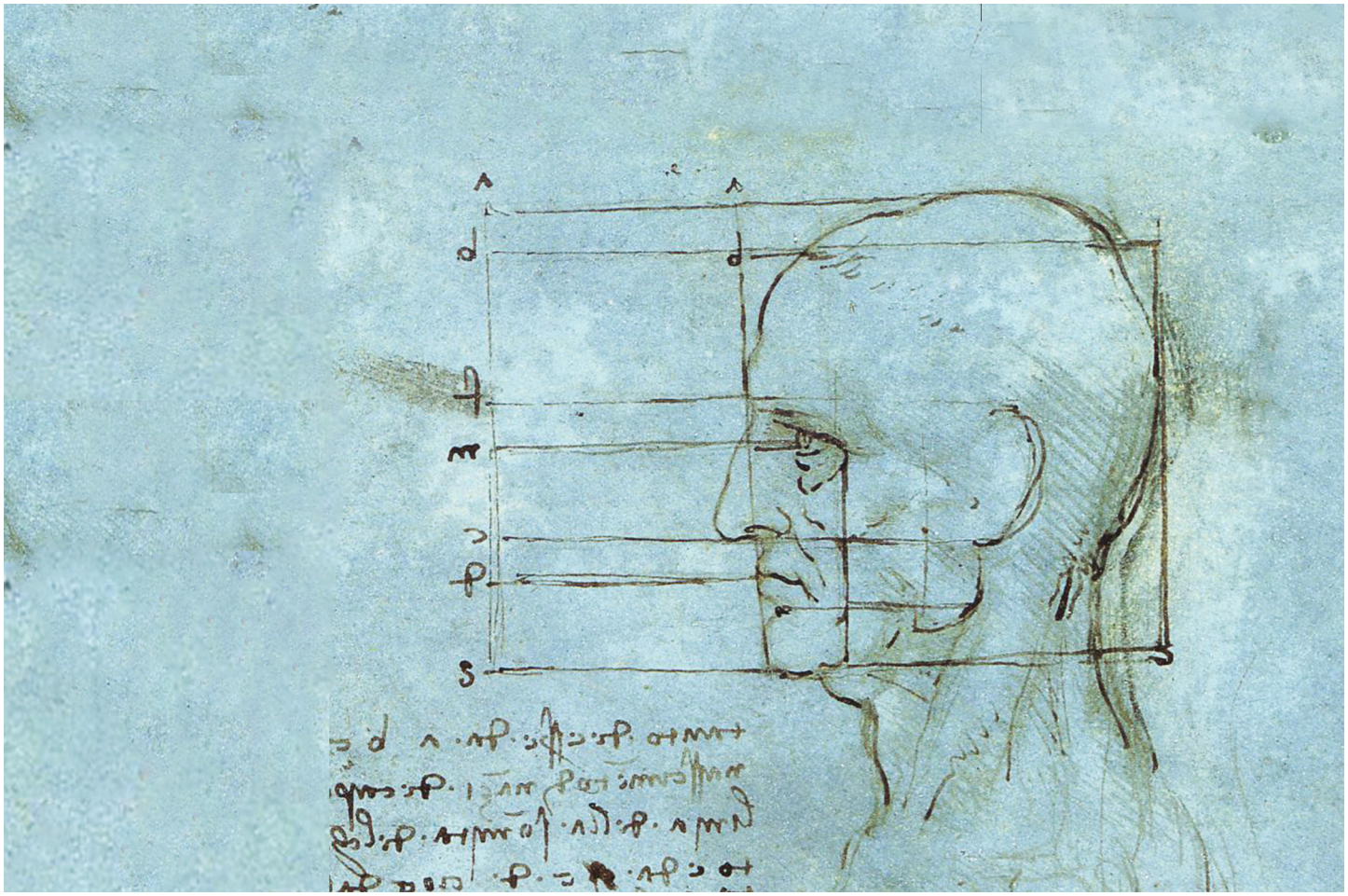
OpenLearn

The Open University

# Introducing consciousness



**About this free course**

This OpenLearn course provides a sample of Level 3 study in Arts and Humanities: http://www.open.ac.uk/courses/find/arts-and-humanities.

This version of the content may include video, images and interactive content that may not be optimised for your device.

You can experience this free course as it was originally designed on OpenLearn, the home of free learning from The Open University - www.open.edu/openlearn/history-the-arts/culture/philosophy/introducing-consciousness/content-section-0.

There you'll also be able to track your progress via your activity record, which you can use to demonstrate your learning.

Head of Intellectual Property, The Open University
Designed and edited by The Open University

# Contents

# Introduction

Consciousness is at once the most important and most baffling aspect of the mind. It is the very heart of our existence yet it is extraordinarily difficult to describe and explain. This course introduces consciousness, and the ‘hard problem’ it presents for a science of the mind.

This OpenLearn course provides a sample of Level 3 study in Arts and Humanities.The material is derived from an Open University course (AA308) that is no longer available. The content has been used as the basis for a section of Part 1 of the MA in Philosophy (A853) qualification.

Tell us what you think! We’d love to hear from you to help us improve our free learning offering through OpenLearn by filling out this short survey.

# Learning Outcomes

After studying this course, you should be able to:

- discuss basic philosophical questions concerning the nature of consciousness
- understand problems concerning the nature of consciousness and discuss them in a philosophical way.

# 1 Consciousness

> Consciousness: The having of perceptions, thoughts, and feelings; awareness. The term is impossible to define except in terms that are unintelligible without a grasp of what consciousness means.… Consciousness is a fascinating but elusive phenomenon: it is impossible to specify what it is, what it does, or why it evolved. Nothing worth reading has been written about it.
>
> (Sutherland 1995, 95)

Consciousness is at once the most important and most baffling aspect of the mind. It is the very heart of our existence – our 'self of selves' as Julian Jaynes puts it – yet it is extraordinarily difficult to describe and explain. This course is an introduction to this slippery phenomenon and the problems it presents. It is in three main sections. The first explains what contemporary philosophers usually mean when they talk about consciousness; the second examines the phenomenon in more detail and highlights some of its puzzling features; and the third sets out the central philosophical problem surrounding consciousness – the so-called 'hard problem' of explaining how it arises and whether it is a physical phenomenon.

# 2 Defining consciousness

## 2.1 Introduction

We use the words ‘conscious’ and ‘consciousness’ in a variety of ways. We talk of losing and regaining consciousness, of being conscious of one's appearance and of taking conscious decisions. We speak of self-consciousness and class-consciousness, of consciousness-raising activities and consciousness-enhancing drugs. Freudians contrast the conscious mind with the unconscious, gurus seek to promote world consciousness and mystics cultivate pure consciousness. These various uses reflect the history of the words. The original meaning of ‘consciousness’ was *awareness* or *knowledge*, either shared or private, and some of our modern uses reflect this. Self-consciousness is awareness of oneself as an individual; class-consciousness is awareness of belonging to a particular socio-economic group; to be conscious of one's appearance is to be very aware of it; and so on. In the seventeenth century, however, philosophers and other writers began to use the word in a more specific sense, to refer to our *inner* awareness of our own mental states – our perceptions, sensations, feelings and thoughts. As the philosopher John Locke (1632–1704) put it, ‘Consciousness is the perception of what passes in a Man's own mind’ (Locke 1961, vol. 1, 87). (Previously ‘conscience’ had been used in a similar way, but that word was coming to be used to refer to an inner moral sense.) Again, some of our modern uses reflect this philosophical usage. The conscious mind is the level of mental activity of which we are aware, in contrast to the repressed unconscious; consciousness-enhancing drugs are ones that alter our mental states in various ways; pure consciousness is mental awareness stripped of all particular content. When contemporary philosophers speak of ‘the problem of consciousness’ they too are using the term in broadly this sense, though with a subtle difference. In this section I shall explain in more detail what they have in mind.

## 2.2 What it's like

Suppose you have just had a dental procedure under general anaesthetic and are coming round. You are aware of a dazzling light above you and of a muffled voice echoing in your ears. There is sickness in your stomach and a sharp metallic taste in your mouth. You feel a moment of panic as you struggle to work out what has happened. Moving your head, you recognise the dentist's face and realise that he is speaking your name and asking if you want a glass of water. Your remember where you are, sit up shakily and take the glass.

Think about what happened as you regained consciousness. Various bodily processes resumed. Your sense organs started functioning again, registering stimuli and sending signals to your brain. Your brain also resumed its normal activity, processing these incoming signals and responding to them. Various brain centres became active, including ones devoted to visual processing, face recognition, emotion, memory, language and conceptualised thought. Signals flew back and forth from region to region and out to your organs and limbs. But this wasn't all that happened when you came round. You also started having *conscious experiences* – experiences with a certain *feel* to them. Imagine having the various experiences I described – seeing a dazzling light above you, hearing a

muffled voice, having a metallic taste in your mouth, feeling a stab of panic and so on. Focus on what it is *like* to have those experiences – on what it feels like from the inside. Each of them, like every other experience, has its own character, which is instantly recognisable but very difficult to describe.

Philosophers use a variety of terms for this aspect of experience. You will find them speaking of an experience's 'qualitative feel', 'phenomenal feel', 'phenomenal character', 'phenomenal content', 'phenomenology' (in some contexts), 'subjective character', 'raw feel', 'what-it-is-likeness' and 'qualia' (a Latin plural meaning 'qualities'; the singular is 'quale'). Some of these terms – 'qualia' in particular – often carry theoretical overtones, but at bottom all refer to the same thing: what a given experience feels like from the inside. When contemporary philosophers speak of *consciousness* it is usually this to which they are referring.

Another way to home in on the phenomenon of consciousness is to contrast conscious mental states with non-conscious ones. Although some philosophers (Descartes for one) have rejected the idea, it is now widely accepted that we are not aware of all of our mental states and processes. This view has been popular among psychologists since at least the nineteenth century, and everyday life provides plenty of evidence for it. Consider driving, for example. One can drive a car, drawing upon one's knowledge of the rules of the road and of the car's controls, without giving any conscious thought to what one is doing. Or think of cases where the solution to a problem pops into one's head some time after one has given up thinking about it consciously. In these cases it seems, some quite complex mental activity must be going on below the surface.

