Philosophy and the Human Situation
Functionalism about the mind

Alex Barber
For a long time, no convincing materialist alternative to Cartesian dualism was available. There were plenty of unconvincing materialist alternatives, and plenty of people willing to advocate them. But they usually did so only because the theories they were advocating were so vague or obscure as to be next to impossible to refute. It wasn't until the late 1960s and early '70s that a genuinely viable alternative to Cartesian dualism came onto the scene. Several papers published around that time, had titles like "The Mental Life of Some Machines", and hinted at a radically different way of thinking about the mind. *Functionalism,* as this way of thinking is now known, quickly became the dominant perspective in the philosophy of mind; it has retained this dominance ever since.

In spite of its huge popularity, the central claims of functionalism can be tricky to grasp. You may find it comforting to keep in mind that discussions in this area can become confusing very quickly, even to people who have spent years participating in them. Do not be alarmed if by the end of the tape you find that you've not taken in all of the points covered, down to the smallest detail.

I'm going to approach functionalism along a deliberately indirect and easygoing route. I'll set the stage by describing two materialist positions out of which functionalism developed in the last century: the *mind-brain identity theory*, and *behaviourism*. These alternative positions are relevant because functionalism managed to keep all the advantages of these earlier theories but drop all their shortcomings. Or rather, that's what functionalists are fond of saying. The truth is that it is possible to regard functionalism as a sophisticated version of either of these earlier theories; I'll be exaggerating the differences for dramatic effect, so as to make clear what exactly functionalism is.

The first ancestor of functionalism I want to talk about is the mind-brain identity theory, which developed in Australia in the 1950s. There seems to be something about the Australian psyche that puts a high value on straightforwardness, and Australian philosophy of mind certainly conforms to this stereotype. Mind-brain identity theorists offer an extremely straightforward materialist account of what mental states are. According to them, mental states are simply *physical states of the brain*.

An example of a mental state is that of being thirsty. To be thirsty, according to the mind-brain identity theory, is for one's brain to be in a particular physical state. Exactly *which* physical state is something that scientists are going to have to uncover for us, using brain scanners and other experimental paraphernalia. Another example of a mental state is the state of *believing that the kettle has just boiled*. To be in this state is for one's brain to be in some *other* physical state, though which other physical state is once again something for the emerging brain-sciences to work out. A popular example from the 1950s, was of the mental state of pain. This was said to be identical with the state of having C-fibres firing off in one's brain. As it turns out, this is no longer plausible even from a narrowly scientific perspective. A scientifically more up to date mind-brain identity theorist could deploy a different example, maintaining that visual consciousness consists in synchronised neuronal firings of around 40 Hz in regions linking the thalamus with the visual cortex. This example, too, is already finding its way into the cupboard of scientific history. But the hope is that mental states may *one* day come to be correctly identified with *something like* these physical brain states. This hope, when transformed into expectation, is what lies behind the appeal of the mind-brain identity theory.

The mind-brain identity theory has two attractive features, but suffers from what is widely regarded as a knockdown objection.

The first attractive feature is that the problem of mind-body causation, the problem that dogs Cartesian dualists, simply evaporates. The problem of mind-body causation should be familiar and it goes something like this. We're all familiar with one physical event causing another physical event. A falling roof tile can cause a windscreen to break, for example. But we need also to be able to make sense of causal interactions between physical events and mental events. My decision to tie my shoelaces is a mental event, and it causes my fingers to move in a certain way, which is a physical event. This physical event, in turn, causes me to see that my laces are tied and to feel satisfied by my achievement. My seeing, and my feeling satisfied, are both mental events. The problem with dualism is that the more it emphasises the separateness of what it regards as two realms, the mental realm and the physical realm, the harder it becomes to see how causal processes could bridge the gap. Many people believe that there is no remotely plausible way for dualists to address this concern.

Mind-brain identity theorists have an elegant way around the problem of mind-body causation. Mental events simply are physical events, they say. Specifically they are physical events involving the brain. The mental event of deciding to tie my shoelaces is the physical event of having particular synapses going off in my brain. So for them, the mental "realm" is really just a sub-part of the physical realm, which is the only realm there is. When a physical event causes a mental event, it's merely causing another physical event and since there's nothing puzzling about one physical event causing another physical event, there's nothing puzzling about a physical event causing a mental event.

I said before that there's a second attractive feature of the mind-brain identity theory, but I'll come back to this in a discussion of behaviourism. Before that, I want to present a notorious objection to the mind-brain identity theory. This objection is known as the multiple-realisability objection. It's a useful objection to know about because it challenges anyone who agrees with the mind-brain identity theory, including those many people who agree with it without realising they do.

I am thinking of people who take it as just obvious that mental states are states of the brain. Being in love, these people might say, is just a characteristic neurophysiological response to a particular hormone; wanting to climb Everest is just for one's brain to posses certain as-yet-unknown physical properties. You may agree with these sentiments. If so, see what you make of the objection. Here it is.

Think about whether something could be made out of a completely different physical stuff than we are, and yet still be capable of having a mental life like our own. Try to imagine an alien organism made out, not of carbon-based materials like we are, but of silicon- or aluminium-based materials, or some unspecified goo. Could such an organism have thoughts? Could it have desires or fall in love? I think: yes, it could – and I hope you think this too. Anyone who thinks that such a creature is not possible is being objectionably provincial or chauvinist in assuming that carbon-based organisms, and humans in particular, are the only entities capable of thinking, desiring, or loving. Most of us are attracted to the more cosmopolitan view according to which organisms made of different physical stuff than us could in principle sustain a mental life much like ours.

