

Also by Stephen Law
The Philosophy Files

The Philosophy Gym

25 SHORT ADVENTURES IN THINKING

Stephen Law

ILLUSTRATED BY DANIEL POSTGATE

Thomas Dunne Books
St. Martin's Press  New York

What to read next

Chapter 2, *What's Wrong with Gay Sex?*, also discusses the principle that we should never use people as means rather than ends.

Further reading

Jonathan Glover (and others),
Fertility and the Family
(London: Fourth Estate, 1989).

John Harris, *Clones, Genes and Immortality* (Oxford: Oxford University Press, 1992).

13

THE CONSCIOUSNESS CONUNDRUM

PHILOSOPHY GYM CATEGORY

WARM-UP

MODERATE

MORE CHALLENGING

Scientists are grappling with 'the problem of consciousness': the problem of explaining how that walnut-shaped lump of grey matter between your ears is capable of producing a rich inner world of conscious experiences. Will they ever solve this mystery? Some think it's only a matter of time. Yet there are arguments that appear to show that consciousness is something that it is *in principle* impossible for science to explain.

The Private Realm of Consciousness

Take a look at something red: a ripe tomato, for example. As you look at this object, you are conscious of having a certain experience – a colour experience. As the philosopher Thomas Nagel explains,* there's *something it is like* to have this experience, something *for you*, the subject.

We spend our lives immersed in a vibrant flow of such experiences: the smell of a flower, the taste of an orange, the rough sensation of wood under one's fingertips, a zinging pain, a melancholic moment. We can focus our attention on the subjective quality of these experiences and savour them. An interesting feature of this rich inner life we lead is that it seems peculiarly hidden from others. Others can observe my body and outward behaviour, but my experiences are hidden inside. Indeed, they would appear to be 'hidden inside' in a very strong sense. For they are not *physically* hidden, as, say, my brain is physically hidden inside my skull. Things that are physically hidden can in principle be revealed. Surgeons might one day be able to open up my skull and observe what physically goes on inside me when I have a colour experience. But they can never enter my mind and observe what the experience is like *for me*, from my point of view.

* Thomas Nagel, 'What Is It Like to Be a Bat?', in Douglas R. Hofstadter and Daniel Dennett (eds), *The Mind's I* (London: Penguin, 1981).

What Is It Like to Be a Bat?

There are also conscious experiences no human being has ever enjoyed. Take bats for example. Bats manage to find their way around in the dark by using echo-location. The bat emits a sound (inaudible to humans), and the loudness of the echoes and the direction from which they come allow the bat to build up a picture of its environment.

Echo-location allows bats to 'see' using sound. Now ask yourself: what must it be like to be a bat, to experience the world as the bat does? No doubt there is something it's like for the bat when it 'sees' using echo-location. It must be a very strange experience, radically unlike any of the experiences we humans enjoy. But, as Nagel points out, we can't know what the experience is like. We could discover everything there is to know about what happens in a bat's nervous system when it 'sees' using sound. But that still wouldn't allow us to know what the experience is like for the bat. Its subjective character would remain unknown to us. It seems the bat's experience, like yours and mine, is essentially private.

The realm of conscious experience is responsible for what continues to be one of the most profound and intractable of mysteries, a mystery with which both philosophers and scientists are currently very much engaged. The mystery concerns how our physical bodies and our conscious minds are related. The problem, as we shall see, is that, on the one hand, it seems your conscious mind must be physical, yet, on the other hand, it seems it cannot be.

Two Competing Theories of Consciousness

Scientists tell us that when you looked at that red object a minute or so ago, the following happened. Light of certain wavelengths was reflected off the object into your eye, where it was focused on to your retina to produce an image. Your retina is covered with millions of light-sensitive cells, some of which are sensitive to differences in wavelength. The light falling on to these cells caused them to emit electrical impulses which then flowed down the nerves linking your eye to the back of your brain. That caused something to happen in your brain.