Many writers also hold that non-conscious *perception* is possible. At first sight this may seem a bizarre claim. How could we *see* non-consciously? The idea is not as odd as it sounds, however. One way to illustrate this is to think about a robotic system. Consider Cog. Cog is a robot that is being built by the Humanoid Robotics Group at The Massachusetts Institute of Technology, under the direction of Rodney Brooks (Figure 1). Cog has a mechanical body (only the upper part so far), powered by electric motors and controlled by microprocessors similar to those found in personal computers. It also has a visual system, consisting of two head-mounted video cameras and a network of microprocessors for analysing their signals. (I say 'it' because the MIT team deny that Cog has a gender.) This gives Cog some basic visual abilities. It can identify faces and other interesting objects, follow moving objects with its eyes and use visual information to guide its hands. But though Cog has a form of vision, no one seriously thinks that it has conscious visual *experiences* of the sort we have when we look at the world around us. We might say that it has *non-conscious* vision: it sees things, but its sight does not have any felt quality to it. The MIT team are also working to equip Cog with auditory and tactile sensors, but again no one expects these to provide it with conscious experiences of hearing and touch. (For more information about Cog, see Adams et al., 2000; Brooks et al., 1998; Humanoid Robotics Group, 2004, online.)

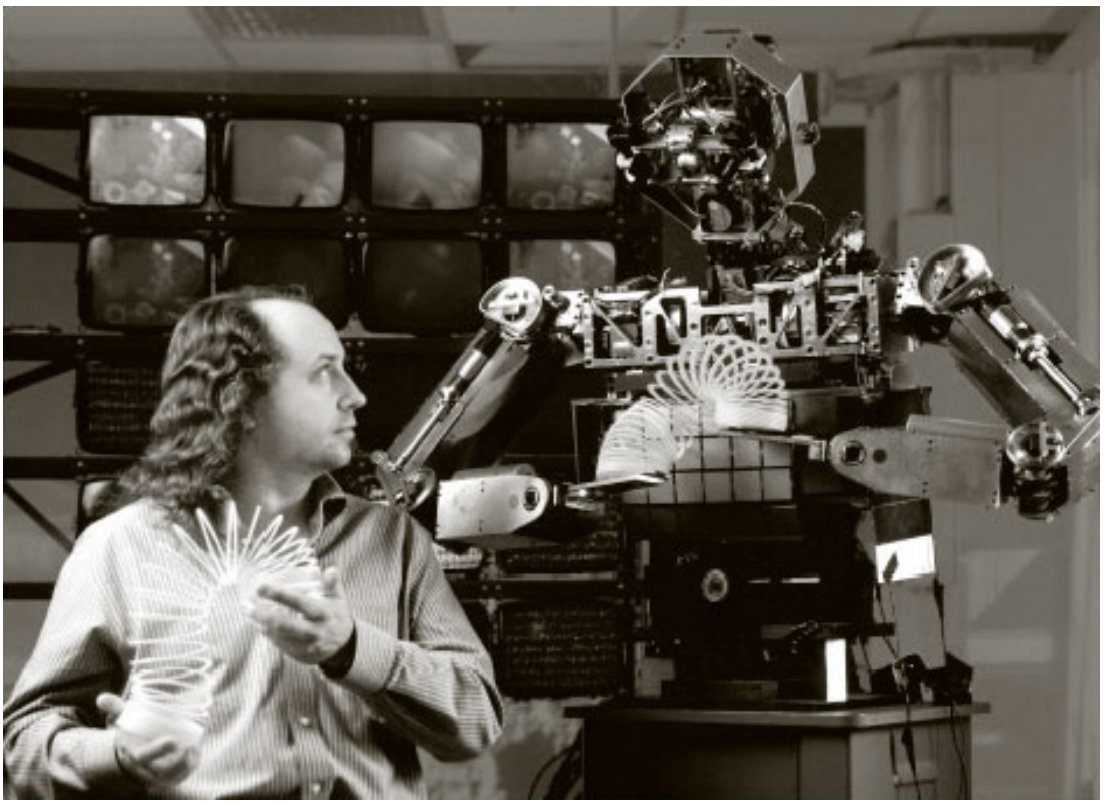


Figure 1 Rodney Brooks and his android robot Cog. Photo by Peter Menzel. Copyright © Peter Menzel/Science Photo Library

There are times when we seem to perceive things in a Cog-like way. Psychologists have shown that it is possible to influence a person's behaviour by means of stimuli that are not consciously perceived (Dixon, 1971). In a typical experiment a word is displayed for a split second, so that the subject has no conscious awareness of seeing it. In subsequent testing, however, the subject makes word associations which are influenced by the word displayed – revealing that they had in fact perceived it at some level. (This is known as *subliminal perception*.) A similar phenomenon seems to occur frequently in everyday life. When driving or walking along a busy street, we continually fine-tune our movements in response to visual cues of which we are not consciously aware – adjusting speed and direction to compensate for the movements of those around us. We also respond in this way to signals from our own bodies, shifting position to avoid cramp or to protect an injury, yet without consciously noticing any discomfort. In these cases, it seems, our brains are registering information and using it to control our behaviour, yet without generating any conscious perceptions or sensations. There are also pathological conditions that seem to involve non-conscious perception. The most famous of these is *blindsight* (Weiskrantz, 1986, 1997). People with this condition have normal eyes but have suffered damage to the visual processing areas of their brains, with the result that they seem to be blind in large areas of their visual field. They say – quite sincerely – that they see nothing in these areas. Yet if presented with an object in the blind area and asked to make a random guess as to its location or orientation, these people usually guess correctly – much to their own surprise when subsequently told the results. It seems that they are visually detecting the objects without any of the felt quality of normal vision.

(You may feel that it is twisting words to talk of non-conscious perceptions. Surely, *seeing* is by definition a conscious experience? This is really a terminological issue, however. If we use the term ‘perception’ in that way, then there are no non-conscious perceptions, just as there are no married bachelors. But it is compatible with this that there are non-

conscious mental states that are very *like* perceptions in the role they play, and calling them 'non-conscious perceptions' is a way of emphasising this. Some writers also talk of non-conscious sensations and experiences, and the same goes for those terms.)

These reflections on non-conscious mentality should help to clarify what philosophers mean when they talk about consciousness. Their focus is not on the nature of perceptions, sensations and thoughts as such, but rather on what is special about those perceptions, sensations and thoughts that have a feel to them. What exactly is this *feel* that conscious experiences have? How does it come about and what is its function? Whatever the answers, the phenomenon is surely tremendously important. As the American philosopher Thomas Nagel (b. 1937) puts it, to say that a creature has conscious experiences is to say that it is *like something* to be that creature – that it has an inner life (Nagel, 1974). A non-conscious being such as Cog might be able to perform sophisticated tasks, guided by information from its sensors, but without conscious experience it would have no inner life. It might detect colours and sounds, but it would never know what it was like to see a brilliant blue sky or to hear leaves rustling in the breeze. It might register when it was damaged or running low on energy, and take appropriate action, but it would never feel real pain or hunger. It might act like us, but it would be dead inside, without any of the conscious experience that accompanies our activities. Life without consciousness would not be life at all as we know it. Indeed, the philosopher Colin McGinn (b. 1950) suggests that the emergence of consciousness was an event of cosmic significance, analogous to the Big Bang. Just as the Big Bang created the physical universe, so the emergence of consciousness – McGinn calls it the 'Soft Shudder' – created a new dimension of mental reality (McGinn, 1999, 15).

### Activity 1

Which of the following do you think have conscious experiences in the sense outlined above: apes, dogs, snakes, fish, insects, bacteria, plants, rocks?

Here is my answer. I find it hard to doubt that apes and dogs have conscious experiences very much like ours. I am not sure what to say about snakes and fish. I am fairly confident that insects do not have conscious experiences and I am certain that bacteria, plants and rocks do not.

Your intuitions may differ from mine of course. You may believe that all animals, even insects, are conscious. (Indeed, as we shall see, some philosophers think that even rocks have a *little bit* of consciousness!) You may be right, but I think you should at least consider the possibility that you are wrong. The fact that animals *behave* like us does not prove that they *feel* like us too. It would be fairly easy to program Cog to detect when it was damaged and issue sounds resembling cries of pain, but it still would not have any conscious pain sensations. And it is possible that animals are the same.

A note on terminology. We are going to need a standard term to refer to the feel of conscious experience. None of the options is unproblematic: 'feel' is ambiguous, 'qualia' has theoretical overtones, and 'what-it-is-likeness' is cumbersome. I shall use 'phenomenal character'. Although the term may sound technical, remember that it denotes something quite simple – the phenomenal character of an experience is *what it is like* to have it. I shall also occasionally speak of an experience's 'phenomenal properties'; this means the same.

## 2.3 Some distinctions

I now want to distinguish consciousness, in the sense outlined above, from some related phenomena. This should help to clarify the concept further and avoid potential confusion. What follows draws in part on distinctions and terminology introduced by the philosopher David Rosenthal (Rosenthal, 1993).

The first distinction I want to make has already been introduced. When I described your experience at the dentist's I spoke both of *you* being conscious and of *your experiences* being conscious. These are different notions of consciousness, of course. When I spoke of you being conscious, I meant that you were awake, as opposed to being asleep or knocked out. When I spoke of your experiences being conscious, I meant that they were of the sort that have a phenomenal character to them. These two sorts of consciousness are sometimes referred to as, respectively, *creature consciousness* and *state consciousness* ('state' here means 'mental state'). Of course, in us creature consciousness involves possession of state-conscious experiences, but perhaps in other creatures the comparable condition does not. When a stunned fish comes round, does it start having conscious experiences? I do not know.

As well as talking of creatures being simply *conscious*, we also talk of them being conscious *of* particular things – as when we say that someone was conscious of a face at the window. This is sometimes referred to as *transitive consciousness*, since it is directed at an object. To be conscious of something in this sense is to be *aware* of it – to be perceiving it or thinking about it. Again, for humans this usually involves having a conscious experience of it, but perhaps for other creatures it does not. Thus, we might say that Cog is conscious of the people around it, in virtue of the fact that it detects their presence and responds to them.

