"Philosophy: Themes & Thinkers
J.W. Phelan
2005 Cambridge

4 Philosophy of mind

Dualism ... is usually adopted on the grounds that it must be true, and rejected on the grounds that it can't be true. (Thomas Nagel)

Aims

On completion of this chapter you should be able to:

- evaluate substance and property dualism
- evaluate the different versions of physicalism (identity theory, eliminative materialism and biological naturalism)
- explain and criticise philosophical behaviourism
- explain and criticise functionalism
- outline evidence in support of each of these theories of mind and understand the tensions between them
- outline the main arguments advanced in the artificial intelligence debate
- understand key notions in the philosophy of mind such as supervenience, qualia, consciousness, privileged access and intentionality
- appreciate the difficulties in explaining how the mind and the body interact
- appreciate the problem of establishing whether other individuals have minds
- appreciate the problem of self-ascription
- appreciate the problem of establishing criteria for continued personal identity through time.

Philosophy has been likened to a conversation spanning over two millennia. If this is so, then the philosophy of mind represents its latest pronouncement, as it combines twentieth-century scientific knowledge with a timeless fascination as to the nature of human thought. The present age sees philosophers in conference with neurobiologists and computer scientists in an attempt to offer an explanation for mental activity, or what is generally termed 'consciousness'. Three approaches are presented below. The first standpoint, known as dualism, holds that there are two separate and mutually exclusive kinds of thing: the physical and the mental. **Physicalism**, on the other hand, holds that consciousness is entirely explicable in terms of physical processes in the brain. The third category of theories, which includes behaviourism and functionalism, tries to sidestep the dualist/physicalist stalemate by arguing that we can only talk meaningfully about minds in terms of behaviour or causal roles. Let us examine the strengths and weaknesses of each position in turn.

Dualism

Dualism is the assumption that there are two worlds: the observable, physical world of brain states and the private, psychical world of mental states. These two realms interact with each other, so that we can talk about a mental event causing a physical one (for example, the desire for a cup of tea causes me to make one) or a physical event causing a mental one (for example, drinking a nice cup of tea causes the feeling of satisfaction). As the opening quotation to this chapter shows, dualism has a popular and intuitive appeal. To assert that we are conscious creatures with private, non-spatial, qualitative thoughts while being physically embodied seems self-evident. Indeed, for a long time the governance of mind over matter was accepted as true without dispute. A popular analogy for the relationship

between mind and body was that of master and slave.[1] One of the earliest written accounts of dualism, recorded in a dialogue attributed to Plato, *Alcibiades I*, claimed that as the soul uses, rules and has authority over the body, any definition of humanity that involves awarding the body significance should be rejected.

Substance dualism

Substance dualism, which is also referred to as Cartesian or classical dualism, asserts that there are two kinds of things or 'substances' in the world. The term 'substance' refers to any entity that does not depend on anything else for its existence, in contrast to features or 'properties', which by definition belong to something else. Descartes is often referred to as the founding father of modern philosophy of mind, and was influenced by the dualism of Plato. Indeed, there are many passages in his *Meditations* (see chapter 9) that could have been penned by the ancient Greek.[2] It is in Descartes that substance dualism finds its most eloquent advocate. Throughout the *Meditations* Descartes reflects on his innermost thoughts in a method known as **introspection**. He retreats from everyday life to the seclusion of his study and then retreats from the surroundings of his study to the dark recesses of his thoughts. This inner world becomes a private laboratory in which he performs certain thought experiments, which confirm that the physical and mental are distinct yet interactive substances. Let us examine his main arguments in favour of dualism.

The argument from doubt

1 I can doubt that my body exists.
2 I cannot doubt that I exist.
3 Therefore, I must be distinct (a different substance) from my body.

There seems to be a property (let us call it indubitability) that I possess but that is not shared by any physical thing. This property assures me of my own existence, and is cited as proof that the mind and body are separate. Many have argued, however, that the above syllogism (logical deduction) does not stand up. Let us make the error conspicuous by forming a parallel argument that makes the same move but is obviously false. Imagine Elizabeth II declaring:

1 I can doubt that a queen of England exists.
2 I cannot doubt that I exist.
3 Therefore, I am not the queen of England.

But at the time of her declaration she is the queen of England! In asserting the above, she makes herself the victim of mistaken identity. Physicalists argue that the same case of mistaken identity surrounds the 'I' in the argument from doubt.

Many philosophers accuse Descartes of committing a fallacy in his use of introspection, as one cannot speak objectively about oneself, but only about things that are outside one's experience. If he is alleging that there is, within his constitution, a part of himself that is neutrally observing and another bit that is being impartially observed, then this seems an arbitrary distinction and one that leads to an infinite regress of observer and observed. One might allege that there is a further part observing the observer, and another observing this observer like a series of Russian dolls. Critics call this error in reasoning the 'homunculus fallacy' as it depicts lots of little observers in one's mind (*homunculus* is Latin and means 'little man').

The argument from indivisibility

Descartes' Sixth Meditation offers an argument in favour of dualism known as the argument from indivisibility, which has in turn led to a sister argument known as the argument from irreducibility. The argument from indivisibility is presented below:

1 The body is divisible into parts.
2 The mind is not divisible into parts.
3 Therefore, the mind must be of an entirely different nature from the body.

While it is not possible to cleave the mind in two with an axe (as the mind is non-spatial), one can perform such a gruesome deed to a brain; therefore, minds and brains must be different types of thing. Yet to assert that the mind cannot be 'divided' appears to fly in the face of the categories that we impose on thoughts, dividing them into perceptions, personality traits, emotions, memories etc. What dualists mean is that these things cannot be reduced to the physical. Yet isn't this a case of pantomime

philosophy where one side asserts one thing, i.e. that mental states cannot be reduced, while the other side asserts the opposite, i.e. 'Oh yes they can'? We wait to see how this disagreement can be resolved.

The argument from irreducibility

1 Mental activities such as reason, intuition, language-use and subjective experience cannot be reduced to physical explanations.
2 Things either have a physical or a mental explanation.
3 There must be non-physical explanations for these activities.

Advocates of the above argument include John Foster[3] and Richard Swinburne,[4] who hold that no physicalist theory could offer an adequate account of the mental. Even if this is correct, the question of how a dualist theory could explain the mental remains and has not been sufficiently well answered.

In his meditations on dualism, Descartes invites us to picture imageless thoughts: our disembodied selves, God, and mathematical figures such as a chiliagon (a thousand-sided figure). Our capacity to conceive of such imageless thoughts gives Descartes sufficient grounds to conclude that our thoughts are not physical entities. This doctrine became known as rationalism. We are born with certain God-given intuitions that are incorruptible by space, time or matter. Dualism fits readily into a religious view of the world, and many theistic philosophers subscribe to dualism, as they cannot see how a physicalist account of human beings is compatible with the survival of the soul. Of course, one does not have to be a member of an institutionalised religion to have a sense of a mysterious, other-worldly realm beyond the material world. Strange occurrences such as telepathy (mind-reading), precognition (seeing the future), telekinesis (thought control of material objects), clairvoyance (knowledge of distant objects), and astral projection (mind leaving the body) have been documented, and so far appear to defy physicalist explanation. There seems to be more in this world than can ever be dreamt of by any physicalist philosophy.

Versions of dualism

Epiphenomenalism is a dualist theory of mind–body interaction that maintains that all mental events are causally dependent upon physical states (i.e. brain states). Mental activity is an epiphenomenon: something that is caused but itself causes nothing, like a cog in a machine that is turned, but which does not itself turn any other cogs. According to this theory, brain events cause mental events, but mental events, although existing in their own right, do not cause any events.

Interactionism, on the other hand, claims that there is a two-way causal connection between mind and body. This means that the body (specifically the brain) can cause mental events such as pain, and also that mental events such as the desire to avoid pain can have a causal influence on the body, as when a person moves her hand away from a fire.

Occasionalism holds that there is no causal interaction between minds and bodies. Instead, mental events are caused by God to occur in synchrony with physical events, so that mind and body only interact via God. This is a version of parallelism, which also maintains that mental and physical events run on a parallel course but do not causally interact with each other. A popular illustration of this theory depicts two watches keeping exactly the same time because they are in harmony and not because there is a causal connection between them.

Criticisms of dualism

Minds are not the genuine article

The term 'mind' occurs in our everyday conversations, but the fact that we talk about minds does not necessarily mean that minds exist.

An objection to dualism is to say that the mind is not a genuine entity, so that to talk about X's mind is comparable to talking about X's sake. When we say, 'Person Y did this for X's sake', the sake is not a genuine object in its own right. Dualists might claim, however, that a relationship exists between a person's brain and their mind analogous to that between an orchestra and its music. The instruments are the only way the music can be presented; music, however, can be considered a genuine entity.

Dualism is never a tenable position for Hume, as it wallows in the muddy waters of metaphysics. He seeks to replace traditional armchair questions such as 'What is the mind?' with questions that can be empirically investigated, such as 'What kind of things are my beliefs about?' and 'What are the sources of my beliefs?' Like Descartes, Hume tries to meditate on the nature of self, but unlike the Frenchman he concludes that there is no self independent of experiences to meditate on. Instead, the self is a bundle of different impressions (see chapter 10). A version of Hume's bundle theory is defended by the philosopher Derek Parfit (born 1942).

The argument from category error

The most famous criticisms of dualism appear in the first chapter of Gilbert Ryle's book *The Concept of Mind* (1949). Ryle (1900–1976) accuses substance dualists of making a category error, i.e. a mistake in classification. The argument from category error separates the meaningful from the meaningless in the style of logical positivism. Whereas for Ryle it makes perfect sense to say 'My football team won the game' or 'There are 15 individual members in the team', it does not make sense to assert 'The team exists in addition to its members'. It is this exact fault that Ryle accuses dualists of making. Thus one cannot say that the mind exists in addition to the mental activity and behaviour displayed. Note that this problem is shared by descriptions of the body, which raise similar metaphysical questions such as: What makes something my body? Am I and my body two things or one? Could two people share a single body?