But what about your experience? According to the philosopher René Descartes (1596–1650), your conscious mind is a distinct entity capable of existing on its own, independently of anything physical. So, in Descartes' view, after something happened in your brain, something else had to happen: your brain

caused something to happen in your mind. Your mind and brain may *interact*. But they are not *identical*.

According to many contemporary scientists and philosophers, however, it's a mistake to think of conscious experience in this way. Professor Susan Greenfield, for example, in her BBC television series 'Brain Story', insists that 'you are your brain'. Your experience isn't something extra – something on top of what happens physically. Rather, the mental *just is* part of what's going on physically.

Certainly, scientists sometimes reveal that what might seem like two distinct things are actually one and the same thing. Take the morning star and the evening star, for example. For a long time we thought these heavenly bodies were distinct. Then astronomers discovered that they are one and the same object seen twice over (the planet Venus).

Scientists have also established that certain properties are identical. For example, they have discovered that heat is a molecular motion, electricity is a flow of electrons and water is H₂O.

So why shouldn't it also turn out that pain just is a certain state of the brain? Admittedly, pain doesn't seem like a brain state. But so what? After all, heat doesn't seem like molecular motion – yet that's just what it is.

Substances and Properties

We have been looking at two competing theories about consciousness. First, there are those who believe that your conscious experiences are nothing over and above what goes on in your brain. Secondly, there are those who, like Descartes, deny this. But before we get to the arguments for and against these two positions, it will be useful if we distinguish two rather different versions of the second position.

In Descartes' view, your mind and body are distinct *substances*: each is capable of existing independently of the other. Your conscious mind could, in principle, be detached from everything physical and exist on its own. This position is called *substance dualism*.

Hardly any scientists or philosophers are now prepared to accept substance dualism. But there are still plenty of philosophers (and at least some scientists) around who believe that there are two distinct and irreducible kinds of *property*: physical properties and mental properties. This position is called *property dualism*.

According to property dualism, there's only one kind of stuff – physical stuff. But objects made out of this physical stuff can have two quite different sorts of

property. In the view of the property dualist, there are both mental properties and physical properties: the mental properties of a human being are extra properties that exist in addition to all of his or her physical properties.

An Argument Against Dualism

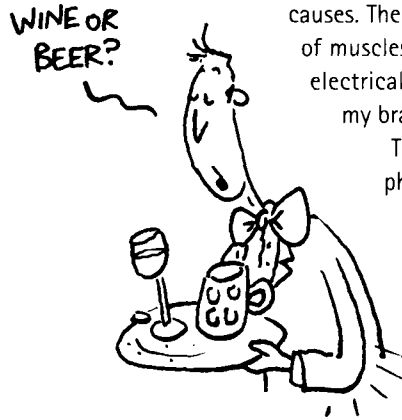
Let's now turn to one of the most popular arguments against all forms of dualism.

In effect, dualists want to introduce an extra layer of facts in addition to the physical facts. There are facts about physical substances and properties. But according to dualists, there are also non-physical substances and/or properties. The facts about these non-physical substances/properties are facts *in addition to* the physical facts. So there are two fundamentally different and irreducible sorts of fact.

Many scientists and philosophers consider the suggestion that there are such 'additional' facts thoroughly unscientific. Why is this?

Suppose that at a dinner party I am given the choice between a glass of wine and a glass of beer.

I like both, but decide on this occasion to have wine. I reach out and select a glass of white.



Scientists tell us that such physical movements have physical causes. The movement of my arm was caused by the action of muscles in my arm, which was itself brought about by electrical activity in the efferent nerves running from my brain.

This electrical activity was in turn caused by physical activity in my brain, which was brought about by further preceding physical causes, including the stimulation of my nervous system by light reflected off the glasses on the tray in front of me and the sound of someone speaking to me. These physical causes in turn had physical causes, which in turn had physical causes, and so on.