The second distinction is between consciousness, in the senses just mentioned, and *self*-consciousness. Self-consciousness is awareness of oneself as an individual. Fully developed, it involves the ability to think about oneself as a thinking, feeling creature, with a history, future and unique perspective on the world. This clearly requires some conceptual sophistication and it seems unlikely that many non-human animals are self-conscious in anything more than a rudimentary way, even if they are fully conscious in the other senses. Self-consciousness raises some important and difficult philosophical issues, but they are tangential to our main topic and I shall not be discussing them in this course.

A third, more contentious, distinction is between kinds of state consciousness. Here the philosopher Ned Block (b. 1942) has argued that we should distinguish between what he calls *phenomenal consciousness* and *access-consciousness* – *P-consciousness* and *A-consciousness* for short (Block, 1995). Phenomenal consciousness is consciousness of the sort we have been discussing: a mental state is phenomenally conscious if it has a phenomenal character. Access-consciousness, on the other hand, is a rather different notion. A mental state is access-conscious if the information it carries is directly available to other mental processes, including reasoning, behavioural control and speech. Normally, of course, our experiences are access-conscious. If I see or hear something, then I am usually able to go on to think about it, tell others about it and decide how to react. But there are exceptions. Blindsighted people cannot draw on their blind-field perceptions in this way and can access them only indirectly, by making guesses. Similarly, subliminal perceptions are only partially available to other mental processes (we cannot report them or draw on them in our general reasoning).

Although phenomenal consciousness and access-consciousness typically go together, Block argues that they are distinct and could in principle come apart. As an example, he suggests that blindsighted people might be trained to make spontaneous guesses about what is present in their blind field, thereby improving their access to the visual information from that region. With enough training, they might find that the information popped into their heads automatically without the need for guessing. So when a red circle was shown in that area they would spontaneously think, 'There is a red circle there' and be able to report the fact and reason about it – even though they still could not *see* the circle in the normal sense. (Block calls this imaginary condition 'superblindsight'.) The information about the red circle would then be access-conscious without being phenomenally conscious. Whether Block is right about this is a matter of considerable dispute. As we shall see, some philosophers hold that phenomenal consciousness is at bottom just a kind of access-consciousness, and that we can explain the phenomenal character of a mental state in terms of its relations to other mental states and processes.

There is one final distinction I want to mention. One can have a phenomenally conscious experience without paying attention to it. For example, all day I have had a slight pain in my left leg. I have not thought about it much, but it has been there, in the background. Occasionally, however, we deliberately focus on our mental states and attend to their features. Now that I have mentioned the pain in my leg, I have started thinking about it and attending to its location, quality and intensity. This sort of inner attention is often referred to as *introspection*, and I shall say that mental states that are the object of it are *introspectively conscious*, whereas states that are conscious in the ordinary way are *non-introspectively conscious*. This distinction is particularly important when thinking about the mental life of non-human animals. It may be that the experiences of dogs, for example, are phenomenally conscious but not introspectively conscious – that dogs do not attend to their experiences in the way that we do.

The seventeenth-century notion of consciousness mentioned earlier ('perception of what passes in a man's mind') is close to that of introspective consciousness. In contemporary discussions, however, the focus is firmly on non-introspective phenomenal consciousness – on ordinary routine experience. What seems mysterious is how experience could have a phenomenal character at all. The fact that we can also deliberately attend to this character is a secondary matter. This is why I said that the modern notion of consciousness is subtly different from the older one.

At this point you may be feeling a bit confused. Surely, even non-introspective consciousness must involve inner awareness of some sort? How could a mental state *feel like* something if one isn't *aware* of its feel? Some philosophers would agree with this, arguing that even non-introspective consciousness involves inner awareness of some sort. But we should not prejudge the issue here. Many writers insist that the phenomenal character of an experience is not an object of awareness at all, but something that *accompanies* our awareness of other things. When we gaze at a beautiful sunset, they claim, we are aware only of the sunset, but our awareness of it has a certain phenomenal character. As Mark Rowlands puts it, what it is like to undergo an experience is not something *of* which we are aware, but something *with* which we are aware (Rowlands, 2002, 159).

### Activity 2

Here is an exercise to help you check your grasp of the distinctions mentioned above. Which meaning of 'consciousness' do the authors of the following quotations seem to

have in mind? (Unless otherwise indicated, the quotations are taken from the *Oxford English Dictionary* entry on consciousness.)

1. It is only to the consciousness of these evils that knowledge and reflection awaken him (F.A. Kemble).
2. We class sensations along with emotions, and volitions, and thoughts, under the common head of *states of consciousness* (Thomas Huxley).
3. [Consciousness is] being aware of oneself as a distinct entity, separate from other people or things in one's environment (C. Evans, *Dictionary of the Mind, Brain and Behaviour*, quoted in Smith, 1985, 129).
4. A state is conscious if it has experiential properties. The totality of the experiential properties of a state are 'what it is like' to have it (adapted from Block, 1995, 230).
5. Consciousness is a word used by Philosophers, to signify that immediate knowledge which we have of our present thoughts and purposes, and, in general, of all the present operations of our minds (Thomas Reid).
6. When the fever left him, and consciousness returned, he awoke to find himself rich and free (Dickens).
7. Content is conscious in virtue of… reaching the Executive system, the system in charge of rational control of action and speech (adapted from Block, 1995, 232).
8. [H]ow it is that anything so remarkable as a state of consciousness comes about as the result of irritating nervous tissue, is just as unaccountable as the appearance of Djin when Aladdin rubbed his lamp… (Thomas Huxley).

1. Transitive consciousness (note the 'of').
2. State consciousness (phenomenal, non-introspective?).
3. Self-consciousness.
4. Phenomenal consciousness (the original reads 'P-conscious').
5. Introspective consciousness.
6. Creature consciousness.
7. Access-consciousness (the original reads 'A-conscious').
8. Phenomenal consciousness.

These are the only distinctions I shall mention for now. But you should remember that the word 'consciousness' has other senses too, both in ordinary speech and in technical writing, and you should always check to see how it is being used. As I explained, our focus in this course will be on state consciousness of the ordinary non-introspective variety – phenomenal consciousness, in Block's terminology. I shall not keep spelling this out, however, and from now on, unless otherwise indicated, the words 'conscious' and 'consciousness' should always be understood in that way.

# 3 The elusiveness of consciousness

Consciousness is, in a sense, the most familiar thing in the world: our lives consist of a succession of conscious experiences. Yet consciousness can also seem elusive and mysterious, and this section contains some activities designed to highlight this. Here is a simple exercise to start us off.

### Activity 3

Think about the different varieties of conscious experience you have and make a list of them. Include perceptual experiences (sight, hearing, etc.), bodily sensations (pain, for example) and any others you can think of. Then turn to Reading 1, which is an extract from the opening chapter of David Chalmers's book *The Conscious Mind* (Chalmers, 1996), and compare your list with his. Do you find Chalmers's descriptions accurate? Are there are any points on which you disagree with him?

Click on the 'View document' link below to read David J. Chalmers on 'A catalog of conscious experiences'.

Chalmers notes that his catalogue is not intended to be exhaustive and you may have included items he omits. His list does, however, cover the main varieties of conscious experience, and it seems to me both evocative and, for the most part, accurate. There are only two points on which I would disagree with Chalmers. First, I think he misdescribes the feel of conscious thoughts (paragraphs 13–14). Such thoughts, he says, often have a distinctive qualitative feel, reflecting their subject matter: thoughts about lions, for example, have a 'whiff of leonine quality' about them. This does not reflect my own experience. I agree that occurrent thoughts often have a phenomenal character, but for me it is primarily the feel of *uttering* the thought to myself in inner speech – a feeling similar to that of saying it aloud, but muted. My thoughts are also sometimes associated with visual images and emotional feelings, though these tend to be vague and ill-defined. Secondly, I am not sure that Chalmers is right to claim that there is a distinct feel associated with the sense of self – a 'background hum', as he puts it, which accompanies our other more fleeting experiences (paragraph 17). For my part, I am not aware of such a feeling but only of specific experiences like those mentioned elsewhere in Chalmers's catalogue.

I suggest you refer back to your list and to Chalmers's catalogue as you work through this course. In philosophical discussions of consciousness it is common to focus on very simple experiences – usually visual ones – but it is important to keep in mind the range and variety of conscious experience, since theories of consciousness are intended to apply to all of them.

Chalmers concentrates on describing the feel of the various experiences he lists, but there is often more to an experience than its feel. Most experiences also carry information, or misinformation, about our environment (misinformation in the case of perceptual illusions, such as when a stick looks bent in water). So, visual experiences tell us about the colours, shapes and movements of things around us; auditory experiences tell us about the location and motion of objects; tastes and smells tell us about the substances present in our food and in the air; bodily sensations, such as pain and thirst, tell us about the condition of our bodies; and so on. States that carry information are known as

*representational states* and the information they carry is known as their *representational content* (the terms 'intentional state' and 'intentional content' are also frequently used, with the same meaning). For example, suppose I have a visual experience as of seeing a blue circle in front of me. The experience has the representational content *that there is a blue circle ahead*. If there is indeed a blue circle there, then this content is true – the experience represents the world accurately. If there is not a blue circle there (if I am hallucinating, say), then the content is false – the experience represents the world inaccurately.

### Activity 4

Do all conscious experiences have representational content? Can you think of any that do not? Does a headache carry information (about the state of blood vessels in the head, perhaps)? Does a buzz of excitement or a rush of euphoria? Does an orgasm?

This question is a controversial one. It is widely held that some bodily sensations and feelings lack representational content. Some philosophers, however, argue that representational content is the essence of consciousness and that all conscious experience possesses it. I am not going to discuss the issue further here; I want you just to bear the question in mind and see if your opinion changes as we go on.

I said that consciousness can seem elusive and mysterious and I want to use the rest of this section to illustrate some aspects of this.

### Activity 5

Look again at Reading 1, especially paragraphs 1–7. Chalmers highlights two ways in which consciousness seems mysterious. What are they?

One point Chalmers mentions several times is that the phenomenal character of many conscious experiences seems *ineffable* – we cannot find words to describe it adequately. Another point he mentions is that, in many cases, the connection between a stimulus and the resulting experience appears *arbitrary* – there seems to be no reason why the experience should have the phenomenal character it does, rather than a different one.