Ryle seeks to reduce the concept of mind to such a degree as to render it redundant. He stresses the absurdity of what he pejoratively calls the 'ghost in the machine theory' by stating that it forces us to think of our lives as running two parallel courses, one mental and one physical, like two disparate autobiographies. In addition, he asserts that when one talks of minds and thoughts occurring 'inside', one is only talking metaphorically, and therefore the content is not an expression of literal truth. One further criticism focuses on causality and language, and runs as follows:

1 Any chain of cause and effect can always be traced back to a single physical cause.
2 A mental event can only cause another mental event if mediated through the physical.
3 It is not necessary to talk in terms of mental activity.
4 The concept of mind is a redundant one.

Scientific evidence

Another criticism of dualism may be taken from the many examples that seem to prove the dependency of the mental on the neurological. One such instance is the case of a nineteenth-century railway worker, Phineas P. Gage, whose story was reported in a publication of the Massachusetts Medical Society (1868), which described the normally placid and reliable man turning into an aggressive alcoholic after surviving an accident in which a steel rod pierced his lower jaw and stuck through his head. Dualists might respond by denying any necessary connection between brain damage and personality change, but case scenarios pointing to the opposite conclusion abound.

There are cases of people whose *corpus callosum*, the conduit of information unifying the left and right hemispheres of the brain, has been severed, resulting in the unusual effect that the two sides of the brain know different things. For instance, the right side of the brain will pick up information from the left visual field and the left side of the brain will pick up information from the right visual field.

This plethora of criticisms has led Daniel Dennett to describe dualism as 'not so much a philosophical position but more of a cliff over which you push your opponents'. **Property dualism** is one theory that attempts to retrieve the situation.

Property dualism

Can we dispense with the idea of a distinct, metaphysical substance called 'the mind' but retain the 'otherness' of thoughts? Property or attributive dualism attempts such a move. The American philosopher Thomas Nagel (born 1937), in his influential essay 'What Is It Like to Be a Bat?' (1974), seeks to reconcile the existence

of subjective experience with the materialism of a scientific age. Nagel asserts that only the physical exists, but that brain waves curiously exhibit a second class of characteristics in addition to material properties, which we might call mental properties. He begins by stating that one cannot reduce consciousness and that consciousness entails subjective experience: 'the fact that an organism has conscious experience *at all* means, basically, that there is something it is like *to be* that organism. We call this the subjective character of experience'.[5] Focusing on the nature of subjective experience, he concludes that it is inexplicable in terms of its purpose, the content of thought or any causal relationship. And any physicalist analysis is at present inadequate in providing an explanation.

Moving on to the illustration promised in the essay's title, Nagel argues as follows:

1 Bats have experience, i.e. there is something like 'being a bat'.
2 Human beings are restricted to the resources of their own minds.
3 Therefore, human beings cannot feel what it is like to be a bat.

The above argument seeks to highlight the exclusivity of subjective experience. It concludes that consciousness exhibits both a private, subjective property and a public, objective property and that these are irreconcilably different. Yet property dualism still appears to skirt around the issue, as it fails to explain what such private, mental properties consist in and how they are related to physical, neuro-physiological activity. It is not clear what the assertion that 'mental properties are not physical properties' actually means other than to say that science in its present form is not advanced enough to understand mental states.

Supervenience

Crude substance dualism gives rise to other theories such as epiphenomenalism and property dualism. These theories attempt to assert in different ways that there exists first and foremost a physical system that gives rise to a mental experience. Respectively, these theories construe the mental as a causal relationship or a distinct set of attributes. Each theory believes that there is a necessary connection between the mental and the physical. This necessary connection is known as **supervenience**, a term that here means 'emergence from'. Supervenient features (the mental) arise out of a subvenient base (the physical). Supervenient features exist in their own right, although they change if the subvenient base changes. Consider the illustrations in figure 4.1; both are made up of different patterns of dots, but one is able to make

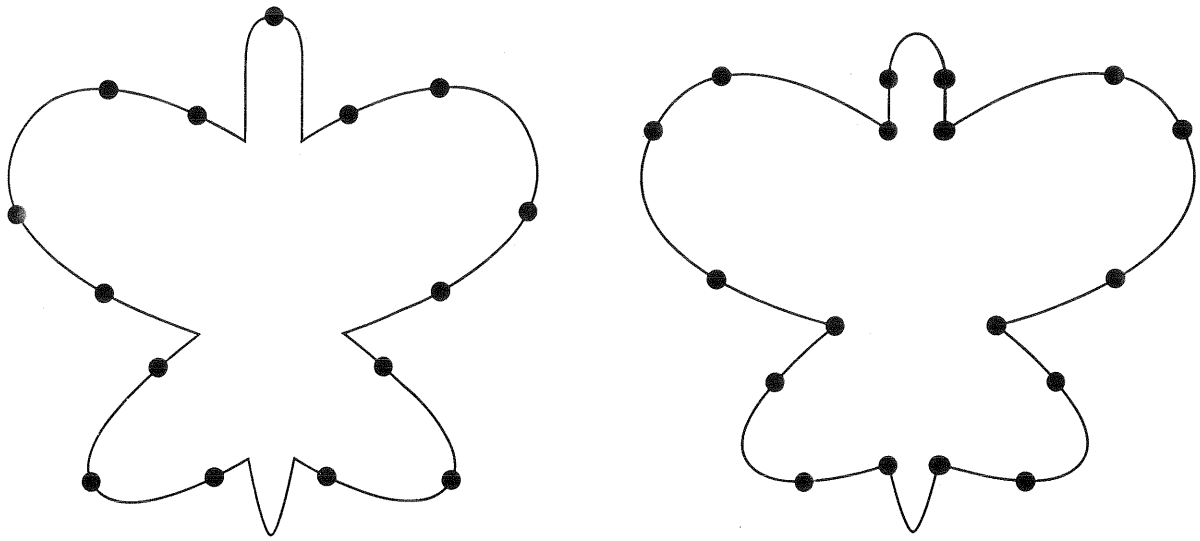


Figure 4.1 *Supervenience*

out the outline of a butterfly in each case. The two different patterns of dots are examples of subvenient bases, while the image of a butterfly may be said to supervene on these patterns. In a similar way, according to supervenience theorists, thoughts can be said to emerge from neural sequences in our brains. Let us move on to examine the physicalist theories of mind.

Review question

Describe and illustrate one argument in favour and one argument against substance dualism.

Physicalism

In contrast to dualism, theories that offer a purely physical explanation of consciousness are called physicalist or materialist theories. Physicalism is a form of **monism** – that is, the belief that the universe is made up of only one kind of substance, physical or mental, but not both. One advantage of monism is that there is no need to try to explain any mysterious interaction between such radically different things as minds and bodies. If you are a monist and an idealist, you think that the only things in existence are minds and their perceptions; if you are a monist and a physicalist, then you think that only material things exist and that philosophers of mind should only be interested in brains. Three versions of physicalism are examined below: identity theory (reductive materialism), eliminative materialism and biological naturalism.

Identity theory (reductive materialism)

Identity theory, or the Australian Heresy as it came to be known, was popular during the 1950s. It should not, however, be viewed as a mere historical stepping stone across the river of progress but a worthwhile spot to pause and fish for arguments. Identity theorists maintain that mental events are identical with events taking place in the brain. This is to say that the relationship between mind and brain, between thoughts and the firing of neurones, is the same as that between, for example, lightning and electrical discharge. This is not to say, however, that the firing of a neurone means the same thing as a thought. Similarly, 'lightning' has a different meaning from the term 'electrical discharge'. I do not look out of the window during a storm and say 'I've just had a certain sequence of neurological firings' but 'I have seen some lightning'. The point is that I can use two different terms, which carry with them different qualitative elements, while still referring to the same thing.

One of the most famous examples of this is the fact that the terms 'morning star' and 'evening star' both refer to the planet Venus, yet have different subjective meanings. The descriptions don't mean the same, or else the ancient Babylonians could have discovered their shared identity *a priori*, without any astronomical research. In the same way, talk about thoughts and the firings of neurones differ in meaning yet have the same reference. Identity theorists argue that mental states are identical to physical states of the brain. The mind is the brain. They believe that brain research will be able to show how consciousness is generated, and the term 'mind-brain' may be favoured in the future to explain this feature. Identity theorists point to the widespread and systematic correlations between mental states and brain states. Yet correlations are not explanations, and many philosophers remain unconvinced.

A problem for identity theory arises when one asserts that if two people are having the same thought about something then they are sharing the same brain state also. If you and I are both looking at a beautiful coastline and we express the thought of its beauty, then identity theory asserts that an identical process will be going on in each of our brains. The problem with this is that my thought about the picturesque coastline is wrapped up with my past experiences of coastlines and with what I consider beautiful in general. For your part, your thought about the coastline is inseparable from all your past experiences in a similar way. If this is true, then the sense in which we can both be said to be having the same thought is weakened, and the sense in which the whole of our mental life can be compared disappears. If we cannot say that two people are having the same thought, then the idea of identifying thoughts with brain processes becomes problematic, because it

assumes that if you can identify the same brain process in one person as in another then you are entitled to hold the belief that they have the same thought.

Leibniz's law

Statements concerning identity have held a peculiar fascination for many philosophers. One was the German philosopher Gottfried Wilhelm Leibniz, who formulated a special law of identity used to test whether two or more seemingly different descriptions are about the same thing. The usual formulation of **Leibniz's law** runs as follows:

> If X = Y, then X has the property F if and only if Y has the property F.

This version presumes that we can first establish that X = Y, which seems to jump the gun. A more accurate phrasing of Leibniz's law may be presented as follows:

> All the properties of X and Y are identical if and only if X = Y, that is to say if X and Y are the same thing.

One can see that this works perfectly in the previous examples, where X might represent lightning and Y electrical discharge.

As it stands, Leibniz's law seems like a good principle of logic, if not a scintillatingly interesting one. However, there have been challenges to it, and the case of the two electrons is cited as a possible exception. One is invited to imagine that the universe consists of only two electrons. They share every property (mass, spin etc.) and, not surprisingly, they are the same distance from each other. Spatial position is irrelevant, as they are the only two things in the universe. According to Leibniz's law, they must be the same electron, but there are two!

If identity theorists are right and mental states are identical to physical states of the brain then, according to Leibniz's law, they must be identical in every way. Opponents of identity theory use Leibniz's law to assert that mental states are not numerically identical (one and the same thing) with brain states. Consider the following two examples.