Indeed, it seems that if scientists were furnished with knowledge of the laws of nature, plus all the physical facts about my body and my environment as they were, say, a minute prior to my deciding to reach out and grasp that glass of wine, it would be possible *in principle* for them to figure out that my arm would

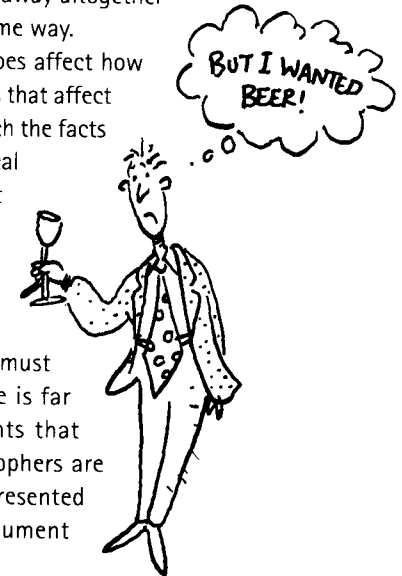
do what it did. That movement of my arm was fixed in advance by how things stood physically.

But if this is correct – if what happens physically is fixed in advance by the preceding physical facts – then there is no possibility of any non-physical fact affecting how things turn out. The non-physical must be causally irrelevant to what goes on physically.

But if dualism is true, then my conscious mind is non-physical. But then it follows that *my mind can make no difference to what goes on physically*. Suppose, for example, that I had decided to pick up a glass of beer instead. Because of the physical facts, my arm would have been compelled to reach out and grasp that glass of wine anyway. Indeed, if dualism is true, you could take my mind away altogether and my body would *still* carry on in exactly the same way.

But this is absurd, surely? My mind can and does affect how my body behaves. But as it is only the *physical* facts that affect how things turn out physically, the only way in which the facts about what happens in my mind can have a physical effect *is if they are themselves physical facts*. But then it follows that dualism (both substance and property) is false.

Many scientists and philosophers are convinced by this and other arguments that the facts about what goes on in the conscious mind must ultimately be physical facts. However, the issue is far from settled. There are also powerful arguments that appear to show that these scientists and philosophers are mistaken. One of the best-known arguments is presented by the philosopher Frank Jackson. Jackson's argument runs as follows.



Mary and the Black and White Room

A girl called Mary is born. Before she has any visual experiences, Mary is placed in a black and white room by scientists who wish to study her. The scientists arrange that Mary never has a colour experience (they hide Mary's pink hands from her by using white gloves, and so on). Mary experiences only black, white and shades of grey.

Mary grows up in her black and white environment. She develops a fascination with science. Indeed, Mary eventually becomes the world's greatest brain expert. She finds out everything there is to know about what goes on inside a human's brain when they describe what they are seeing as 'red'. She discovers all the physical facts about the brains of colour perceivers: how their neurons are firing, how the brain chemistry is balanced, and so on.

Then, one day, one of the scientists studying her brings a ripe tomato into her black and white world.

Mary is stunned. She now has an experience that she's never had before. She finds out what it is like to have a colour experience. Mary discovers a new fact: the fact that the experience of red is *like this* (I'm looking at that red object again). But Mary previously knew all the physical facts. So the fact that the experience is *like this* is not a physical fact. Facts about the qualitative character of our conscious experiences – about what it is like to have them – are not physical facts.

The Explanatory Gap

Jackson seems to have shown that there are more facts than just the physical facts. But there's a further conclusion you might wish to draw. Jackson's story also appears to show that *not everything can be explained or understood by science*. We can't explain or understand why red things look *like this* by appealing only to the physical facts about us. We come up against what contemporary philosophers call an *explanatory gap* at this point.

Contrast the case of heat. Identifying heat with vigorous molecular motion allows us to deduce the various properties of heat. Discovering what's going on at the molecular level allows us to understand why objects that are heated char and blacken, why they tend to make nearby objects hot, and so on.