Let us consider these claims in more detail, beginning with ineffability. Chalmers's point is that it is often hard to describe an experience in a way that really conveys what it is like and that would be informative to someone who had never had it. This is not just because experiences can be very complex. Indeed, complex experiences may be easier to describe than simple ones, since we can break them down into more basic components. For example, a wine critic may describe the bouquet of a wine by saying that it contains scents of peach, anti-freeze and grass clippings. But these more basic sensory experiences seem indescribable. How could we describe the smell of grass clippings? It is distinctive and easily recognisable but seemingly impossible to characterise. (Of course, we can describe it *indirectly* as 'the smell you get from grass clippings', but how could we describe what it is like *in itself*?)

It is worth dwelling on this point a little. Take a simple visual experience – looking at a blue surface, say. How could I set about conveying the quality of this experience to someone

who had never had it? I might try comparing it to other experiences – saying, for example, that it is more like the experience of seeing green than it is like that of seeing yellow. But such descriptions would be of use only to someone who had already had those other experiences. What description could I give to someone who was congenitally blind? The only option, it seems, would to be make comparisons with non-visual experiences, but it is hard to find informative ones. (A famous example, cited by John Locke, is that of a blind man who had the idea that scarlet resembled the sound of a trumpet [Locke, 1961, vol. 2, 30]. Although not bad as such comparisons go, this still falls a long way short of capturing what it is like to see something red.) The same goes for experiences involving other sense organs.


Figure 2 The difficulties of describing phenomenal properties. Cartoon by James Thurber. Copyright © 1933, 1961 Rosemary A. Thurber. All rights reserved

In this connection is it interesting to note that we do not have distinctive words for phenomenal properties themselves. Take the experience a normally sighted person has when looking at a ripe tomato. What term should we use for the phenomenal character of this experience? We might loosely call it 'red' – in everyday speech we do sometimes talk of having red sensations or red afterimages. (An afterimage is the sensation one gets after staring at a bright light and then looking away.) But the *experience* – the mental state – is not really red, at least not in the same sense the tomato is. The experience is not *coloured* red. (It is true that if experiences are states of the brain, as many philosophers believe, then the neural tissues involved will have colours. But there is no reason to think they will be *red*. Your brain doesn't change colour depending on what you are looking at!) To get round this difficulty some writers coin special terms for phenomenal properties. The philosopher Joseph Levine, for example, uses 'reddishness' for the property possessed by experiences of red (Levine, 2001). Thus Levine would say that red things cause reddish experiences.

The claim that conscious experience is ineffable is closely related to another claim often made about it – namely, that it is *subjective*. The phenomenal character of an experience, it is claimed, can only be appreciated from the inside, from the first-person point of view. We might study the brain processes involved in a certain type of experience in the most minute detail, but we would not learn what the experience was really *like* unless we were to have it for ourselves. To emphasise the point, the phenomenal aspect of experience is often referred to as its *subjective character* – in contrast to the objective, publicly observable features of the brain states involved (Nagel, 1974).

Turn now to Chalmers's second point – the arbitrariness of phenomenal character. The idea here is that, in many cases, the connection between what an experience is *of* and the way it *feels* seems arbitrary. 'Why should *that* feel like *this*?' we are tempted to ask, reflecting on a stimulus and the experience it causes. Of course, as Chalmers notes, this is not true of all experiences. It is surely not arbitrary that the experience of seeing a cube and that of seeing a sphere should feel the way they do, rather than the other way round. But in many cases the connections do seem arbitrary. Colours, sounds and smells offer good examples. Why should the light reflected from a ripe tomato produce a reddish sensation (to use Levine's terminology) rather than a greenish one? Why do the sound

waves produced by a telephone cause us to hear a ringing sound, rather than, say, a squeaky one? Why do the chemicals in newly mown grass produce the particular smell they do, rather than another?

Of course, there is much we do not know about the brain processes involved in sense perception. But even if we knew everything about them it is still not clear that this sense of arbitrariness would be removed. We might still be at a loss to know why particular brain processes give rise to the particular experiences they do – why nerve firings in a certain region of the visual cortex (the area of the brain associated with vision) give rise to a reddish sensation, rather than a greenish one, or why the stimulation of certain cells in the olfactory bulb (the brain region associated with smell) causes a smell of grass clippings, rather than, say, one of linseed oil.

The apparent arbitrariness of phenomenal character suggests a strange possibility. If the links between stimuli and the experiences they cause really are arbitrary, then perhaps the same stimuli do not produce the same experiences in everyone. Perhaps when other people look at blue objects they have an experience of the sort I have when I look at yellow ones – so that for them looking at a cloudless summer sky is like looking at a vast sandy desert. How would I know?

### Activity 6

Could I tell by questioning other people if the phenomenal character of their blue and yellow experiences is inverted in this way? Pause and think for a few minutes.

It seems unlikely that I could. Asking them if the sky looks blue or yellow won't help. They will say it looks blue – since they have learned to call things that produce experiences with this phenomenal character ‘blue’. The question is whether they associate the word ‘blue’ with the same phenomenal character I do. Nor will it help to ask them to describe the experience itself. For, as we have seen, it is very hard to describe simple experiences in a way that conveys their phenomenal character. Perhaps the best option would be to ask them to make comparisons between colours. If blue things produce yellowish experiences in them, then they will say that blue things look more similar to orange things than to green things, whereas if they produce bluish experiences, they will say it is the other way round. Even this would not be conclusive, however. For it might be that their other colour experiences are switched round too – so that, for example, the experiences they have when looking at orange things are like those I have when looking at green things and vice versa. If the whole range of their colour experiences was systematically inverted, then – arguably – all colour comparisons would be preserved and the inversion would be undetectable. This is referred to as the possibility of *interpersonal spectrum inversion* (Figure 3).

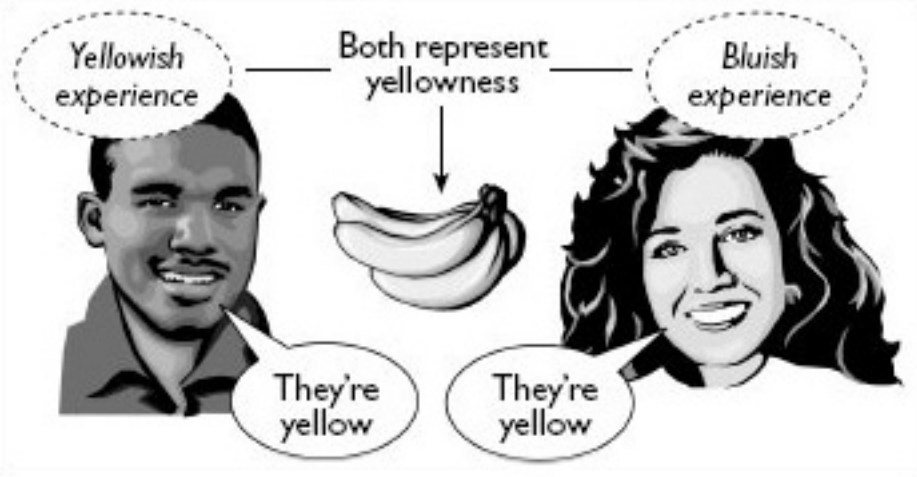



Figure 3 Spectrum inversion

The possibility of spectrum inversion is sometimes said to show that the phenomenal character of an experience is independent of its representational content. The thought is that differences in phenomenal character need not make any difference to the way we classify objects and use colour words. You and I could agree on which things are yellow, even if these things produce different experiences in each of us. And the resulting experiences would represent the same thing – namely, the presence of something of the sort we both call 'yellow' – despite their difference in phenomenal character. We can imagine inversions in other sense organs, too. For example, patchouli might produce in you the smell sensation that almond oil produces in me. Again, however, both experiences could *represent* the same thing – the presence of patchouli. Considerations like these lead some people to say that the phenomenal character of an experience is an *intrinsic*, or *non-relational*, property of it – that is, one which it possesses in its own right, independently of its relations to other things. The representational properties of an experience, on the other hand, are not intrinsic, but determined by its relations to the object or property represented.

The conception of phenomenal properties just outlined – as ineffable, subjective and intrinsic – has been very influential in philosophical thinking about consciousness. Not everyone agrees that the conception is correct, however. Although consciousness can seem elusive when we reflect on it from the first-person point of view, many philosophers believe that our intuitions in this area should be treated with caution. Some writers deny that we are aware of intrinsic, non-representational features of our experiences, and many believe that conscious experiences are states of the brain that are, in principle, publicly observable and describable in physical terms.

# 4 The problem of consciousness

## 4.1 Introduction

Let us turn now to the problem of consciousness. What exactly is the issue here that so divides philosophers and is the focus of such vigorous debate? In broad terms, it is the question of the place of consciousness in the world – the question of how it arises and how it is related to processes in the brain. It is hard to deny that consciousness is closely dependent on the brain. Changes in the brain can affect consciousness (think of the effects of anaesthetics and psychedelic drugs), and damage to the brain can remove it permanently (think of blindsight, for example). But *how* does the brain generate consciousness? How *could* conscious experiences arise from the activity of brain cells – individually not much different from any other cells? As Colin McGinn puts it, it seems like magic:

> How is it possible for conscious states to depend upon brain states? How can technicolour phenomenology arise from soggy grey matter? What makes the bodily organ we call the brain so radically different from other bodily organs, say the kidneys – the body parts without a trace of consciousness? How could the aggregation of millions of individually insentient neurons generate subjective awareness? We know that brains are the *de facto* causal basis of consciousness, but we have, it seems, no understanding whatever of how this can be so. It strikes us as miraculous, eerie, even faintly comic. Somehow, we feel, the water of the physical brain is turned into the wine of consciousness, but we draw a total blank on the nature of this conversion. Neural transmissions just seem like the wrong kind of materials with which to bring consciousness into the world, but it appears that in some way they perform this mysterious feat. The mind-body problem is the problem of understanding how the miracle is wrought, thus removing the sense of deep mystery. We want to take the magic out of the link between consciousness and the brain.
>
> (McGinn, 1989, 349)

This is the problem of consciousness – the problem of taking the magic out of its link to the brain and, thereby, to the rest of the natural world. For many philosophers, solving this problem would involve giving a *reductive explanation* of consciousness, and I shall begin by saying something about explanations of this kind and the reasons for seeking them.

