like Zeus and the square root of minus one, which do not exist. No mere physical thing could be said to be, in a literal sense, ‘about’ something else. The nature of the mental is both queer and elusive. In Ryle's deliberately abusive phrase, the mind, as the dualist conceives of it, is a ‘ghost in a machine’. Ghosts are mysterious and unintelligible: machines are composed of identifiable parts and work on intelligible principles. But this contrast holds only if we stick to a Newtonian and common-sense view of the material. Think instead of energy and force-fields in a space-time that possesses none of the properties that our senses seem to reveal: on this conception, we seem to be able to attribute to matter nothing beyond an abstruse mathematical structure. Whilst the material world, because of its mathematicalisation, forms a tighter abstract system than mind, the sensible properties that figure as the objects of mental states constitute the only intelligible content for any concrete picture of the world that we can devise. Perhaps the world within the experiencing mind is, once one considers it properly, no more—or even less—queer than the world outside it.

5.2 The Unity of the Mind

Whether one believes that the mind is a substance or just a bundle of properties, the same challenge arises, which 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. Neither tradition has been notably successful in this latter task: indeed, Hume, in the appendix to the *Treatise*, declared himself wholly mystified by the problem, rejecting his own initial solution (though quite why is not clear from the text).

5.2.1 Unity and Bundle Dualism

If the mind is only a bundle of properties, without a mental substance to unite them, then an account is needed of what constitutes its unity. The only route appears to be to postulate a primitive relation of co-consciousness in which the various elements stand to each other.

There are two strategies which can be used to attack the bundle theory. One is to claim that our intuitions favour belief in a subject and that the arguments presented in favour of the bundle alternative are unsuccessful, so the intuition stands. The other is to try to refute the theory itself. Foster (1991, 212–9) takes the former path. This is not effective against someone who thinks that metaphysical economy gives a prima facie priority to bundle theories, on account of their avoiding mysterious substances.

The core objection to bundle theories (see, for example, Armstrong (1968), 21–3) is that, because it takes individual mental contents as its elements, such contents should be able to exist alone, as could the individual bricks from a house. Hume accepted this consequence, but most philosophers regard it as absurd. There could not be a mind that consisted of a lone pain or red after-image, especially not of one that had detached itself from the mind to

which it had previously belonged. Therefore it makes more sense to think of mental contents as *modes of a subject*.

Bundle theorists tend to take phenomenal contents as the primary elements in their bundle. Thus the problem is how to relate, say, the visual field to the auditory field, producing a 'unity of apperception', that is, a total experience that seems to be presented to a single subject. Seeing the problem in this way has obvious Humean roots. This atomistic conception of the problem becomes less natural if one tries to accommodate other kinds of mental activity and contents. How are acts of conceptualising, attending to or willing with respect to, such perceptual contents to be conceived? These kinds of mental acts seem to be less naturally treated as atomic elements in a bundle, bound by a passive unity of apperception. William James (1890, vol. 1, 336–41) attempts to answer these problems. He claims to introspect in himself a 'pulse of thought' for each present moment, which he calls 'the Thought' and which is the 'vehicle of the judgement of identity' and the 'vehicle of choice as well as of cognition'. These 'pulses' are united over time because each 'appropriates' the past Thoughts and 'makes us say "as sure *as I exist*, those past facts were part of myself". James attributes to these Thoughts acts of judging, attending, willing etc, and this may seem incoherent in the absence of a genuine subject. But there is also a tendency to treat many if not all aspects of agency as mere awareness of bodily actions or tendencies, which moves one back towards a more normal Humean position. Whether James' position really improves on Hume's, or merely mystifies it, is still a moot point. (But see Sprigge (1993), 84–97, for an excellent, sympathetic discussion.)

5.2.2 Unity and Substance Dualism

The problem is to explain what kind of a thing an immaterial substance is, such that its presence explains the unity of the mind. The answers given can be divided into three kinds.

(a) The 'ectoplasm' account: The view that immaterial substance is a kind of immaterial stuff. There are two problems with this approach. First, in so far as this 'ectoplasm' has any characterisation as a 'stuff'—that is, a structure of its own over and above the explicitly mental properties that it sustains—it leaves it as much a mystery why *this* kind of stuff should support consciousness as it is why ordinary matter should. Second, and connectedly, it is not clear in what sense such stuff is immaterial, except in the sense that it cannot be integrated into the normal scientific account of the physical world. Why is it not just an aberrant kind of physical stuff?

(b) The 'consciousness' account: The view that consciousness is the substance. Account (a) allowed the immaterial substance to have a nature over and above the kinds of state we would regard as mental. The consciousness account does not. This is Descartes' view. The most obvious objection to this theory is that it does not allow the subject to exist when unconscious. This forces one to take one of four possible theories. One could claim (i) that we are conscious when we do not seem to be (which was Descartes' view): or (ii) that we exist intermittently, though are still the same thing (which is Swinburne's theory, (1997), 179): or (iii) that each of us consists of a series of substances, changed at any break in consciousness, which pushes one towards a constructivist account of identity through time and so towards the spirit of the bundle theory: or (iv) even more speculatively, that the self

stands in such a relation to the normal time series that its own continued existence is not brought into question by its failure to be present in time at those moments when it is not conscious within that series (Robinson, forthcoming).

(c) The ‘no-analysis' account: The view that it is a mistake to present any analysis. This is Foster's view, though I think Vendler (1984) and Madell (1981) have similar positions. Foster argues that even the ‘consciousness’ account is an attempt to explain what the immaterial self is ‘made of’ which assimilates it too far towards a kind of physical substance. In other words, Descartes has only half escaped from the ‘ectoplasmic’ model. (He *has* half escaped because he does not attribute non-mental properties to the self, but he is still captured by trying to explain what it is made of.)

Foster (1991) expresses it as follows:

> …it seems to me that when I focus on myself introspectively, I am not only aware of being in a certain mental condition; I am also aware, with the same kind of immediacy, of being a certain sort of thing…
>
> It will now be asked: ‘Well, what *is* this nature, this sortal attribute? Let's have it specified!’ But such a demand is misconceived. Of course, I can give it a verbal label: for instance, I can call it ‘subjectness’ or ‘selfhood’. But unless they are interpreted ‘ostensively’, by reference to what is revealed by introspective awareness, such labels will not convey anything over and above the nominal essence of the term ‘basic subject’. In this respect, however, there is no difference between this attribute, which constitutes the subject's essential nature, and the specific psychological attributes of his conscious life…
>
> Admittedly, the feeling that there must be more to be said from a God's eye view dies hard. The reason is that, even when we have acknowledged that basic subjects are wholly non-physical, we still tend to approach the issue of their essential natures in the shadow of the physical paradigm. (243–5)

Berkeley's concept of *notion* again helps here. One can interpret Berkeley as implying that there is more to the self than introspection can capture, or we can interpret him as saying that notions, though presenting *stranger* entities than ideas, capture them just as totally. The latter is the ‘no account is needed’ view.