Gottfried Leibniz (1646–1716)

Gottfried Wilhelm Leibniz was the son of a professor of moral philosophy from Leipzig, Germany. His father died when he was just six years old, and the young Gottfried made an effort to learn Latin and Greek from an early age in order to read his father's books. Leibniz was brought up in the German Lutheran tradition and became interested in theological questions. He published his work entitled *Theodicy* in 1710, which inspired Voltaire's satirical novel *Candide* in which Leibniz is caricatured as the indefatigable Dr Pangloss, who believed that everything was for the best in the best of all possible worlds, despite evidence to the contrary. During his lifetime, Leibniz was described as a stooped figure of medium height, broad-shouldered but bandy-legged, who was capable of sitting in the same chair and thinking for several days. This thinking paid off, and Leibniz became an immensely influential figure in mathematics and logic, developing the binary system of arithmetic as well as differential and integral calculus. Like many philosophers, he never married. He worked in Hanover from 1676 at a strange selection of jobs, from library work and teaching to designing a wind-driven pump to drain water from a nearby mine. Leibniz was also a prodigious letter writer and had over 600 scholarly correspondents across Europe. He died in 1716, plagued with gout but leaving the world of philosophy a richer place.

Significant works

New Essays on Human Understanding (c. 1705)
Theodicy (1710)
Monadology (1714)

Example 1: a Cartesian argument

1 All entities fall into two mutually exclusive categories: space-occupying entities and conscious, non-extended entities.
2 The spatial location of brain processes can be specified with a high degree of accuracy.
3 The spatial location of thoughts or emotions cannot be identified at all.
4 If the brain is spatial and the mind is non-spatial, then, according to Leibniz's law, the brain cannot be numerically the same as the mind.

Example 2: an argument from intentionality

1 Mental states are about states of affairs external to themselves, e.g. beliefs, desires, dreams, hopes, thoughts are about something apart from the thought itself.
2 For a thought to be about something means that it represents something.
3 No physical state can possess representational content.
4 Therefore, mental states cannot be, according to Leibniz's law, numerically identical with physical states.

The debate takes the following twists and turns. First, identity theorists inform us that the mind is the brain. This is a statement of identity, and as such should fit with Leibniz's law. Critics of identity theory assert that it doesn't, and cite examples such as spatial location and representational content as instances where minds and brains differ.

Task

Can our thoughts be reduced to brain waves?

Consider whether the following statements are successful in identifying a difference between mind and brain:

1 Brain waves occupy space; thoughts do not.
2 Thoughts are about something, whereas brain waves are not.
3 I can keep my thoughts private, but a neuroscientist can measure my brain activity in public.
4 I can attribute value to my thoughts (e.g. a naughty suggestion), but my brain activity exists in a realm of fact, not value.
5 Our thoughts have distinct felt qualities, e.g. pangs of jealousy, but our brain activities do not.
6 I can know what my thoughts are by virtue of the fact that I am me, but in order to understand my brain activity I need to study neuroscience.

The type–token distinction

Identity theorists have attempted to refine their theory using the type–token distinction.

The term 'type' is any kind of thing, which is shared by many particular entities. The term 'token' refers to a particular instance or exemplar of a type, which is located in some particular space or time. Consider the following sentence:

The bat sat on the mat.

There are six individual words in this sentence that we might call specific instances or tokens. When we look closer we see that there are two specific instances of the word 'the', so the word 'the' can be classed as a type. Type identity refers to the kind of thought we might be experiencing, whereas token identity describes specific thoughts. For example, type identity theorists argue that pain and the firing of C-fibres in my brain are identical. Both arms of the identity (pain and the firing of C-fibres) are construed as the same type of thing. This account runs into difficulties if it turns out that pain can be realised in another way, perhaps by the activity of B-fibres. In this case the claim that pain and the firing of C-fibres are identical is proved false. In order to maintain identity theory one needs to frame a version in terms of tokens. So, for instance, one would say that a particular token of pain is identical to C-fibres firing, while a different token is identical to B-fibres firing.

Kripke's criticism

A further criticism waged against identity theory issues from the American philosopher Saul Kripke, who argues that 'The mind is the brain', as a statement of identity, should be a necessary truth but isn't. This accusation is based on the fact that statements of identity are necessary truths and that if different names refer to the same thing (for example, the terms 'morning star' and 'evening star' both refer to the planet Venus),

then this is a necessary truth in every possible world. The fact that one can conceive of a possible world in which the mind is not the brain means that identity theory does not advance anything that may be classed as a necessary truth. Let us phrase this objection formally:

1 Identity theorists assert that 'mental states are physical states' is a necessary truth about the world.
2 To be a necessary truth about the world, a statement needs to be true in this actual world and true in all possible worlds (e.g. water is H_2O).
3 We can, however, imagine a world where there is brain activity, e.g. C-fibres fire in the brain (physical state), but the individual, for whatever reason, does not feel pain.
4 Therefore, it is possible to conceive of a brain state occurring without the presence of a corresponding mental state. Therefore, mental states and physical states are not identical.

The fallacy of intentionality

Identity theorists attempt to dodge the above criticisms by catching their critics out on a technicality. They assert that all the examples cited in the task on page 79 commit the fallacy of intentionality. This occurs when an argument refers to the arguer's beliefs and not to the object of those beliefs. So, for instance, if I were to argue that my belief in God proves that God exists, this would commit the fallacy of intentionality, as my belief does not affect God's existence in any way but only tells you something about me. Identity theorists assert that one's belief in doubt, privacy, value-attribution, after-images, subjective experience and introspection go a long way in telling us about our own thoughts, but this does not prove identity theory wrong, as it does not conclusively show that these features belong to minds rather than brains. A more hardline theory that challenges the notion of mental states altogether is eliminative materialism.

Eliminative materialism

In order to understand **eliminative materialism**, one needs to establish exactly what is being eliminated. The theory is in the tradition of logical positivism, the twentieth-century movement of scientifically minded philosophers who sought to distinguish the factually meaningful from the factually meaningless according to whether the statement under analysis is true by definition or provable by experiment (these are meaningful statements; other types of statement are meaningless). Contemporary eliminativists, such as the Canadian philosophers Paul and Patricia Churchland (born in 1942 and 1943 respectively) and the American philosopher Daniel Dennett (born 1942), seek to perform the same linguistic operation on descriptions of mental states. Their conclusion is that any notions of free-floating thoughts or beliefs are factually nonsensical. They believe that such 'hornswoggle' or 'gobbledegook' is the product of an unscientific method, and should be heaped under the pejorative label of **folk psychology**. This phrase may conjure up images of bearded fiddle players discussing the works of Freud, but here it means any non-empirical theory that conjectures a non-physical realm of thought.

According to the Churchlands, it is the job of science to eliminate conceptual frameworks that misinterpret the phenomena. Historical examples abound, such as the concept of witches and demonic possession in the field of mental health or ancient alchemical arguments over how base metals are turned into gold. Such superstition has been dispelled by science as nonsensical and has now only minor historical significance. Eliminativists believe that the same fate awaits the concept of mind:

> Eliminative materialism is the thesis that our common-sense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced.[6]

The Churchlands' argument against what they call folk psychology proceeds along the following lines. First, they classify folk psychology as a scientific theory, as it uses rules and laws to make predictions. Then they point out that the theory fails in offering an explanation in a number of

important cases, such as mental illness, the psychological function of sleep, and child development. Indeed folk psychology as a research project, which has its origins in ancient Greece, has been failing for over two thousand years. Further confirmation that folk psychology is redundant is found in the fact that it does not cohere with the current canon of scientific theories, including particle physics, atomic and molecular theory, organic chemistry, evolutionary biology, physiology and materialistic neuroscience. While eliminativists such as Dennett see some role for folk psychology's continuation in the future, the Churchlands believe that its obvious explanatory poverty should spell elimination. This presents a problem for them, as it is not clear how we could communicate without a hopelessly uneconomical and complicated reliance on neurological facts. We are not all neurologists.

At first glance it seems strange that a group of philosophers of mind should become famous for denying the existence of minds. What does this leave them with? The answer is philosophy, which they view as complementary to science, and they hold that it is only through neuroscience that philosophical problems concerning consciousness can be solved. Indeed, Patricia Churchland's *magnum opus* is entitled *Neurophilosophy* to emphasise the synthesis of philosophy and neuroscience. She identifies a further candidate for elimination in her article 'The Hornswoggle Problem' (1996). In this piece of work she berates the false distinction between the difficult, philosophical question concerning the nature of consciousness (termed the 'hard problem') and all other unsolved mysteries in neuroscience (termed 'easy problems'):

> Although consciousness is, certainly, a difficult problem, difficulty *per se* does not distinguish it from many other neuroscientific problems such as how the brains of homeotherms keep a constant internal temperature despite varying external conditions, or the brain basis for schizophrenia and autism, or why we dream and sleep.[7]

The author invites us to imagine a ledger of two columns entitled respectively 'hard problems' and 'easy problems'. To cite only one entry, i.e. the philosophical problem of consciousness, under the heading 'hard problems', while listing every other research project concerned with the human brain in the other column is, for the author, ungrounded and ill-informed. If one studies scientific progress, then it is reasonable to assume that thirty-first century neurobiology will have explained what consciousness is. Not to accept this position is to commit the informal fallacy *argumentum ad ignorantiam* (an argument based on ignorance) as one asserts:

1 Phenomenon P (in this case the mind) is not understood.
2 Therefore, we know that phenomenon P can never be understood.

This argument is fallacious because we cannot say that we know that we will never know at all.

Criticisms of eliminative materialism

Eliminativists seem to commit two informal fallacies themselves: firstly, in defending their standpoint with an argument from analogy, e.g. in comparing consciousness to witchcraft; and, secondly, in presuming that the answer to the problem of consciousness is a physical one. Like many physicalists who approach the problem of personal identity, an unfounded shift is made by moving from asserting (1) I am a physical thing to the conclusion (2) I am essentially a physical thing or I am only what I can be proved to be through science.

The argument from the conservation of energy

Eliminativists argue that if the physical domain is a closed causal network, then there is no place for non-physical entities such as the mind. In other words, there cannot be an interaction between mind and matter in terms of energy exchange, as dualists would have to assert. This is known as the argument from the conservation of energy and may be summarised as follows:

1 The total net energy of any closed system always remains the same (the law of conservation of energy).
2 The physical world is a closed system.
3 If the physical world interacted with something non-physical (e.g. a mind) then its total net energy would increase or decrease.