## 4.2 Naturalism and reductive explanation

There is a widespread commitment among contemporary philosophers and scientists to a *naturalistic* view of the world. In broad terms, naturalism is the view that everything is scientifically explicable – to put it crudely, that there are no miracles. (Note that I am using 'naturalism' here for a *metaphysical* position – a view about the nature of the world. It is also used for a *methodological* position – a view about how the world, or some aspect of it, should be studied. Indeed, the word has a variety of meanings and should be used with care.) Thus, naturalists deny the existence of supernatural entities and powers and

assume that everything that happens is causally explicable by reference to scientific principles and laws. Some naturalists also make a further assumption. They assume that natural phenomena form a hierarchy and that higher-level ones can be explained by reference to more basic ones, right down to the level of chemistry and physics. *Reproduction*, for example, counts as a high-level phenomenon, which can be explained in terms of more basic genetic and cellular processes, which can themselves be explained in chemical and physical terms. This sort of explanation, where a phenomenon at one level is explained in terms of those at a lower level, is called *reductive explanation*. The notion of reductive explanation is a key one in the modern debate about consciousness and it is important to understand how reductive explanations work. An example may be useful.

Suppose that a person who has had no previous contact with modern civilisation is shown a working television set, displaying images of events taking place many miles away. They are astonished by the device and declare that it must be magic. How would we convince them otherwise? The answer, of course, is by explaining how a television set works. We might begin by describing a television camera – explaining that it uses a lens to focus a moving image onto a light-sensitive plate, which then generates a stream of electronic pulses, corresponding to the pattern on the plate. We would then explain how this electronic signal is amplified and broadcast – explaining what radio waves are and how they can be used to carry an electronic signal. Finally, we would turn to the television set itself and explain that it detects radio waves via an antenna, decodes the signal and uses it to modulate the beam of a cathode ray tube, causing the tube to emit patterns of light which correspond to the images in the camera and which are perceived by the human eye as a moving picture. Of course, in order to make all this comprehensible we would have to provide a lot of further information about the underlying physical processes – about light, optics, electricity, radio waves and so on – but with time and access to reference books we could surely satisfy our hearer that there was nothing magical about the television.

In doing all this we would have *reductively explained* the television's power to display moving images of distant events. That is, we would have shown that this property follows from more basic, lower-level properties of the television – its possession of various mechanical and electronic components. These properties explain the television's power to display moving images of distant events because it is obvious that they are sufficient for it. Nothing more is needed in order for the television to have that power than for it to possess those properties. In a widely-used phrase, the lower-level properties *realise* the higher-level one: the television possesses the latter *in virtue* of the fact that it possesses the former.

It is important to distinguish reductive explanation from *reduction*. To say that a property can be *reduced* to a lower-level one is to say that it can be identified with it across the board – that they are in fact the same property, under different names. (Or at least that is one common meaning of ‘reduction’.) For example, the property of being water reduces to that of having the molecular structure $H_2O$. Reductions like this are quite rare, however, since most properties can be realised in more than one way (‘multiply realised’). Different kinds of television, for example, work in different ways and are made of different materials (some have plasma screens instead of cathode ray tubes, some receive the signal by cable instead of aerial, older models use vacuum tubes or transistors instead of integrated circuits and so on). So we cannot identify the property of being a television with that of having a particular set of components. *Any* components will do, provided they do the job. However, the fact that a phenomenon cannot be reduced to a lower-level one does not mean that it cannot be reductively explained in lower-level terms. Each *instance* of the

phenomenon may be realised in lower-level properties and explicable in terms of them, even if these properties are not the same in every case.

Now as I said, many philosophers and scientists assume that all phenomena above the level of basic physics can, in principle, be reductively explained. They view the natural world as a unified structure, whose higher levels of organisation emerge in a thoroughly comprehensible way from lower-level ones and ultimately from basic physical states and processes. I shall refer to this view as *strong naturalism*.

Strong naturalism has considerable plausibility. It is a remarkable fact that just about every phenomenon scientists have studied has turned out to yield to reductive explanation. Take life, for example. Until the middle of the nineteenth century it was common for biologists to maintain that life was not the product of more basic inorganic processes, but dependent on a special vital spirit or force – a view known as *vitalism*. It is easy to see why they thought this. Inanimate structures tend to decay steadily, whereas living things are able to sustain, repair and reproduce themselves. Given the undeveloped state of biological knowledge, it was not implausible to think that this amazing regenerative ability could not be the product of mere physical processes. In fact, of course, this was quite wrong. As biologists studied organic processes in more detail, they discovered that they were nothing more than complex chemical reactions, which could be replicated in the laboratory. With time, more and more biological phenomena yielded to reductive explanation, and today vitalism is wholly discredited. What is special about living things, it turned out, is not that they possess a non-physical ingredient but that they involve a unique and very complex *organisation* of physical elements.

What proved true in biology has proved true in the other sciences, too. Almost everywhere scientists have been able to explain higher-level phenomena in terms of lower-level ones.

### Activity 7

Can you think of any properties that seem unlikely to yield to reductive explanation? (Set aside mental ones for the moment.)

The most obvious candidates, I think, are moral and aesthetic properties. Can the rightness and wrongness of our actions be reductively explained by reference to their physical characteristics – when and where and in what manner they were performed? It seems unlikely: the very same action, physically characterised, might be disloyal, say, in one context but not in another. Similarly, can we explain why objects have the aesthetic properties they do – why they are graceful or elegant or ugly, for example – by reference to their physical properties – their colours and shapes and so on? Again, many would say no: we cannot read off an object's aesthetic properties from its physical ones.

It may be, then, that moral and aesthetic properties cannot be reductively explained. But even if this is so, there is no fatal objection here to strong naturalism. For defenders of the doctrine may simply deny that moral and aesthetic properties are real properties of actions and objects and claim instead that they are just projections of our own responses to them. Indeed, for some people, the very resistance of these properties to reductive explanation is a reason for denying their reality. If a phenomenon cannot be reductively explained – if we cannot see how it could arise from lower-level processes – then, these people would say, that is a good reason for thinking that it is not real or has at least been seriously mischaracterised.

(It is worth stressing at this point that a strong naturalist need not claim that reductive explanation is the only legitimate kind of explanation. A reductive explanation shows how a phenomenon is constituted, but there are other types of explanation with different functions. For example, much scientific explanation involves explaining processes at a high level of description without going into the details of how they are constituted. This is the case with explanations in the so-called *special sciences* – the sciences devoted to specific phenomena above the level of basic physics – biology, chemistry, psychology and so on. Strong naturalists need not deny the legitimacy or usefulness of these other types of explanation, though they will claim that there are reductive explanations of *why* they hold.)

Here is an activity to reinforce the points just made.

### Activity 8

Which of the following claims would strong naturalists endorse?

1. Everything that exists is natural.
2. Everything that happens can be scientifically explained.
3. Science can only deal with natural processes; supernatural ones are beyond it.
4. All phenomena above the level of basic physics can be explained in lower-level terms.
5. All phenomena can be reduced to physical ones.
6. Reductive explanation is the only legitimate kind of explanation.

---

Strong naturalists would endorse (2) and (4). (1) is ambiguous. If 'natural' means 'not supernatural', then strong naturalists would endorse it. If it means 'not man-made' then of course they would not. Naturalism has nothing to do with the contrast between the natural and the man-made. (3) is a misunderstanding of the naturalist position, as I have characterised it. Naturalists do not claim that supernatural processes are beyond science; they claim that there *are no* supernatural processes. As for claims (5) and (6), I have already explained that strong naturalists need not endorse them. The strong naturalist need not maintain that higher-level phenomena can be reduced to physical ones, merely that they can be reductively explained in physical terms. Nor need they deny the legitimacy of other kinds of explanation.

## 4.3 The easy problems and the hard problem

What implications do naturalism and strong naturalism have for the study of the mind? There are two. First, naturalists will deny the existence of souls, spirits and other psychic phenomena and maintain that the mind is part of the natural world, subject to natural laws. This view is shared by most modern philosophers of mind. Secondly, *strong* naturalists will hold that mental phenomena can be reductively explained in terms of processes in the brain, which can themselves be explained in terms of lower-level processes at the chemical and physical level. Although not as widely accepted as the first, this view is also common among contemporary philosophers, and, indeed, there is a strong case for it. All other high-level phenomena seem to be reductively explicable; why should the mind be any different?

But how *could* brain processes give rise to minds and mental states? How could collections of neurons and synapses generate beliefs and desires, hopes and fears, pains and pleasures? Much of contemporary philosophy of mind has been devoted to trying to answer this question – to constructing a naturalistic theory of the mind – and though we are still a long way from fully understanding how the mind works, there are plenty of theories as to how mental states and processes might be realised in the brain.

An important development was the idea that many mental states and processes can be defined *functionally*, in terms of the causal role they play in the operation of the mind – the view known as *functionalism*. So, for example, a belief is a state that is generated by perception or inference, serves as a premise in reasoning and prompts actions that would be rational if it were true; a desire is a state that is caused by bodily needs, serves as a goal in reasoning and tends to produce behaviour that will satisfy it; perception is a process in which information about the environment is acquired through the receipt of sensory stimuli; and so on. If we think of mental states and processes in this way, then it is not too difficult to see how a brain could support them. It would just have to possess states and mechanisms that play the appropriate causal roles.

Another source of inspiration was the development of computers, which provided models of how reasoning could be performed mechanically, through the manipulation of symbols. This suggested that the brain itself might be a biological computer operating on symbols in an internal language, and a new field of research opened up devoted to modelling mental processes in computational terms. Again, on this view it is not too difficult to see how brain tissue could support a mind; it would simply need to be organised in such a way as to implement the relevant computational processes. This approach may not be the right one (there are rivals to it) and many problems remain – in particular, that of explaining how symbols in the mental language get their meaning. But it does suggest that there is no obstacle *in principle* to providing reductive explanations of many mental phenomena.