But a full understanding of the goings-on in the human brain will not allow us to understand why pain feels the way it does or explain why ripe tomatoes produce *this* sort of visual experience. Mary knows everything there is to know about what goes on in the brains of colour perceivers, but this knowledge does not allow her to understand what an experience of red is actually like. Indeed, none of the physical facts she discovers go any way towards explaining why such physiological states should be accompanied by conscious states *at all*.

The Analogy with Life

Jackson's argument appears to show both that

1. there are more facts than the physical facts, and
2. it is *in principle* impossible for the physical sciences to account for consciousness.

But many scientists are dismissive of such conclusions. They often suggest that the current situation with respect to consciousness is similar to the situation 200 years ago with respect to life. Life at that time constituted a great mystery. We simply had no idea how mere physical matter could be organised in such a way as to produce an animate, living thing. Many thought that something extra – a mysterious and supernatural 'vital force' – had to be added to a physical object in order to imbue it with life.

Today, of course, the explanation of life is mostly within our grasp. Darwin's theory of natural selection, advances in genetics, and so on, have allowed us to explain many of the properties of life. Even where a scientific explanation of some particular feature of life currently eludes us, we can at least now see how such an explanation might in principle be constructed just by appealing to physical facts.

Many scientists argue that, similarly, just because a scientific explanation of consciousness *now* eludes us doesn't mean that no such explanation is possible. There's no need to suppose that consciousness must be something mysterious and supernatural that exists *in addition to* what we find within the natural, physical world. These are early days in the scientific investigation of consciousness. Our current inability to imagine how consciousness might be explained by appealing only to physical facts may simply be due to our lack of an adequate theory, just as in the case of life.

Conclusion: A Mystery

We have been grappling with the mystery of how to accommodate consciousness within the physical universe. Many scientists believe that consciousness must ultimately be reducible to and explicable in terms of the physical. Indeed, given that the conscious mind is able causally to affect what goes on physically, it seems it must itself be physical.

But there are powerful objections to this belief. Jackson's story about Mary and the black and white room seems to show that it is in principle impossible for the facts about the character of our conscious experience to be reduced to and explained in terms of physical facts. It seems there must be more facts than just the physical.

Many scientists reject all forms of dualism out of hand. But unless they can show what is wrong with Jackson's argument (and, indeed, the other very convincing-looking arguments that are around), their dismissive attitude towards dualism looks hasty. Blindly to reject such arguments looks more like prejudice than a rationally held position.

Of course, it may be that there's something wrong with Jackson's argument (see the box below). But the onus is on those who reject all forms of dualism to show precisely *what* is wrong with it. And, of course, showing what's wrong with such arguments is the job not of empirical science, but of logic and philosophy.

So can science ever solve the mystery of consciousness? The answer is: perhaps, but not by itself. Science will need the help of philosophy.

Thinking Tools: The Masked Man Fallacy

This section explains what may be wrong with Jackson's argument. There is a popular form of argument often used to establish that two things are not identical. You search for a property that one of the two things has that the other lacks. If you can find such a property, it follows that the items in question are non-identical.

For example, if you want to establish that K2 and Everest are distinct mountains, all you need to do is to find a property one mountain possesses that the other lacks. You might argue like this:

- Everest has the property of being over 29,000 feet high.
- K2 doesn't have the property of being over 29,000 feet high.
- Therefore Everest is not identical to K2.

This is a sound argument: each of the two premises is true, and the logic is impeccable. The argument really does establish that Everest and K2 are distinct.

Those who argue that mind and body are non-identical often appeal to the same form of argument. Here, for example, is an argument (often attributed to Descartes) called *the argument from doubt*:

- I don't doubt that I exist. After all, by trying to doubt that I exist, I demonstrate that I do exist, so my attempt at doubting that I exist must inevitably be self-defeating.
- I do doubt that my body exists. It seems to me that I might be a disembodied mind, with all my experiences being generated by some sort of malevolent demon (for more on this sort of doubt, see Chapter 3, Brain-Snatched).
- But then it seems that my body has a property that I lack: my body has the property of being *something the existence of which I doubt*. I lack this property. So it surely follows – by an argument analogous to that about Everest and K2 – that I'm not identical with my body.