4 Therefore, the physical world cannot interact with anything non-physical.

This seems like a good argument, but it isn't. We do not know that the physical world is a closed system; science has not shown that every physical event must have a physical cause. The argument from the conservation of energy demonstrates the confidence that eliminativists have in science to provide a Theory of Everything (TOE). This is by no means a clear-cut assertion.

A further criticism, which is often dismissed by philosophers as *ad personam* (an argument based on personal interest), is that a theory such as eliminative materialism is counter-intuitive. Many feel that if they accept such a clinical diagnosis of the human condition then life would lose its value. How was it possible, they might argue, for Paul and Patricia Churchland to fall in love? Was their union sealed in a heady biochemical surge of testosterone, serotonin and nothing more? One physicalist theory that leaves the door ajar for qualitative experience is biological naturalism.

Task

Imagine a society where any references to minds or emotions have been eliminated and replaced with scientific jargon. Write a short dialogue between two lovers replacing any expressions of feeling with scientific language. Do we lose anything by changing register from the romantic to the biological? If so, then what is lost?

Biological naturalism

The description 'biological naturalism' refers to the fact that this school of thought holds that consciousness is completely explicable in terms of the physical sciences. Biological naturalists believe that the question 'What is the mind?' cannot be separated from the question 'What is human nature?' Their answer casts us in the role of 'human animals' who interact with our environment, reproduce, eat, sleep and perform other biological processes. For biological naturalists, our mental life is as natural a part of this material existence as respiration or defecation. Two arguments are employed by biological naturalists to defend this position.

The first is an argument from evolution that states that consciousness, like any other biological process, is the product of evolution, and that it is this fact that enables us to account for the complexity and sophistication of the human brain. Expressed formally it would run:

1 Every aspect of human beings (like other animals) has purely physical origins.
2 Consciousness is an aspect of human beings.
3 Consciousness has a purely physical origin.

Secondly, biological naturalists distinguish between human activity on a macro-level, such as social behaviour, and human activity on the ordinarily unobservable level of our micro-particles, which includes consciousness. They maintain that the workings of the mind can be explained most effectively on a cellular level and that we ought to resist the temptation to search out non-scientific explanations for macro-level activity. A comparison is drawn between hydrogen and oxygen combining on a molecular level to create the feeling of wetness on a macro-level. The same can be said of C-fibres firing on a micro-level to produce feelings of pain on a macro-level scale.

Biological naturalists such as John Searle (born 1932) attempt to provide a holistic account of human thought that is neither reductive nor eliminative, but includes the distinctive quality of experiences and acknowledges that perceptions have representational content. Consider what happens when someone looks at a bright light. The light entering the eye causes a reductive chemical reaction in the cells of the retina on which it falls; the chemical reaction causes an electrical current in the neighbouring nerve cells; this current stimulates more nerve cells in the brain and elsewhere in the nervous system; some of the nerve cells stimulated are connected to muscles, for instance in the iris, and the current in them causes these muscles to contract. A chain of cause and effect is generated, with each event being the cause of further events. The cells of the retina, nervous system, muscles and so forth are like components in a very complex machine that enables the human body to respond in a more or less sophisticated way to stimuli. In fact, machines can be built to mimic some features of

the human body. Certain cameras, for instance, can automatically match the size of their aperture to the intensity of the surrounding light, just as the human eye does. But something else happens when a person looks at a bright light, provided that the person's sight is not impaired in any way. The person 'sees' the light. They have something we call a visual experience.

We know what such experiences are like because we are having them all the time. Biological naturalists assert that, whereas scientists such as anatomists, biochemists and neurologists can explain how we perceive things, there is in addition a value-added component, which we term 'consciousness'. This remains mysterious at present, but science will eventually be able to give an account of what it is. Biological naturalism is an attractive theory that maintains a degree of scientific credibility (although there is some disagreement as to what constitutes science) while admitting that there is a quality about experience that is part of the distinctly human aptitude to understand the representational content of thoughts. The term used by philosophers to describe the representational content of thought is 'intentionality', and Searle argues that this quality cannot be reduced to a physicalist, non-philosophical explanation and that thought-content implies qualitative experience.[8]

In some ways, John Searle is the natural successor to Descartes, in that both authors compare and contrast human intelligence to non-human intelligence. In the case of Searle it is **artificial intelligence** and in the case of Descartes it is animal intelligence. Indeed, it is this fact that leads us to the accusation that biological naturalism is really dualism in sheep's clothing. Biological naturalists want to remain scientifically credible while acknowledging subjective experience (or what philosophers term **qualia**) and intentionality. This standpoint does not escape the mind–body problem. Indeed, the similarity is clearly seen when one concedes that any coherent dualist philosophy of mind would never deny that consciousness is inextricably linked to physical brain activity. Descartes' view of science might have been the mechanistic, clockwork conception promulgated in the seventeenth century but it was, nevertheless, an attempt to explore the relationship between thought and matter. In many of the most important respects biological naturalism is indistinguishable from property dualism, as advanced by Nagel, who ends his essay 'What Is It Like to Be a Bat?' on a note of scientific optimism while terming his position 'dualist'.

This highlights a depressing fact about the philosophy of mind. The fact is that thinkers are still searching in the dark for an elusive explanation, and have to date only offered an incomplete picture of how the brain works to which they attach (in the case of biological naturalism), superimpose (as in the case of supervenience theory) or dangle (in the case of Davidson's anomalous monism, which is discussed later in this chapter) subjective experience and the content of thoughts. It is hard not to read such attempts as nothing more than restatements of dualism.

Review question

Describe and illustrate one physicalist account of the mind.

Behaviourism and functionalism

Philosophical behaviourism

Philosophical behaviourism is from the same stable as logical positivism, and holds that we can only talk meaningfully about what can be verified. Behaviourists believe that due to the private nature of mental states philosophers of mind should concern themselves with behaviour. We can talk meaningfully about behaviour because it can be directly observed. In practice, what this means is that a word like 'pain' does not refer to some set of inner mental events but to a certain set of behaviour such as wincing, shrieking and hopping around. Psychological statements are verified by observing how people behave. It is important to note that philosophical behaviourists do not dismiss the idea of mental phenomena completely, and, while the theory is compatible with staunch physicalism, it does not entail it.

Knowing how and knowing that

Ryle sought to dismiss dualism, which he described in the first chapter of his classic study *The Concept of Mind*, as the dogma of the 'ghost' in the machine'. Yet Ryle runs into a problem in reducing everything to acts and utterances, as to how we might distinguish between 'knowing how' and 'knowing that'. The category of 'knowing how' is not problematic as it is concerned with practical, behavioural skills, but 'knowing that' appears non-behavioural and suggests a mind that does the knowing. Ryle attempts to meet this criticism by explaining away intelligence as nothing above or beyond knowing what is required in a given situation. 'Knowing that' is the act of having practised certain strategies and, as a consequence, having met certain performative standards. He states the following examples:

> It was because Aristotle found himself and others reasoning now intelligently and now stupidly and it was because Izaak Walton found himself and others angling sometimes effectively and sometimes ineffectively that both were able to give their pupils the maxims and prescriptions of their arts.[9]

Conditional statements

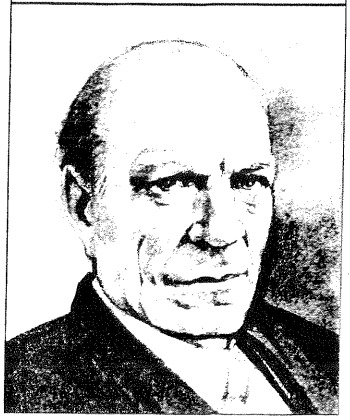
Ryle and like-minded philosophers set out to develop a theory that would allow one to paraphrase any sentence about mental states in a way that completely captures the meaning of the original but that makes no use of mental language. The fact that such an operation is possible in theory, though not necessarily practicable, holds philosophical significance for behaviourists. It would enable human beings to completely eliminate mental language, and with it the pseudo-questions that arise from a mistaken understanding of the function of such terms. The best way of seeing how the paraphrasing of mind-talk into talk about behaviour works is to cite an example.

Behaviourists seek to offer definitions of mental activity using 'If ... then' formulae. Consider the following sentence:

> If X were put in unsaturated water, then X would dissolve.

This is an operational definition of solubility, and similar operational definitions can be offered to explain behaviour:

Gilbert Ryle (1900–1976)



The most famous philosophical behaviourist was the Oxford philosopher Gilbert Ryle, although Quine and Wittgenstein (later in his career) had behaviourist leanings. Ryle was the epitome of a pipe-smoking, tweed-clad Oxford professor. The son of an Anglican bishop, Ryle studied as an undergraduate at Christ Church, Oxford, and stayed on as tutor in philosophy. He greatly aided A. J. Ayer in his career by persuading him to visit Vienna on honeymoon and to find out what the Vienna Circle was doing. He also introduced Ayer to Wittgenstein when they visited Cambridge. On another occasion reported by Ayer, the two scholars were driving across France on their way to a European philosophy conference when the conversation quickly dried up. In an effort to start it again, Ayer asked the shy professor whether he was a virgin, to which he answered 'yes'. Ayer then asked Ryle whether he preferred boys or girls, to which he replied 'probably boys' and remained in stony silence for the rest of the journey. Like Ayer, Ryle was a staunch advocate of analytic philosophy, and believed that many philosophical dilemmas were caused by logical or grammatical mistakes. He edited the philosophy journal *Mind* for nearly twenty-five years before his retirement.

Significant works

The Concept of Mind (1949)
Dilemmas (1954)
Collected Papers (1971)

If X were asked, then X would give answer Y.

According to behaviourists, every mental ascription is equivalent to asserting an 'If ... then' statement. The assertion 'Person X is thirsty' is equivalent to asserting 'If there were water available, then Person X would drink some'.

'If' clauses are always about stimuli, whereas 'then' clauses are always about behaviour. The fact that both stimuli and behaviour are physical makes this linguistic operation compatible with physicalist theories of mind.

In summation, any description about my beliefs can be paraphrased in a statement about how I would behave in various circumstances. Thus we can use the idea of dispositions to give a better statement of the behaviourist position, which is that the real function of mental language is to provide an economical way of talking about the ways people behave if they are in a given situation. According to the behaviourist, any sentence in which a mental term occurs can be paraphrased in a way that makes no reference to the mental but that instead refers to behavioural tendencies.