When it comes to consciousness, however, the functional/computational approach runs into problems. Although some of the things we call ‘consciousness’ may be explicable in functional/computational terms (access-consciousness, for example), it is very hard to see how *phenomenal* consciousness could be. This problem has been recognised since the development of functional approaches to the mind in the late 1960s, but it was powerfully restated in the 1990s by the Australian philosopher David Chalmers (b. 1966), who has famously dubbed it the ‘hard problem’ of consciousness. I shall let Chalmers outline it himself, in an extract from one of his first papers on the topic.

### Activity 9

Follow the link to David Chalmers' article ‘Facing Up to the Problem of Consciousness and read sections 2 and 3 (‘The Easy Problems and the Hard Problem’ and ‘Functional Explanation’). Then answer the following questions.

1. In paragraph 2 Chalmers lists various phenomena associated with the word ‘consciousness’. Which of the terms introduced earlier (‘creature consciousness’, ‘access-consciousness’, ‘transitive consciousness’, etc.) corresponds best to each of the items in the list? (Note that in some cases the correspondence is not exact.)
2. What does Chalmers mean by ‘experience’? (Paragraphs 5–6)
3. Why, according to Chalmers, are the easy problems easy? (Paragraphs 9–13)
4. Why is the hard problem hard? (Paragraphs 14–16)

1. The phenomena line up roughly as follows. The first (the ability to discriminate, categorise and react to stimuli) is a state of general awareness, so it falls under the heading of creature consciousness. The second, third and fourth items (the integration of information, reportability and internal access) involve the passing of information between internal systems, so they can be grouped under access-consciousness. The fourth phenomenon (attention) is a perceptual process, so it comes under the heading of transitive consciousness (awareness *of* something). A deliberate action is one performed with reflective awareness, so the fifth item (the deliberate control of behaviour) involves introspective consciousness (and perhaps also self-consciousness). The last item (wakefulness) corresponds to creature consciousness again.
2. He means *phenomenal consciousness* – the subjective aspect of our experiences, *what it is like* to have them.
3. The easy problems are easy because the phenomena to be explained are functionally definable and we can explain how a system exhibits them by describing the mechanisms that perform the relevant functions. These mechanisms might be described either in neurological terms or in more abstract computational ones. (In the latter case, to give a full explanation we would also have to specify the neural mechanisms which implement the computational processes, but that would be just another 'easy' problem.) Thus, for example, if we can identify the brain mechanisms that give us the ability to make verbal reports of our beliefs and other mental states, then we shall have explained the phenomenon of reportability.
4. The hard problem is hard because it goes *beyond* the performance of functions. Even when we have explained all the various functional processes that occur when we perceive things, we would still not have explained why these processes are accompanied by conscious experience – that is, why our perceptions have a phenomenal character. This looks like a much more difficult problem.

In this extract Chalmers is appealing to intuition rather than offering arguments, and you should not take his comments to be the final verdict on functionalism. But the intuition to which he appeals is certainly strong. Put simply, functionalism characterises mental states by what they *do*, rather than by how *they feel*. And it seems that a brain state could play the functional role of an experience without having any phenomenal character to it. Take pain, for example. Pains have a distinctive functional role: they are caused by bodily damage and cause characteristic behavioural reactions. Yet, it seems, a brain state could play this role without actually *hurting*. Think about Cog again. Suppose that damage to Cog's body activates an internal subsystem which registers the location and extent of the damage and initiates appropriate action, such as protecting the damaged area, withdrawing from the source of the damage and emitting the word 'Ouch!' from a speech synthesiser. Then when this subsystem is activated, it would be appropriate to say that Cog is *in pain*, in the functional sense, even though Cog doesn't actually feel anything. Similarly for other perceptions and experiences. It seems that a brain state could play the functional role appropriate to a visual perception – say, of a bright blue light – without having the phenomenal character normally possessed by such a state, or indeed with a quite different phenomenal character. So, it seems, functionalism leaves the mystery of consciousness untouched: how do some brain states come to have phenomenal character?

We can look at the same problem from another perspective. Suppose the MIT team wanted to give Cog conscious experiences. What would they have to do? Would it involve new programming? Or new circuitry? Or what? There are many things they could do to improve Cog's visual system – increasing the sensitivity of its camera-eyes, boosting the power of its visual processors and upgrading their software – but it is not clear what they could do to give its visual processes phenomenal character. Where would they start? If we have no idea how nature produces conscious experiences, then how can we set about trying to produce them artificially? It is worth noting that practically all the research programmes currently being pursued by the MIT team and other roboticists are devoted to equipping robots with specific functional capacities – capacities to discriminate, categorise, learn, perform everyday tasks and so on. None is devoted directly to making it *conscious*. Indeed, the MIT team say they try to avoid using the 'c-word' in their labs!

Let me repeat that you should not take this as the final verdict on functionalism. Many functionalists think that their approach *can* explain consciousness. When properly understood, these writers claim, the functional processes involved in experience do explain its phenomenal character. And, of course, even if functionalist explanations fail, a reductive explanation in other terms might still be possible. But it is undeniable that there is a serious problem here, and some people believe that consciousness is resistant *in principle* to reductive explanation. Here, they claim, strong naturalism reaches its limits.

## 4.4 Physicalism and the hard problem

I introduced the hard problem as an explanatory problem – the problem of explaining how consciousness arises. But it can also be presented as a metaphysical problem – the problem of saying what kind of phenomenon consciousness is, and, more specifically, whether it is a *physical* one. In this section I shall say something about this aspect of the hard problem and its relation to the explanatory one.

The terms 'physical' and 'physicalism' (the view that everything is physical) are used in a number of different senses and it is easy to become confused by them. Some writers who count as physicalists in one sense of the term count as anti-physicalists in another. (The same goes for 'materialism', which is sometimes used interchangeably with 'physicalism'.) I shall distinguish some important senses and give them labels, but you should note that other writers draw the distinctions in different ways and use the terminology differently.

In one form, physicalism is the view that everything in the universe is composed wholly of the basic entities and forces postulated by modern physics (electrons, protons, gravity, electromagnetism and so on). It is the view that, as John Haugeland puts it, 'if you took away all the atoms, nothing would be left behind' (Haugeland, 1982, 96). I shall refer to this view as *substance physicalism*. (*Substance dualism*, on the other hand, is the view that the universe also contains other entities and forces in addition to the basic physical ones – immaterial souls or psychic energy, for example.) Now I am going to assume that substance physicalism is true. This reflects the prevailing view among contemporary philosophers of mind, including some who would describe themselves as being, in another sense, non-physicalists. The modern debate over physicalism focuses on other claims, not about *substances*, but about *properties*. (If you are sympathetic to substance dualism, you should not conclude that the rest of this course will be irrelevant to you; I shall explain why shortly.)

Suppose substance physicalism is true. Still, questions remain about the *properties* of things. Let me begin by introducing the notion of a *basic physical property*. By this I mean a property invoked by physicists, such as mass or electrical charge, or a property that can

be defined in terms of the properties invoked by physicists, such as that of being composed of atoms of a certain kind. Now if physics describes the basic components of the world, then there is a sense in which the basic physical properties of things are the fundamental ones. But, of course, things possess many other properties in addition to their basic physical ones. Take me. I have various basic physical properties. For example, I have a certain mass, I am composed of millions of molecules arranged in elaborate structures, I am the site of numerous complex electrochemical processes. But I also have many other properties: I am alive, I am human, I have a digestive system, I belong to blood group O, I like Bob Dylan, I currently have a slight headache and so on. Let us call these *high-level properties*. But how are these high-level properties related to my basic physical ones? The question could also be phrased in terms of facts. For each of my properties there is a corresponding fact – the fact that I possess the property. So another way of putting the question would be to ask how the high-level facts about me are related to the basic physical facts about me. Most modern debates about physicalism are about the answers to questions like these.

*Property physicalism* is the view that high-level properties are not fundamentally distinct from basic physical ones. They are not new features of the world, in addition to the basic physical ones, but just those same features under different guises. So, for example, my having a digestive system is not an extra property of mine, over and above the basic physical ones. Rather, it *consists in* my having certain basic physical properties – having certain basic physical components arranged in a certain way and performing certain functions. Similarly for all other high-level properties. Or, putting it in terms of facts, high-level facts are not extra facts, over and above the basic physical ones; rather, they are just redescriptions of the basic physical ones. There is, it is true, a sense in which high-level facts plainly are different from basic physical ones: the claim that I have a digestive system does not *mean the same* as the claim I have certain basic physical properties. But – the physicalist will say – there is just one underlying state of affairs which makes both claims true. This is sometimes expressed by saying that once God fixed the basic physical facts, he fixed all the facts; there was no more work for him to do (Kripke, 1980). (The reference to God need not be taken literally – it is just a vivid way of making the point about the relation between the different properties.)

Property physicalism contrasts with *property dualism*. This is the view that some high-level properties *are* fundamentally distinct from basic physical ones – that they are additional features of the world, over and above the basic physical ones. Or, putting it in terms of facts, some high-level facts are extra facts, distinct from the basic physical ones. So when God fixed the basic physical facts, he still had more work to do: he still had some high-level facts to fix. If he had pleased, God could have created a world which was exactly like ours in all its basic physical details but which didn't have the same high-level properties. It is worth stressing that property dualists do not claim that *all* high-level properties are distinct from basic physical ones, but only that *some* are. Thus, while property physicalism is a general claim about all high-level properties, property dualism comes in different versions, each concerned with a different high-level property or group of high-level properties.

### Activity 10

Consider the property of being alive.

1. What would a property dualist about life say about the relation between my basic physical properties and the fact that I am alive?

2. What would a property physicalist say?

1. The dualist about life would say that being alive is an extra property of mine over and above my basic physical ones. God could have made a creature that was exactly like me in all its basic physical properties but was not alive.
2. The property physicalist would say that my being alive is not an extra property of mine, over and above my basic physical ones. Given that I have all the basic physical properties I do, I could not fail to be alive: there is nothing more to it.