Here's the argument laid out more formally:

- My body possesses the property of *being something the existence of which I doubt*.
- I don't possess the property of *being something the existence of which I doubt*.
- Therefore I am not identical with my body.

This sort of argument has convinced many that mind and body are non-identical. But despite the similarity to the Everest/K2 argument, this is a bad argument. What we have here is an example of the *masked man fallacy*. Here's another example of the fallacy. Suppose I witness a bank being robbed. This leads me to believe that the masked man robbed the bank. Later, detectives inform me that their lead suspect is my father. Horrified, I try to prove that my father cannot be the masked man. I point out that the masked man has a property my father lacks: he's someone I believe to have robbed the bank. I argue like this:

- The masked man has the property of *being someone I believe robbed the bank*.
- My father lacks the property of *being someone I believe robbed the bank*.
- Therefore my father is not identical with the masked man.

This is obviously a bad argument, despite sharing the same form as the sound Everest/K2 argument. It could yet turn out that my father *is* the

masked man, despite the fact that both premises are true. Why is this?

The problem is that this form of argument does not work for *all* kinds of property. It works for properties such as being more than 29,000 feet high. It does not work with properties such as being someone I believe to have robbed the bank. More generally, this form of argument is invalid whenever the property in question involves *someone's psychological attitude towards a thing*.

For example, in the masked man case, I try to show that my father and the masked man are distinct by pointing out that I have an attitude towards one that I don't have towards the other: I believe one robbed the bank but not the other. But such attitudes are incapable of revealing whether or not the items in question really are distinct. Here are a couple of other examples:

- John Wayne is someone Michael knows appeared in *True Grit*.
- Marion Morrison is not someone Michael knows appeared in *True Grit*.
- Therefore John Wayne isn't Marion Morrison.
- Heat is widely recognised as something with which to cook food.
- Molecular motion is not widely recognised as something with which to cook food.
- Therefore heat isn't molecular motion.

Both these arguments have true premises but false conclusions ('John Wayne' is the stage name of Marion Morrison). The problem, again, is that what someone may know or believe or recognise about one thing but not another is not the sort of property one can use to establish the non-identity of those things. The argument from doubt involves the same fallacy.

What of Jackson's argument about Mary? Does it also involve the masked man fallacy? I think that, as it stands, it does. But you should check for yourself. Of course, none of this is to say that I believe dualism is now defeated. There may be better arguments for dualism than Jackson's, arguments that don't involve the masked man fallacy.

What to read next

This chapter might usefully be read in conjunction with Chapter 6, *Could a Machine Think?* Look for where some of the arguments overlap.

In Chapter 15, *Do We Ever Deserve to Be Punished?*, I briefly discuss the discovery that the universe is not after all governed by strict and exceptionless laws, but merely by probabilistic laws. So it turns out that the most that someone furnished with full information about my physical body and environment could ever predict about my future behaviour is what I will *probably* do. After reading Chapter 15, you might wish to return to this chapter to consider the question: does this discovery undermine the argument against dualism presented above? Even if it does, might some version of that argument still be salvaged?

Further reading

Jackson's story about Mary and the black and white room appears in:

'Epiphenomenal Qualia', in W. Lycan (ed.), *Mind and Cognition* (Oxford: Blackwell, 1990).

For a breezy and yet quite thorough introduction to the problem of consciousness, see:

David Papineau and Howard Selina, *Introducing Consciousness* (Cambridge: Icon, 2000).

An interesting collection of pieces on the mind can be found in the now quite old but nevertheless still excellent:

Douglas R. Hofstadter and Daniel Dennett (eds), *The Mind's I* (London: Penguin, 1981).

The Mind's I includes Thomas Nagel's famous paper, 'What Is It Like to Be a Bat?', which is also included as Chapter 38 of Nigel Warburton (ed.), *Philosophy: Basic Readings* (London: Routledge, 1999).