Criticisms of behaviourism

Behaviourism is an inadequate linguistic theory

The best that behaviourists can offer is an open-ended set of hypotheticals: 'If X were asked, then X may give answer Y.' Behaviourists cannot assert, as they claim, 'If X were asked, then X will/must give answer Y'. One could imagine a state of affairs where someone might be contrary, weak-willed, careless or having a siesta, and would not act in the way predicted. Behaviourists can only offer a 'for the most part' prediction. Imagine if in the above example an ascetic deliberately sought to be thirsty. The list of conditionals needed before we can make any kind of accurate prediction seem problematically long, open-ended and apparently unpredictable in the face of choice.

Pretence

A further problem with behavourism is that people can pretend to be in pain as well as actually being in pain. I can wince, shriek, hop around and, if my behaviour is all there is, then you would have grounds for supposing that I am in pain despite the fact that I might be malingering. Actually, this is not such a problem for behaviourists, as they include in the term 'behaviour' all sorts of reactions that are non-voluntary. To be in pain means not just crying out or limping but also evincing a range of physiological characteristics such as uncontrollable sweating or a high temperature. If we include all the non-voluntary facets of behaviour in our definition of pain, then the behaviourist becomes less vulnerable to the charge of deception through pretence.

Paralysis

A criticism which is more difficult to meet concerns people who are paralysed. Such individuals may not be able to move in even the simplest of ways, and they may need ventilator machines to keep them alive. From the behaviourist point of view, there is nothing to suggest that such people have any inner mental life at all. Yet we know that such individuals do have a mental life. The French journalist Jean-Dominique Bauby, for instance, after suffering a massive heart attack and slipping into a coma, dictated his reflections by moving his left eyelid, signalling his choice of letter as the alphabet was read to him; the result was his book *The Diving Bell and the Butterfly*. If behaviour is the sum total of meaning and in these cases there is no behaviour, then no account can be given about these patients with regard to mind or meaning.

A criticism from introspection

An odd consequence of behaviourism is that it seems to imply that if I want to know about my own feelings or beliefs I should not do so by introspection but by observing my own behaviour. Behaviourism seems to deny the 'inner' aspects of mental states, for example that pain has an intrinsic, qualitative, introspective nature separate from pain behaviour. Behaviourist assertions about dispositions seem frustrating in ignoring such 'inner aspects'. A strictly behaviourist account seems to omit meaningful distinctions such as the difference between being thoughtful and being taciturn, as the resultant behaviour remains the same. Talk about dispositions can only be hypothetical,

when something positive, categorical and explanatory is needed. Would behaviourists have to check in a mirror to see whether they were in a good mood or not? And would a behaviourist lover ask, 'I know it was good for you but was it good for me?'

Philosophical behaviourism is nevertheless of historical significance in the philosophy of mind and has left as its legacy an important theory known as functionalism.

Functionalism

Functionalists are concerned with the causal role of mental states, expressed in terms of cause and effect. They want to know what sort of job the mind performs. Functionalism retains the main insight of behaviourism, in preserving the link between mental states and behaviour. Yet the functionalist avoids some of the principal pitfalls of behaviourism in maintaining that mental states have an inner component. Functionalists, such as Hilary Putnam (born 1926), accept that mental language is about what it appears to be about, i.e. mental states. Functionalism attempts to give a complete account of mental events through highlighting the causal relationship between input, what may be called 'inner workings' or psychological states, and output. To make this clearer, let us consider how a function works.

Consider the workings of a security system. Having been burgled, you decide to install a security system, which involves lights on the outside of the house and sensors, which are activated by movement within a certain range. This system can be described in terms of its purpose or function as in figure 4.2.

Functionalists describe the workings of our brains in terms of a function. The experience of pain, for example, may be summarised as in figure 4.3.

Functionalists point to the fact that many things are functionally equivalent: one might, for instance, employ a watchful student armed with a hooter instead of installing a burglar alarm at great expense. In this case the student and the alarm are functionally equivalent. This presents a useful response on the part of functionalists to the problem of multiple realisability. They assert that it does not matter whether a creature's brain is made of something totally unlike our own; as long as it does the same job then it can be classed as a mind. Functionalism appears to have offered us a fairly holistic picture of mental properties despite its rather formulaic appearance.

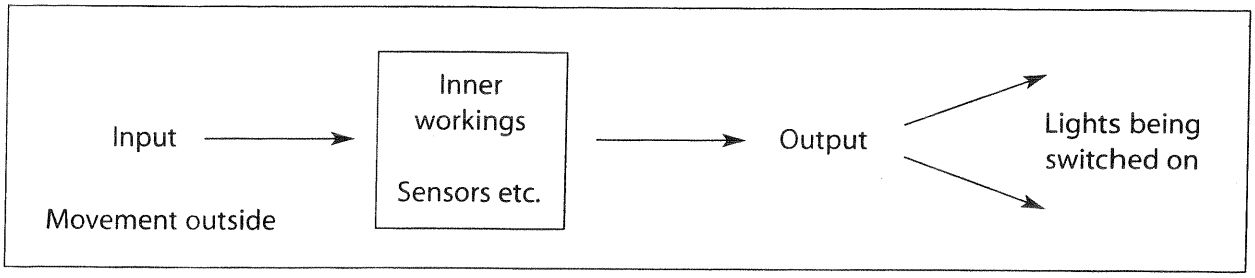


Figure 4.2 *Function of a security system*

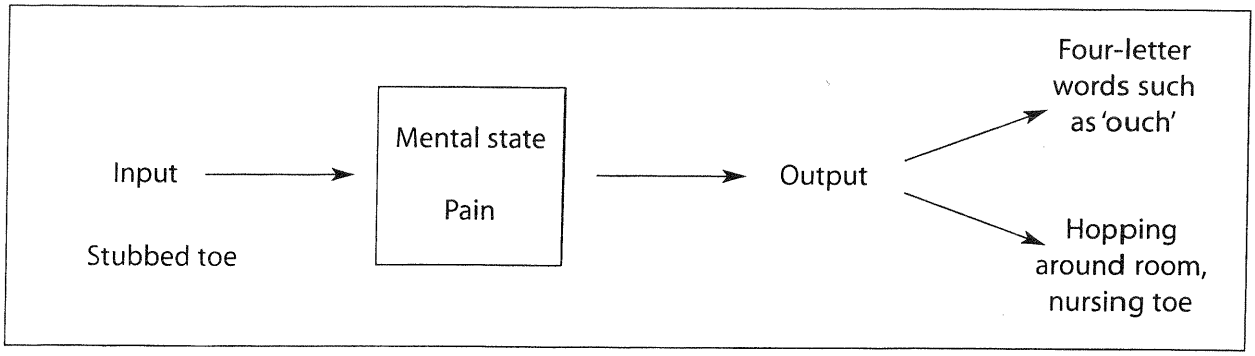


Figure 4.3 *The experience of pain*

Task

Complete the table below with likely scenarios to show the relationships between input, mental states and output.

Input	Mental state	Output
Tissue damage	Pain	Groaning, wincing, nursing injury
	Perception	
	Anger	
	Dream	
	Hope	
	Jealousy	
	Love	

Functionalists also find it useful to talk in terms of belief and desire, which are termed 'propositional attitudes'. Beliefs and desires seem to be inferred from an analysis of stimuli and behaviour. So, for instance, if I observe you stubbing your toe and hopping around I can justifiably infer that you are experiencing pain and desire not to have such an unpleasant experience again.

Criticisms of functionalism

There are two main criticisms of the functionalist account. The first is that the theory fails to account for the subjective 'feel' of these mental states. Philosophers term this ineffable 'what-is-it-like-ness' of an experience a 'quale' (plural 'qualia'). Secondly, functionalism seems comfortable with the notion that humans are functionally equivalent to machines, a view not shared by many thinkers. Let us explore each of these areas in more detail, starting with qualia.

Qualia

Qualia are the various subjective qualities in our sensations that cannot be expressed in causal terms. I might have studied types of exotic fruit but until I sink my teeth into a juicy guava I cannot claim to have enjoyed the qualitative sensory experience of tasting one. Qualia are derived from the senses, but are part of the process of introspection. They vary between individuals despite stemming from the same stimulus; thus, having halved the guava, I might offer you some, which you duly savour but appear to relish less than I do. We are experiencing different qualia, although there is a necessary connection between the properties of the fruit and the mental properties of our thoughts.

Mental states in possession of qualia may be categorised under the following headings:

- perceptual experience, e.g. seeing red;
- bodily pain, e.g. the itchiness of an itch;
- felt reactions, e.g. the pangs of jealousy;
- felt moods, e.g. a roller-coaster weekend of elation and despair.

Since ancient times, human beings have been intrigued by the possibility of other beings imitating humans in such a way as to remain undetected. Thought experiments surrounding the notion of qualia have considered the case of zombies and mutants that display all the roles and reactions that we do without any of the 'felt quality'. Such meditations are intended to raise the question as to whether any significance needs to be attached to qualia at all. Some philosophers such as Daniel Dennett have argued that qualia are nothing more than a philosophical myth, and consequently do not present any real problems for functionalists. However, the Australian philosopher Frank Jackson tells the following story in order to prove that qualia do exist.

The two Marys

Jackson was not convinced by his fellow Australian physicalists (although he has subsequently succumbed to their arguments), due to the felt quality of experience appearing to

be something more than the physical. He invites us to imagine that we live in an age where science has explained everything and in which there are no more mysteries. Imagine two scientists who are both called Mary, who have spent their lives in a strictly controlled room that displays only various shades of black, white and grey. The Marys are so pale themselves as to appear on this black and white spectrum and they dress in black and white clothes. Despite the absence of colour in their lives they have become world authorities on the physics of colour, and understand everything there is to know about wavelengths and the various chemical compounds needed to yield each subtle hue. Imagine what would happen if one of the Marys were suddenly released from her black and white chamber and shown a colourful object such as a vibrant red tomato! She would experience redness, rather than just know what the colour is in theory, and it is this personal experience, termed qualia, that presents functionalists with a problem. Let us call the cloistered Mary Mary 1 and the released Mary Mary 2. Mary 1 and Mary 2 are functionally equivalent as a computer and human brain might be, according to the functionalist account of things, yet their mental states are different. Mary 2 enjoys qualia, for example 'redness', yet for Mary 1 qualia are absent.