Of course, the dualist about life may accept that there is a regular *correlation* between the presence of certain basic physical properties and the presence of life. After all, we don't find life in just any old physical structure but only in certain highly organised ones. The dualist may even say that it is a law of nature that when certain basic physical properties are present, then life is, too. But, they will say, it is a contingent fact that this law holds and it could have been different. (Compare the way that light has a certain speed in our universe but could have had a different one.) The physicalist, on the other hand, will deny that it is a contingent fact that certain basic physical properties are correlated with life. Rather, they will say, there is *nothing more* to being alive than having the right set of basic physical properties, and the latter could not occur without life.

There is one more distinction to make before we move on. Property physicalism is the view that high-level properties are not fundamentally distinct from basic physical ones, but there are stronger and weaker versions of this view. According to the stronger version, high-level properties *reduce* to basic physical ones. That is, each high-level property can be identified with a single basic physical property (or single set of such properties) in all its instances. This is sometimes expressed by saying that the two properties are *type-identical*. This strong form of property physicalism is not plausible. *Some* high-level properties reduce to basic physical ones; for example, having blood group O reduces to having blood of a certain molecular composition. But, as I mentioned earlier, most high-level properties can be constituted in more than one way. For example, the digestive system involves quite different structures in different animals. The same goes for mental properties, too. An alien might suffer from headaches and like Bob Dylan despite having a completely different brain chemistry from me. It does not follow, however, that these properties are fundamentally distinct from basic physical ones; it remains open that they are *realised in* basic physical properties. That is, every instance of these properties might be identical with an instance of some basic physical property – the nature of the latter varying from case to case. So a weaker and more plausible version of property physicalism holds that high-level properties *either* reduce to *or* are realised in basic physical ones.

This weak form of property physicalism is the most popular contemporary version of physicalism and the chances are that when you come across the word ‘physicalism’ (or ‘materialism’) in a contemporary book or article it is this weak form of property physicalism, or something very close to it, that is meant (though there will be exceptions and you should always check to make sure). For convenience, I too shall use the term ‘physicalism’, without qualification, to refer to this position. I shall also assume that physicalists in this sense endorse substance physicalism, though I shall not be discussing this aspect of their position. Two more terminological points. I shall use the term ‘physical’ in a broad sense to refer both to basic physical properties and also to high-level properties that are realised in basic physical ones. Likewise, I shall use the term ‘physical facts’ both for facts about basic physical properties and also for facts about high-level properties that

are realised in basic physical ones. Thus, the physicalist, in the sense just defined, holds that all properties are physical properties and that all facts are physical facts.

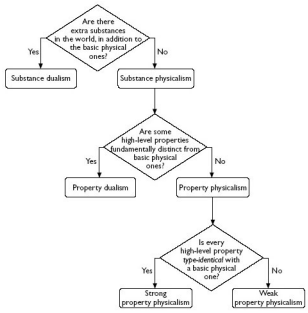



Figure 4 Varieties of dualism and physicalism

## Activity 11

Which of the following statements would be endorsed by a physicalist, in the sense just defined?

1. All high-level properties reduce to basic physical ones.
2. Each instance of a high-level property is identical with an instance of some basic physical property.
3. All high-level properties either reduce to, or are realised in, basic physical ones.
4. Only basic physical properties are real.
5. All objects are composed wholly of basic physical entities.

---

Physicalists would endorse (2), (3) and (5) (the first two of these say much the same thing). (1) is a stronger claim than the one physicalists make. Physicalists accept that many high-level properties can be multiply realised and so cannot be reduced to basic physical ones. Nor is there any reason for a physicalist to endorse (4). Physicalists do not hold that high-level properties are unreal – only that they are not the fundamental ones. Indeed, for a physicalist, showing that a high-level phenomenon is physically constituted amounts to a demonstration of its reality – proof that it is not illusory but firmly grounded in physical reality. Finally, (5) is a statement of substance physicalism, which we are assuming physicalists also endorse.

It is worth stressing that most writers accept that *many* phenomena are physical in the broad sense defined above. It is widely accepted, for example, that chemical, biological, neurological and functional properties are physical ones, and in the rest of this book I shall assume that this is the case. The question is whether consciousness is a further, non-physical property, over and above these.

I now want to link up this discussion of physicalism with the earlier discussion of reductive explanation. The crucial thing to note is that strong naturalists are committed to physicalism, in the sense just defined. For they hold that all high-level properties can ultimately be explained in basic physical terms. And such explanations will be possible only if high-level properties are not fundamentally distinct from basic physical ones. Recall how we explained the television's power to display moving images. The explanation worked because it was clear that the properties mentioned in the explanation were *sufficient* for the existence of the power. With those components, working in that way, the

set simply *could not fail* to display moving pictures. There was nothing more to it. In short, the properties cited explained the power because they *realised* it.

For the same reason, if a high-level property *is* distinct from basic physical ones, then it will not be possible to explain it in basic physical terms. Suppose property dualism about life were true – that life was an extra property, over and above basic physical ones. Then it would not be possible to give a reductive explanation of life in basic physical terms. For even if we were to identify all the various chemical and physical processes associated with life, there would still remain a mystery: *Why* does life emerge when those properties are present? If life were something extra, over and above those properties, then identifying them would not fully explain its existence. (I should add that very few people are property dualists about life; as I mentioned earlier, it is widely accepted that organic processes can be reductively explained in terms of chemical and physical ones.)

To sum up: if you believe that all high-level phenomena can be reductively explained in basic physical terms, then you are committed to the view that high-level properties are not fundamentally distinct from basic physical ones – that is, you are committed to physicalism.

### Activity 12

What implications does the conclusion just reached have for the hard problem of consciousness?

If solving the hard problem involves reductively explaining consciousness in lower-level terms and ultimately in basic physical ones, then a solution will be possible only if consciousness is a physical phenomenon – that is, only if phenomenal properties are physical properties.

The converse also applies. If consciousness is a physical phenomenon, then it should be possible to reductively explain it by identifying the physical properties in which it is realised. When these properties are described in detail, it should be obvious that they are sufficient for the existence of consciousness – just as it was obvious that the various electrical and mechanical components of the television were sufficient for it to have the power to display images of distant events. The question of whether consciousness is a physical phenomenon and that of whether it can be reductively explained are thus two sides of the same coin: considerations for and against the one claim count equally for and against the other.

I shall close this section by addressing some worries that may have been raised by the preceding discussion. First, I mentioned that I was going to assume that substance physicalism was true. The question we shall be considering is whether the *property* of consciousness is a physical one, not whether the *substance* which possesses this property is. I shall assume that this substance is simply the brain. But what if you do not share this assumption? What if you believe that we have a non-physical soul as well as a brain? Won't you find the debate about properties irrelevant? I don't think you should. For it is quite possible to hold *both* that we have a non-physical soul *and* that consciousness is a property of the brain. *Some* mental properties may be physical, even if others are not. (And, as I noted earlier, it is hard to deny that consciousness is at least very closely related to the brain; we know that chemical changes in the brain can affect consciousness and that damage to the brain can remove it.) If, on the other hand, one of your reasons for believing in substance dualism is that you think that consciousness is *not* a physical

phenomenon, then you will need to consider the arguments for the view that it is, and think about how you would respond to them. You might also want to consider whether you really need to endorse *substance* dualism, as opposed to the weaker property-based version.

Secondly, you may feel uncomfortable with the emphasis I have placed on a reductive, physicalist explanation of consciousness. Why should we bother to look for such an explanation? Isn't the whole approach dehumanising?

Wouldn't it be better to turn to novelists and poets for an account of our inner lives, rather than to scientists? There is something in this. If we want descriptions of what consciousness is *like* – the subtle shades of perception, sensation, emotion and thought – then a reductive explanation is likely to leave us cold. This is true across the board. An account of how a phenomenon is constituted will not tell us much about how it affects us. If you want to know about the emotional impact of the *Mona Lisa*, then a chemical analysis of the paint will not satisfy you. But it would be a mistake to think that this is an objection to reductive explanation. High-level descriptions of a phenomenon are not made redundant by a reductive explanation of it. The two serve quite different purposes: one tells us how the phenomenon appears to us, the other how it is constituted. Indeed, the two approaches may be complementary; close observation of a phenomenon may provide hints as to how it is constituted, and learning how it is constituted may illuminate our observations of it. A chemical analysis of a painting may help us to see it in a new way – to see new relations between its colours and to understand why they have the effects they do. There is no need, then, to fear that a reductive explanation of consciousness will replace the descriptions of novelists and poets; indeed it may help to enrich them.

Thirdly, is physicalism really a coherent position? Some people object that it is not, since it is not clear what counts as a basic physical property. After all, physicists keep revising their catalogue of the basic physical entities and forces, and they will probably continue to do so for some time (Crane and Mellor, 1990). This is a fair point, but not, I think, a fatal one. Physicalists can respond that their position involves an open-ended commitment – a commitment to regard as fundamental the properties posited by our current physics, whatever these happen to be. (Compare how being a law-abiding citizen involves an open-ended commitment to abide by the laws in force at the time, whatever these may be.) This means that physicalism is not a tightly defined doctrine, but in practice this does not matter too much, especially so far as philosophy of mind is concerned. For what is chiefly at issue in debates about the mind is whether mental properties are identical with, or realised in, *non-mental* ones. The exact nature of these non-mental properties is a secondary issue. So physicalists can live with some vagueness in their position (Papineau, 1993, 2002). Of course, if physicists were to decide to include *mental* phenomena in their catalogue of the fundamental entities and forces, then it would be a different story. But in that case the whole debate would take on a completely new aspect and everyone would have to rethink their position.

## 4.5 The function of consciousness

There is another problem I want to mention briefly. What is the *function* of consciousness? What difference does it make to have phenomenally conscious experiences?

This may seem an odd question. Surely, the answer is obvious: the function of consciousness is to provide us with information about our environment – about colours, shapes, sounds and so on. But this is too swift. We do not need to have conscious experiences in order to acquire perceptual information about our environment. Cog's sensors provide it with information about colours and shapes and sounds, too – it is just

that this information does not have a phenomenal character to it. What is added by supplementing this information with phenomenal character? We can put the same point in terms of the distinction between access-consciousness and phenomenal consciousness. It is obvious why it is useful for a creature's experiences to be access-consciousness – to be available to the processes controlling reasoning and behaviour. But why is it useful for them to be phenomenally conscious, too?