The possibility of inverted qualia is even more problematic for functionalists. This point has its origins in Wittgenstein's private language argument. Consider the possibility of two people's qualia being different.

Person	Quale observed	Quale termed
Bill	Red	Red
Ben	Green	Red

Both Bill and Ben are the same in terms of output in so far as they might exclaim 'I understand redness!' but they have different experiences. On the surface this argument seems straightforward, but the more you mull it over in your mind, the more you realise what difficulties it creates.

The Chinese room argument

In a similar vein, the American philosopher John Searle offers an argument from intentionality or 'aboutness' as a criticism of the functionalist view of intelligence. It is known as the Chinese room argument. Searle does not agree that humans and machines are functionally equivalent, due to the latter being incapable of feeling and understanding. The following thought experiment is intended to illustrate what is happening when a machine, e.g. a computer, is solving a problem. Imagine that it is someone's job to live in a room all by themselves and to reply to messages posted to them through a slot on one side of the room. The messages are written in Chinese characters and correspond to a set of replies, which are stored inside the room. The person who is replying, however, does not speak Chinese and only recognises which replies to post back by virtue of the same squiggles appearing on both message and reply. The worker can become really fast at posting the right answers to the questions that he is receiving but can never learn the meanings of those symbols.

The point of Searle's thought experiment is that the inmate is imitating what goes on in the central processing unit (CPU) of a computer, and this symbol manipulation cannot lead to any kind of understanding of the characters that are being manipulated. Searle draws a distinction between the 'syntactic' manipulation of symbols and the 'semantic' understanding of ideas. Searle's Chinese room argument rests on the inadequacy of the person in the room to assign meaning to the symbols. Critics have tried to rectify this by building a bigger and better computational model called a Chinese gym, which enables the inmate to consult with a speaker of both Chinese and English, and eventually familiarise himself with what is going on.

Artificial intelligence

A central feature of functionalism is that human intelligence is viewed as functionally equivalent to other types of intelligence. According to functionalists, to deny this is to be guilty of speciesism in the case of animal intelligence or bio-chauvinism in the case of artificial (silicon-based) intelligence. George Graham illustrates what functionalists would consider as unfounded prejudice in the following story:

> Imagine that a creature from another planet lands in your backyard and tells you (in perfect English) that it is going to enroll for courses at Princeton University. Then the creature dresses up like a human (suppose that otherwise it looks most unhuman and that the creature is devoid of anything which we would classify as a brain) and enrolls at Princeton. Within three months it is awarded a Ph.D. in Comparative Literature! What should you say? Should you say that brains are essential for belief or that they are inessential?[10]

This story may seem outlandish to more down-to-earth readers, but it intends to make the philosophical point that it is conceivable for something to exist that does what we do but has a very different anatomy and psychology. Functionalists draw on arguments from technological development to conjecture that machines essentially behave like brains, and so have the potential to become conscious. If you think that conjecturing what computers will be like in the twenty-fifth century is mere science fiction, then more recent reports should be of interest. An article in the *New Scientist*[11] describes how artificial intelligence experts are aiming to produce self-managing, self-repairing computers by engineering the machines to mimic the biochemistry of genes and proteins. Such 'software enzymes' evolve in the same way as biological enzymes, suggesting a future possibility of computers that grow to be more intelligent than their designers. Indeed, the American artificial intelligence expert Marvin Minsky believes that computers will eventually dominate society and keep us on as their pets.

The Turing test

The myth of Pygmalion[12] describes a statue coming to life and winning the heart of the eponymous hero. Of course the statue wasn't truly human, but its successful deception raises important questions. If X looks like a human and is functionally equivalent in terms of roles and reactions, then how are we justified in not classifying X as one of us? The same dilemma is faced in differentiating between carbon-based human intelligence and silicon-based artificial intelligence. If computers can do everything that we do then why not include them in our community of intelligence?

Alan Turing proposed in the journal *Mind* (1950) that if a computer could communicate with a human being in such a way that the human being could not tell the difference between conversing with the computer and with another human being, then it would be arbitrary (mere bio-chauvinism) to deny that the computer can think simply because it is a computer. By following an algorithm (a series of simple, discrete steps), computers can perform many of the functions that humans perform. What then is the difference between brains and motherboards, the computational equivalent of brains? Turing suggested that a two-way input–output device be set up between a person and two rooms. There would be a human being in one of the rooms and a pre-programmed computer in the other. If the person outside the rooms who was communicating with the inhabitants of each could not decide which room contained the human and which the machine, then one is forced to admit that there is no functional difference between them, at least inside these narrowly defined parameters.

Task

Are there any absolute differences between minds and machines?

Alicebot was a winning attempt at satisfying the Turing test. Search for winners of the Loebner prize (this is awarded each year for the program which comes closest to passing the Turing test) on the Internet and converse with artificial intelligence. Record what differences you notice between talking to Alicebot and a human. Vague, irrelevant and incoherent responses do not count as differences! Draw up a list of general differences between minds and machines. For example, the philosopher Hubert Dreyfus has suggested that 'common sense' is an attribute of the minded but not of the mechanical.

Review questions

1 Outline and illustrate the view that talk about mental states is talk about actual or potential behaviour.

2 Outline and illustrate what is meant by the term 'multiple realisability'.

Problems in the philosophy of mind

Let us next examine the four main problems encountered by theories of mind: the mind–body problem, the problem of other minds, the problem of self-ascription and the problem of personal identity.

The mind–body problem

Questions concerning the relationship between the mind and body are controversial from the outset, as they assume the existence of minds. Hard-nosed physicalists and behaviourists with materialist leanings dispute the existence of minds, and as a consequence assert that there is no such problem. But for those thinkers who do posit a mind, two central questions are raised:

- What is the mind? This is the problem of classification.
- What is the mind's relationship with the body? This is the problem of interaction.

There are many more tributary questions feeding into these main headings. One might ask when classifying the mind: What is it made of? What size is it? Where is it located? Similarly, when exploring how the mind relates to the body one might ask: Does the mental depend upon the physical or vice versa? Where do thoughts come from? What gives thoughts their content? Any cursory look at the mind–body problem will yield the conclusion that there isn't just one problem but many. Looking at their resolution helps to clarify the philosophical views already discussed in this chapter. These views are represented in the following table.

Task

For each of the theories listed across the top of the table below, write a paragraph summary of its response to the mind–body problem. Which do you find the most convincing reply?

The mind–body problem has spawned other interesting responses, one of which is called anomalous monism and is an account of the mind–body interaction proposed by the American philosopher Donald Davidson (1917–2003). As its name suggests, the theory holds that there is only one existent substance; Davidson believes that this is physical. The description 'anomalous' derives from the Greek *anomie*, meaning 'lawlessness', and refers to Davidson's conviction that there are no law-like connections between thoughts and brain states. Davidson believes that given the uncertainties and plasticity surrounding consciousness we can only talk about accidents, coincidences and exceptions.

As a term, 'consciousness' is often used in the philosophy of mind; yet it requires further analysis. It may accurately be defined as an awareness of external phenomena, an access to memories and an ability to monitor one's feelings. In addition, theorists have drawn a distinction between intransitive and transitive accounts of consciousness. The former refers to the sense of being awake or aware of one's faculties, and may be termed 'self-consciousness'; whereas the latter may be characterised by statements that take the form, 'I am conscious that there is something there'. Two rival camps emerge in the debate as to how consciousness

	Cartesian dualism	**Identity theory**	**Eliminative materialism**	**Behaviourism**	**Functionalism**
Do thoughts exist?	Yes	Yes	Brainwaves exist	Most say no	Yes
How are thoughts caused?	No answer	Possibly physical	Brains	Caused physically	No answer
What does it mean to have a mind?	To possess a soul?	To possess brain states	Nothing	To behave	Input relates to output
Does the mental depend on the physical?	No	Yes	Only the physical exists	Yes	No answer
Do thoughts have both mental and physical properties?	No, only mental	Yes	No, only physical	Yes	No answer

may be explained. The first is the view that consciousness is a mysterious, non-physical substance along the lines of an aura or soul, which does not seem to fit into any of the physical stories we tell about what goes on in the body. This theory, sometimes called transcendentalism or mysterianism, holds that conscious experiences occur in the mind and, in some more extreme versions, suggest that God is responsible for the mind's interaction with the body. The best that can be said for this view is that it is an ignorance-avoidance theory that does not contain any logical contradictions. The contrary view is founded in physicalism and holds that consciousness is nothing more or less than brain waves. An example of this might be the token-identity theory (see p. 79).

The problem of other minds

We have direct experience of our own sensations, feelings and perceptions. If we consider the experiences of others, however, we do not have access to such directness. This gives rise to the problem of other minds, which constitutes a sceptical challenge that states that we can never know whether other human beings have minds or whether their inner mental life is the same as our own. We become aware of other people's feelings and thoughts through observing their behaviour; from this we infer whether they are in a good mood or not. Behaviourists boast that their theory renders the problem of other minds redundant, but, as we have seen, there are many other difficulties with behaviourism, such as pretence and deception. And any reference to predisposition is as hard to prove as the existence of other minds.

On the face of it, these ideas may seem eccentric, but to dismiss the problem of other minds on the grounds that it does not fit in with our common sense is not to give a reasoned refutation of it. The problem of other minds, as one philosopher[13] has pointed out, sheds uncertainty on whether we can establish the minded status of aliens, animals and automata as well as other human beings. Expressed more formally and with specific reference to other human beings, the contention runs:

1 I cannot know exactly what sorts of experience other people have but can only observe their behaviour.
2 There is no necessary connection between mental states and behaviour.
3 We cannot possess knowledge of another person's mental states.

This conclusion leaves the door open for some hard-line sceptics to advance an extreme position known as solipsism (see pp. 16–17). Certain solutions have been suggested; one of the most popular until recently was the argument from analogy.

The argument from analogy

Traditionally, philosophers have tried to meet this sort of sceptical position by arguing from analogy. Arguments from analogy rely on the comparison of similar things. If one thing is like another in some respects, then it is thought it will be like it in others. Since I am like other individuals in my appearance and behaviour, the argument from analogy suggests that it is reasonable to assume that other individuals have the same mental life as I do. So when they do something to themselves that would cause me pain, such as stubbing a toe, I believe that it is likely that they are in pain. One may back this assertion up with scientific evidence concerning their nervous systems. The argument from analogy may be summarised as follows:

1 I am familiar with my own mind and body.
2 I am familiar with other people's physical behaviour and aware that it resembles my own.
3 I can infer that other people's minds resemble my own.

Despite being a commonsensical argument, it runs into problems if one pushes the possibility that other people's minds don't resemble one's own mind. There are many possible entities that are minded but do not resemble the human mind, e.g. gods, floridly psychotic human beings, animals and aliens. Why do proponents of the argument from analogy such as Bertrand Russell presume that, just because people's behaviour resembles his, their minds also resemble his mind?

Wittgenstein and language use

Ludwig Wittgenstein (1889–1951) was one of the most subtle and engaging thinkers of the twentieth century, and frequently turned his attention to questions in the philosophy of mind. His rebuttal of the argument from analogy has been labelled 'the beetle in the box criticism', in reference to a famous passage occurring in *Philosophical Investigations*:

> Suppose everyone had a box with something in it: we call it a 'beetle'. No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at *his* beetle. – Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. – But suppose the word 'beetle' had a use in these people's language? – If so it would not be used as the name of a thing. The thing in the box has no place in the language-game at all; not even as a *something*: for the box might even be empty.[14]

Let us consider an analogy along similar lines. Imagine entering a warehouse that is full of crates. All the crates look similar to each other in that they are all the same size and weight and have the same labels. We are curious about the contents of the crates, so we decide to open one up to see what is inside. By analogy from the one instance we have knowledge of, we may expect all the other crates to have similar contents to the one we have opened. However, any such assumption would be philosophically unfounded. Just because we know the contents of one of the crates does not mean that we can infer the contents of the others. The one difference between the case of the crates and our attempt to claim knowledge of the existence of other minds is that, in the case of the crates, we at least have the possibility of opening up a few more to confirm our suspicions. In the end, we may even decide to open all the crates and make absolutely sure that the contents of the first, second or third are substantially similar to all the rest. In the case of minds, we are only able to open and directly look into the one 'crate' that is our own mind. As far as other people are concerned, we have to rely on the markings on the sides of the crates or, to put it another way, the behaviour of other people. Wittgenstein goes further in inviting us to imagine lots of people peering into the various crates and advancing descriptions about what they contain, but having no method of verifying whether they are talking about the same thing. Such descriptions do not abide by the rules of normal language, according to Wittgenstein, and it would be irresponsible to draw conclusions from it. The argument from analogy does not solve the problem of other minds.

This depressing thought is salvaged by Wittgenstein when he goes on to show that the existence of language implies the existence of other minds. According to Wittgenstein, words have meanings because of the rules that are associated with their use. The word 'chair' can be used correctly or incorrectly depending on whether we use it within the bounds of the rules for its application or not. If I assumed that I was the only being in the universe then it would be impossible for me ever to be certain that I was using the words of my language correctly. All that I would have to go on to tell me that I was applying a word today in the same way that I applied it yesterday would be my own memory. To use Wittgenstein's own example, I would be like a person who in seeking to confirm his opinion that a train arrives at a particular time could only do so by referring to his own memory rather than to a timetable that exists independently of that memory. The private language argument attempts to show that built into our language about our own consciousness is a set of assumptions about the existence of other minds, and that to raise sceptical doubts about the existence of such minds is therefore to undermine our ability to say anything not just about other people's experience but about our own.

Wittgenstein believes not only that any private language invented by oneself is impractical, but also that it is an impossibility. Why is private language impossible? First of all, we would not have the power of mind to create it on our own. More significantly, any language could not be verified or falsified if we were the sole language-user. He compares this state of affairs (*Philosophical Investigations* (265)) to someone buying several copies of the same newspaper to check the validity of stories they read in the first. Equally, there could be no sustained rules of the

language game, and nothing would stop the solitary language-user changing the rules at will. But language does have rules, and the fact that we infer from such rules that other minds exist depends on the use of language and not its meaning. For Wittgenstein, meaning is use; thus to talk meaningfully about love stems from the fact that we have learnt how to use the language of love. We have been taught by other language-users the words for certain experiences and sensations. We have been able to compare these descriptions with the descriptions from people around us. We have been able to measure the intensity of these feelings, and have been corrected by others when we have made mistakes in describing our story. Such language use would be impossible without other minded individuals.[15]

The problem of self-ascription

Imagine that you are standing in front of a mirror, looking at your reflection. You might assert:

Ludwig Wittgenstein (1889–1951)

Wittgenstein has been hailed as the most important philosopher of the twentieth century and someone who breathed fresh ideas into the subject. He remained concerned with how language represents both ideas and the external world throughout his life, although he changed his approach after publishing the *Tractatus*, leading commentators to refer to early and later Wittgenstein. He is remembered for many things: as the heir to a multi-million-pound fortune inherited from his steel baron father; as someone who gave up his inheritance to ensure the security of his sisters from Nazi anti-semitism; as an ascetic who renounced attachment to material possessions and sought solitude variously in Ireland, Iceland and Norway; as a talented mathematician and logician; as an architect, inventor, choleric primary school teacher (the only work other than the *Tractatus* published in his lifetime was a children's dictionary); as a suicidal, tortured genius with ambiguous sexuality. One might say that he had it all. He is most famously associated with the Cambridge school of philosophy, and is considered as the immediate successor to Bertrand Russell. He found the atmosphere at Trinity College precious and restrictive, but inspired a band of acolytes who dressed as he did with open-necked shirts and adopted his militaristic, purposeful walk and mimicked his gestures, contorted grimaces, catchphrases and Austrian accent. It was at a famous meeting of the Moral Sciences Club at King's College in October 1946 that Wittgenstein was said to have threatened the visiting speaker, Karl Popper, with a poker. Reports differ as to the accuracy of this story, and it is noteworthy that this encounter took place in a dim, firelit room where most people were smoking. Hence it must have been hard to ascertain important details. If a reliable account is unobtainable, the story attests to the fiery, domineering egomania that made Wittgenstein such a charismatic and ebullient character. He sadly died of cancer in 1951. During his illness he would regularly take a walk, despite his growing frailty, to visit a tree which grew near Churchill College. He found comfort in hugging this tree, and when it was cut down to make room for a car park was heard to exclaim 'My tree, my life!' The famous philosopher died the day after. For many, Wittgenstein exhibited an other-worldly intelligence and mysticism; yet for many of his critics he was a confidence trickster who left philosophy no richer by his presence.

Significant works

Tractatus Logico-Philosophicus (1921)
The Blue and Brown Books (1958)
Philosophical Investigations (1953)

1 I have grey hair.
2 I am feeling contented.

Both statements may be described as 'self-ascriptions', but whereas the first statement is a bodily self-ascription the second statement is a mental self-ascription. Our concern in this section is with the latter.

Many philosophers believe that we can self-ascribe about perceptions, feelings, memories and the like through an inner sense. One is present to oneself as an object of introspective self-awareness. On the basis of this self-awareness any first-person statement such as 'I am feeling contented' is immune from error through misidentification. It is impossible to be wrong about your own mental state. This conviction is linked to three notable doctrines in the philosophy of mind: immediate, privileged and infallible accessibility. Let us define each in turn:

- *Immediate accessibility* is the first-hand, transparent quality of self-awareness. It would not make sense to assert, 'I don't know whether I am contented'.
- *Privileged accessibility* refers to the fact that I am in a uniquely authoritative position to comment on my own mental state. This is not enjoyed when commenting on other people's mental states (the problem of other minds).
- *Infallible accessibility* refers to the truth value of my mental self-ascriptions. By virtue of the fact that I am the subject of experience X, I know experience X is true.

How then is it possible to question self-ascriptions and why is it a problem? The problem lies in the fact that mental self-ascription carries with it the baggage of a dualist conception of the self. Hardline sceptics such as Lichtenberg (in his criticism of Descartes) and Hume (in his bundle theory) deny the existence of a self and as a consequence deny the validity of self-ascription. Thus the 'I' in the statement 'I am feeling contented' does not correspond to any independent reality but is just a feature of language. A weaker sceptical argument holds that a self exists but is caught in an introspective twilight and thus cannot pronounce on mental states. The fact that many mental states are unconscious may be cited in support of this argument. The line being advanced here is that any view of introspection as a searchlight that illuminates our experiences with exact detail is plainly not true. Such an argument fans the flames of Colin McGinn's conviction that we cannot meaningfully advance theories about our own mind as we only have our mind as the medium of investigation.

Colin McGinn (born 1950) argues that human beings cannot understand how the mind works, as the only tools they have at their disposal are minds, and this creates a cognitive closure. We cannot understand how technicolour phenomenology arises out of soggy, grey matter, or as he puts it, 'how the water of the physical brain is turned into the wine of consciousness'[16] because we are like a five-year-old child trying to learn about the theory of relativity. We do not have access to any mind–brain link, as we cannot get out of our minds.

Task

Discuss possible answers to the following questions on self-ascription:

1 What would a behaviourist say about mental self-ascription?
2 Do animals self-ascribe?
3 At what stage do humans gain the power to self-ascribe?

Review questions

1 Outline and illustrate one response to the mind–body problem.
2 Outline and illustrate one response to the problem of other minds.

The problem of personal identity

In essence, philosophy is an exercise in self-reflection: How should I behave? What do I know and believe? And more fundamentally still, what is it to be human? Not surprisingly, physicalists respond to this last question in a physicalist vein and dualists in a dualist one. Respectively, one might call these physical and psychological accounts of personal identity. There is a third approach, most recently advocated in the philosophy of mind by Derek

Parfit, which is termed 'bundle theory'. The problem of personal identity comes to a head in examining philosophical difficulties in establishing our continued identity through time. Is there an enduring self? Scientists inform us that over a period of time all the molecules in our body change. Cells are replaced in the gut after two days, red blood cells after one hundred days and brain cells over a period of years. If Sleeping Beauty fell asleep for seven years then all of the cells in her body would have been replaced by the time she woke up. So would she remain the same person? Can philosophers suggest watertight criteria that prove she was the same person at time 1 (t_1), when she was put under the spell of the evil witch, as she was later at time 2 (t_2), when kissed by the handsome prince?