It might be suggested that the phenomenal character of an experience affects our *reaction* to it. Pain, for example, not only tells us that our body has been damaged, but also induces us to react to the damage in a certain way. If I touch something hot, then the pain moves me to withdraw my hand. Smells, tastes and colours also provoke characteristic reactions. Again, however, this is too swift. For a sensory state could cause a reaction without having any phenomenal character. As I mentioned earlier, it would be possible to program Cog to take avoiding action when it detects damage to itself – so that if someone pokes it in its eye, for example, it registers the fact, withdraws its head quickly and says 'Ouch!'. Yet it might still not actually *feel* anything – not have any conscious sensations of pain. So what is the point of consciousness? Provided Cog reacts to damage in the right way, why need it feel pain as well?

There is a general problem here. Whatever effects a conscious mental state has, it seems, a non-conscious one could also have. ('Conscious' here means 'phenomenally conscious' of course.) So why did evolution equip us with conscious experiences, rather than non-conscious ones? What survival advantage does phenomenal consciousness confer? Does it do anything at all? Or is it just a by-product of other processes, like the exhaust from a car's engine, which does not play any useful role?

This problem is closely connected with that of providing a reductive explanation of consciousness. Reductive explanations of mental phenomena typically exploit the fact that mental states can be characterised in functional terms – in terms of the role they play in mental processing and behavioural control. If a mental state can be characterised in this way, then we can identify it with whatever brain state plays the role in question. But if consciousness does not have a function, then this approach is a non-starter.

You may be feeling that something must have gone wrong here. Surely it is absurd to suggest that consciousness has no function – that the painfulness of a pain makes no difference to its effects. The suggestion is certainly counterintuitive; but that does not necessarily mean that we should rule it out. Even our strongest intuitions can mislead us (it seemed obvious to our ancestors that the earth was flat and that the sun moved through the sky), and we may have to escape the confines of our familiar outlook if we are to understand consciousness.

# 5 Conclusion

This course has laid the groundwork for a study of consciousness. We have identified the phenomenon in which we are interested, looked at some of its mysterious features and considered the problems it poses. You may have been wondering exactly what philosophers have to contribute here. Isn't explaining consciousness a matter for scientists – requiring the formation and testing of empirical hypotheses, not conceptual analysis and a priori reasoning? There are at least three aspects to the answer.

First, in posing a challenge to strong naturalism and physicalism, consciousness raises questions that go *beyond* science. Can everything be reductively explained? Are we capable of understanding all natural phenomena? Is physicalism true? Can there be a science of consciousness? These are metaphysical and epistemological questions – questions about the fundamental nature of reality and about the limits of our knowledge – and as such they fall squarely within the province of philosophy.

Secondly, even if a reductive explanation of consciousness is possible, there are some very general theoretical questions that need to be addressed at an early stage. What overall shape should the explanation take? What kind of mechanisms should it postulate? Could phenomenal consciousness be a form of access-consciousness? Does it involve inner awareness of some kind? These questions, too, are ones that philosophers are currently trying to answer. They see themselves as working in conjunction with scientists, helping to establish an outline theory of consciousness that will provide a framework for future empirical work. This does not require detailed scientific knowledge – though philosophers of mind do draw on empirical work. Rather, it involves thinking at a very general level about the facts of consciousness and how best to explain them.

Thirdly, it may be necessary to rethink the very way we pose the problem of consciousness. Some philosophers believe that our commonsense view of consciousness involves serious misconceptions, which blight philosophical and scientific work on the subject, and that philosophers have a role to play in exposing these and developing a better conception of the phenomenon.

This OpenLearn course provides a sample of Level 3 study in Arts and Humanities.The material is derived from an Open University course (AA308) that is no longer available. The content has been used as the basis for a section of Part 1 of the MA in Philosophy (A853) qualification.

# Keep on learning

## Study another free course

There are more than **800 courses on OpenLearn** for you to choose from on a range of subjects.

Find out more about all our free courses.

## Take your studies further

Find out more about studying with The Open University by visiting our online prospectus.

If you are new to university study, you may be interested in our Access Courses or Certificates.

## What's new from OpenLearn?

Sign up to our newsletter or view a sample.

For reference, full URLs to pages listed above:

OpenLearn – www.open.edu/openlearn/free-courses

Visiting our online prospectus – www.open.ac.uk/courses

Access Courses – www.open.ac.uk/courses/do-it/access

Certificates – www.open.ac.uk/courses/certificates-he

Newsletter – www.open.edu/openlearn/about-openlearn/subscribe-the-openlearn-newsletter

# References

Adams, B., Breazeal, C., Brooks, R. and Scassellati, B. (2000) 'Humanoid robots: a new kind of tool', *IEEE Intelligent Systems*, 15, 25–31.

Block, N. (1995) 'On a confusion about a function of consciousness', *Behavioral and Brain Sciences*, 18, 227–47.

Block, N., Flanagan, O. and Güzeldere, G. (eds) (1997), *The Nature of Consciousness: Philosophical Debates*, Cambridge, Mass., MIT Press, pp. 1–67.

Brooks, R., Breazeal, C., Marjanovic, M., Scassellati, B. and Williamson, M. (1998) 'The Cog project: building a humanoid robot', in C. Nehaniv (ed.), *Computation for Metaphors, Analogy and Agents* (vol. 1562 of Springer Lecture Notes in Artificial Intelligence), Heidelberg, Springer-Verlag.

Carruthers, P. (2000) *Phenomenal Consciousness: A Naturalistic Theory*, Cambridge, Cambridge University Press.

Chalmers, D.J. (1995) 'Facing up to the problem of consciousness', *Journal of Consciousness Studies*, 2, 200–19.

Chalmers, D.J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, New York, Oxford University Press.

Crane, T. and Mellor, D.H. (1990) 'There is no question of physicalism', *Mind*, 99, 185–206.

Dennett, D.C. (1991) *Consciousness Explained*, Boston, Little Brown.

Dixon, N.F. (1971) *Subliminal Perception: The Nature of a Controversy*. New York: McGraw-Hill.

Haugeland, J. (1982) 'Weak supervenience', *American Philosophical Quarterly*, 19, 93–103.

Humanoid Robotics Group, http://www.ai.mit.edu/projects/humanoid-robotics-group/ (accessed 1 November 2007).

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

Levine, J. (2001) *Purple Haze: The Puzzle of Consciousness*, New York, Oxford University Press.

Locke, J. (1961) *An Essay Concerning Human Understanding*, 5th edn, J.W. Yolton (ed.), London, Dent (first published 1706).

McGinn, C. (1989) 'Can we solve the mind–body problem?', *Mind*, 98, 349–66.

McGinn, C. (1999) *The Mysterious Flame: Conscious Minds in a Material World*, New York, Basic Books.

Nagel, T. (1974) 'What is it like to be a bat?', *Philosophical Review*, 83, 435–50.

Papineau, D. (1993) *Philosophical Naturalism*, Oxford, Blackwell.

Papineau, D. (2002) *Thinking about Consciousness*, Oxford, Oxford University Press.

Rosenthal, D.M. (1993) 'State consciousness and transitive consciousness', *Consciousness and Cognition*, 2, 355–63.

Rowlands, M. (2002) 'Two dogmas of consciousness', *Journal of Consciousness Studies*, 9, 158–80.

Smith, A. (1985) *The Mind*, Harmondsworth, Penguin Books.

Sutherland, S. (ed.) (1995) *The Macmillan Dictionary of Psychology*, London, Palgrave Macmillan.

Tye, M. (1995) *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*, Cambridge, Mass., MIT Press.

Weiskrantz, L. (1986) *Blindsight*, Oxford, Oxford University Press.

Weiskrantz, L. (1997) *Consciousness Lost and Found*, Oxford, Oxford University Press.

# Further reading

For a more advanced introduction to the topic of consciousness, which includes an historical survey of philosophical and psychological work on the topic and a survey of recent debates, see:

Güzeldere, G. (1997) 'The many faces of consciousness: a field guide', in N. Block, O. Flanagan and G. Güzeldere (eds), *The Nature of Consciousness: Philosophical Debates*, Cambridge, Mass., MIT Press, pp. 1–67. (The collection in which this essay appears is a useful one, which reprints many important papers on consciousness.)

The following chapters also contain useful introductions – though each reflects its author's own theoretical preoccupations and assumptions:

Carruthers, P. (2000) *Phenomenal Consciousness: A Naturalistic Theory*, Cambridge, Cambridge University Press, Chapter 1. (Introduces some useful distinctions and concepts and provides a route map of contemporary theories of consciousness.)

Chalmers, D. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, New York, Oxford University Press, Chapters 1–2. (Chapter 1 is an accessible introduction to the problem of consciousness; Chapter 2 is technical but includes useful discussion of reductive explanation.)

Dennett, D.C. (1991) *Consciousness Explained*, Boston, Little, Brown, Chapter 2. (Argues against substance dualist approaches to consciousness.)

McGinn, C. (1999) *The Mysterious Flame*, New York, Basic Books, Chapter 1. (Emphasises how strange and inexplicable consciousness can seem.)

Tye, M. (1995) *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*, Cambridge, Mass., MIT Press, Chapter 1. (Introduces phenomenal consciousness and outlines a number of problems surrounding it.)

Finally, a useful general resource, which you may like to explore at your leisure, is David Chalmers's website, at http://consc.net/chalmers. This contains a wealth of material on consciousness, including a list of online papers, bibliographies of philosophical and scientific work and many of Chalmers's own papers.

# Acknowledgements

This course was written by Dr Keith Frankish



Course image: Leonardo da Vinci in Wikimedia available in the public domain.

Reading 1 Chalmers, D. (1996), *The Conscious Mind*, Oxford, Oxford University Press.

Figure 2 Cartoon by James Thurber. Copyright © 1933, 1961 Rosemary A. Thurber.

Adapted from Adam Siegal: http://www.flickr.com/photos/dopeylok/38814742/ [Details correct as of 23rd October 2007]

**Don't miss out:**

If reading this text has inspired you to learn more, you may be interested in joining the millions of people who discover our free learning resources and qualifications by visiting The Open University - www.open.edu/openlearn/free-courses