Physical continuity

Physicalists maintain that we recognise people as individual and distinct by physical criteria of identity, in other words that they are 'flesh and blood'. One might summarise the argument:

1 Personal identity consists in my being a flesh-and-blood body and being biologically alive.
2 There is continuity between my being alive at t_1 and t_2; therefore, I am the same person at t_1 and t_2.

One might even take one step further and cite DNA as the best criterion for physical continuity, with the unique hallmark of our existence contained in each of our cells. Yet our physicality can be manipulated, either in the realm of science, or at least through the philosopher's imagination in the realm of science fiction. One might conceive of someone who had their DNA removed from each cell, leaving the cell structure intact, before someone else's DNA was inserted. Would they remain the same person? This operation might seem highly improbable, but it illustrates the problem of advancing purely physical grounds for personal identity.

One refinement, advanced by Thomas Noonan, of the physical continuity theory admits that any person at t_1 and t_2 does not have to be materially identical, but suggests that all the matter constituting the person at t_2 needs to have resulted from that constituting the person at t_1 by a series of more or less gradual replacements. This still creates an identity crisis for recipients of organ donations, and raises the question of whether those people in the future who have received transplants from other species will be classed as only partly human.

Psychological continuity

Theories of psychological continuity inherit a long tradition in western philosophy, stretching back to Plato's *Phaedo* and including theistic notions of body and soul, with the latter possessing the ability to survive death. Later in the tradition, atheistic existentialists argue that if we are free to create ourselves as works of art then this unique blend of character traits and the ability to choose which course of action to take constitutes ourselves. Memory has been hugely important in all psychological accounts of personal identity and may be outlined as follows:

1 Personal identity consists in my remembering the events that have occurred throughout my life.
2 There is a continuity in my memory at t_1 and t_2; therefore, I must be the same person at t_1 and t_2.

Certain criticisms challenge this notion:

- Memory can be manipulated through hypnosis or suggestion. Conceive a Ministry of Truth, in the spirit of Orwell's *1984*, which manufactures a series of photographs with you airbrushed in, despite the fact that you weren't there. Through such devilry, the men from the Ministry set about conditioning you to believe that you were present. Memory can be artificially manufactured, and as a consequence seems a weak candidate for identity.
- Memory can also be lost, as in the case of amnesia or just general forgetfulness. Thomas Reid (1710–1796) offers the following story in disagreement with an argument from memory proposed by John Locke (1632–1704). There was a soldier who, when he was a boy, had stolen apples from an orchard and was flogged as a consequence. That same boy later became

a brave ensign who won honours in battle. Afterwards, the ensign gained promotion and subsequently became a famous general. If memory is supposed to be necessary for personal identity, what happens if the brave ensign can remember stealing the apples as a boy and the general can remember winning honours in battle but not stealing apples? Are we to say that they are not the same person? Clearly, personal identity is an all-or-nothing thing, but are we really expected to remember every experience? One might counter Reid's argument by asserting that there is still a continuity of connectedness between boy, ensign and general; therefore, they are one and the same person. Noonan suggested physical continuity based on gradual replacement; one might make a similar move here and advance an argument from overlapping memories.

- One might argue that it is wrong to force an 'either/or' decision between physical and psychological approaches. Instead, one might champion the existence of an indivisible unity between mind and body that constitutes the whole. Such a unity exists even in the case of multiple personalities, which pose an exception to purely psychological accounts; any such unity is vehemently denied by bundle theorists.

Bundle theories

The contemporary bundle theorist Derek Parfit contrasts both physical and psychological accounts, which he collectively terms 'ego theories', with a standpoint called the bundle theory. The thoughts we experience are not connected through a separate 'ego' but exist in a haystack-like bundle and are loosely linked through their contiguity in space and time. Parfit argues against the notion of a unified, underlying person. He states that when we talk of personhood we usually mean a single entity but that such an account is unintelligible. Parfit cites two cases of bifurcation as evidence: one physical and one psychological. Firstly, we are to imagine a case of artificial replacement when parts of our body are replaced or swapped. Secondly, we are to imagine certain facts about our brain (content, psychological make-up) being scanned and duplicated in an exact replica. In both cases, we do not remain a single, unified entity. Parfit argues that these cases prove problematic for ego theories, which posit an underlying self, but not for bundle theories, which define personhood in terms of a changing collection of thoughts, sensations and experiences.

In western philosophy, bundle theories are the natural heirs to Georg Lichtenberg (1742–1799) and Hume (1711–1776), but many counterparts exist in eastern philosophy, where the position is termed the doctrine of *anatman* (no self). The Buddha likens the mistaken belief in a self to someone who believes that a carriage exists in addition to its parts. Ryle would have said the person was making a category error.

As well as Buddhists, certain contemporary thinkers have denied that a self exists at all and have put forward the view that the idea is nothing more than the product of social convention. The psychologist Susan Blackmore (born 1951) has argued from a physicalist perspective that the notion of selfhood is a lie, an illusion or an evolutionary trick.[17] She argues that thoughts are generated physically, and that such beliefs as religion, science and the idea of personal identity have originated by accident but are so beneficial to the survival of the species that they are passed on like genes from person to person. These skills, habits and ideas that are imitated through the medium of language are termed 'memes', though it remains unclear why each person and their *a priori* receptiveness to memes cannot count as grounds for personal identity.

Revision questions

1 In what ways can the mental and the physical be distinguished?

2 Assess the strengths and weaknesses of eliminative materialism as a theory of mind.

3 Assess the problems involved in arguing from our own case to the existence of other minds.

4 Describe and illustrate three difficulties in establishing our continued identity through time.

5 Examine the difficulties that intentionality presents for theories of the relationship between mind and body.

Discussion questions

- Describe and illustrate what is meant by qualia.
- Assess the arguments for and against the survival of disembodied consciousness.
- Behaviourism, in none of its versions, can account for the fact that an inner mental life exists with no behavioural manifestations. Discuss.
- Assess the implications of artificial intelligence for the view that the mind is the brain.
- Assess whether any form of dualism offers a plausible account of the nature of mind.

Glossary of key terms

Artificial intelligence: the study of silicon-based systems that display skills such as problem-solving.

Eliminative materialism: the view that one can eliminate the concept of the mental without loss of meaning.

Epiphenomenalism: the dualist doctrine that the physical can cause the mental but the mental cannot cause the physical, e.g. you feel hot because you are hot not because you are thinking about heat.

Folk psychology: a pejorative term coined by eliminativists to describe homespun psychological theories that have evolved to explain other people's behaviour.

Intentionality: the content of a thought. Mental expressions have to be about something, and what a specific thought or belief is about is termed its intentionality.

Introspection: the process whereby we gain information about ourselves by thinking about our own thoughts.

Leibniz's law: Leibniz asserted that for two things to be the same then they must be numerically identical. This may be expressed formally: If every property of X is also a property of Y then X = Y.

Monism: the view that there is only one substance of which the universe is made. This may be mental, as in the case of Berkeley's immaterialism, or physical, as in the case of modern materialist philosophies of mind.

Physicalism: the group of philosophical theories which hold that questions concerning consciousness and the mind can be resolved by physical explanations.

Property dualism: the view, advanced by thinkers such as Nagel, that asserts that the mental and the physical exhibit different attributes and should, therefore, be considered as distinct.

Qualia: the qualitative, subjective characteristics of experience – for example, how something tastes to a particular person.

Supervenience: this refers to the process whereby a further layer of meaning emanates from particular entities. For example, if one calls someone's face cute, one is describing cuteness supervening or emanating from the particular features of the individual's face.

Suggested reading

- Churchland, P. (1999) *Matter of Consciousness*, London and New York, MIT Press
  A short volume introducing the various schools of thought, including the Churchlands' own eliminativism
- Graham, G. (2000) *Philosophy of Mind: An Introduction*, Oxford, Blackwell
  The most accessible introduction to the subject with interesting links to other philosophical issues
- Lyons, W. ed. (1995) *Modern Philosophy of Mind*, London, Dent
  An anthology of essays and articles covering a wide spectrum of thought
- Maslin, K. T. (2001) *An Introduction to the Philosophy of Mind*, Oxford, Polity
  An in-depth survey of all the relevant arguments
- Smith, P. and Jones, O. R. (1986) *The Philosophy of Mind*, Cambridge, CUP
  An undergraduate text on the philosophy of mind that nevertheless includes some useful introductory material

Notes

[1] Ryle. G. (1949) *The Concept of Mind*, Harmondsworth, Penguin, p. 25.

[2] Descartes, R. (2000) *Meditations and Other Metaphysical Writings*, Harmondsworth, Penguin, pp. 20, 22, 33, 40, 52, 56 and 58.

[3] Foster, J. (1991) *The Immaterial Self*, London, Routledge.

[4] Swinburne, R. and Shoemaker, S. (1989) *Personal Identity*, Oxford, Blackwell.

[5] Nagel, T. (1995) 'What Is It Like to Be a Bat?', in W. Lyons, *Modern Philosophy of Mind*, London, Dent, p. 160.

[6] Churchland, P. (1999) 'Eliminative Materialism and the Propositional Attitudes', in W. G. Lycan (ed.), *Mind and Cognition*, Oxford, Blackwell, p. 206.

[7] Churchland, P. (1996) 'The Hornswoggle Problem', Internet.

[8] Searle, J. (1994) *The Rediscovery of Mind*, Massachusetts, MIT Press.

[9] Ryle, G. (1949) *The Concept of Mind*, Harmondsworth, Penguin, p. 31.

[10] Graham, G. (2000) *Philosophy of Mind: An Introduction*, Oxford, Blackwell, p. 88.

[11] 'Evolve or Die', *New Scientist*, 27 October 2001.

[12] Ovid (1988) *Metamorphoses*, Harmondsworth, Penguin, Book X.

[13] Churchland, P. (1984) *Matter and Consciousness*, London, MIT, p. 67.

[14] Wittgenstein, L. (1958) *Philosophical Investigations*, Oxford, Blackwell, p. 100.

[15] For a detailed analysis of Wittgenstein's private-language argument see A. Kenny (1973) *Wittgenstein*, London, Penguin, pp. 178–202.

[16] McGinn, C. (1995) 'Can We Solve the Mind–Body Problem?', in W. Lyons, *Modern Philosophy of Mind*, London, Dent, p. 272.

[17] Blackmore, S. (1999) 'Meme, Myself, I', *New Scientist*, 13 March, pp. 40–44.