These envisaged alien organisms are not in the same physical brain state as us – how could they be if they are made of a different physical stuff? This shows that being in a particular physical brain state is not essential to being in a particular mental state; yet this is exactly what mind-brain identity theorists claim. To make a deliberate pun, the mind-brain identity theory is too narrow minded about what is involved in having mental states.

I said that this objection is known as the multiple-realisability objection. A word of warning is appropriate, here: the English world 'realise' is ambiguous and is being used with its less common meaning. More commonly, it means 'come to know', as in 'I realised he wasn't my biological father after all'. The less common usage is illustrated in the sentence 'The United States was the first country to realise the political ideals of the French enlightenment'. 'To realise' in this sense means something like to implement or, more literally, to make real.

With that clarification, it should be clear why the objection just sketched is called the multiple-realisability objection. Our intuitions tell us that completely different kinds of physical stuff could realise the same mental state; this shows that mental states cannot be identified with any particular physical brain state. Your belief that cassette tapes have two sides is realised in carbon-based human-brain goo; an alien organism's beliefs about the double-sidedness of cassette tapes is realised in different stuff than the stuff it is realised in, in you. The mind-brain identity theory is incompatible with this multiple realisability of mental states.

We'll see later on how functionalists are immune to this multiple realisability objection.

The second ancestor of functionalism I want to look at is behaviourism. My presentation of behaviourism is inevitably going to be something of a caricature, and behaviourism in this crass form has largely died out. Very few people these days call themselves behaviourists. On the other hand the legacy of behaviourism is very much alive and various mutations of the original idea still exist but are known under different names.

Behaviourists claim that to be in a mental state is to have a disposition to behave in a certain way. Which way will depend on which mental state it is we're talking about, but the basic idea can be illustrated with the example of thirst. To be thirsty is to be disposed to reach out for water and bring it towards one's mouth. What being thirsty is definitely not is what Cartesian dualists say it is: a mysterious inner state or feeling, directly known only to the person in that state.

Talk of "behavioural dispositions" can be understood in terms that will be familiar. Someone is in a particular mental state, that is, they have a particular behavioural disposition, if a certain conditional, an if-then proposition, is true of them. For you to be thirsty is for the following conditional to be true of you:

> If water is put in front of your eyes, then your hand will reach out to that water.

Your thirst is not a hidden, inner state that causes this conditional to be true of you. Your thirst consists in this conditional being true of you.

I'm now going to present two apparent advantages of behaviourism, and after that, one major shortcoming.

Advantage number one is this. If behaviourists are right about what mental states are, then the study of mind can be put on a secure scientific footing. And in fact behaviourism is often described even by its critics as the earliest science of the mind. What lies behind this generous description is that behavioural dispositions are publicly observable and hence available for objective empirical investigation. Scientists can test attributions of mental states by watching to see whether or not the relevant conditionals are true. They do not need any kind of magical telepathy, mind-reading machine, to work out what people have on their mind. I can test to see if you are thirsty by putting water in front of you and seeing if you reach out for it. This contrasts favourably with the Cartesian view of mental states as something locked away and so unavailable for objective study.

A second advantage of behaviourism is its capacity to accommodate the multiple realisability of mental states, a problem for the mind-brain identity theory as we saw. Behaviourists have no difficulty with multiple realisability because things made out of very different kinds of stuff – carbon-based goo or silicon-based goo – could nevertheless possess exactly the same range of behavioural dispositions, and so the same range of mental states.

The main disadvantage of behaviourism may well have occurred to you already. Can it really be true that being thirsty is the same as having a behavioural disposition to reach out if water is put in front of you? It's not difficult to produce counterexamples. You could be thirsty but not reach out because you think the water is poisoned, or because you want to test your ability to withstand thirst. Rabid dogs, supposedly, are desperately thirsty but do not drink water because they are hydrophobic. Conversely, just because you have a disposition to reach out for water doesn't mean that you're thirsty. You could be pretending to be thirsty.

Or perhaps your earlier thirst has been thoroughly quenched already but, you think that more water is always better for you, and so you just keep on drinking.

Pressing home this objection is harder than you might think. Behaviourists have a ready reply to counterexamples like the ones I just came up with. Thirst, they insist, is a behavioural disposition; it's just not the disposition to reach out for water. That was just meant to illustrate the basic idea. Really, they say, thirst is a far more complex disposition, namely, the disposition to reach out for water when and only when certain other factors are not interfering, such as delusional views on the health value of never-ending amounts of water.

Against this quick response, it's tempting to ask a behaviourist to tell us more about what these supposed 'other factors' are. If we're not told, the theory appears to lose the precision we expect of properly scientific theories, as behaviourism is claimed to be. But the appeal to complexity looks particularly hopeless in the face of a more extreme counterexample, the case of the bodiless-brain.

Imagine we're in the future. Medicine has developed, but so have a number of virulent diseases affecting the human body. Someone, call him Fred, is suffering from such a disease and has opted to have his limbs carefully amputated, one by one. Next, his torso is surgically removed and disposed of, followed eventually by his face and skull until all that remains of Fred is a brain, floating in a vat of nutrients at a life-sustaining temperature. When the anaesthetic is removed, Fred feels as if he has been paralysed and left in a sensory-deprivation chamber. At first he's just happy to be alive, but he quickly becomes bored and depressed. Eventually he reconciles himself to his predicament, and has a reasonably rich mental life, filled with reflective, emotional, and imaginative episodes.

The plausibility of this scenario suggests that mental life cannot be equated with a disposition to behave. Fred, in his vat, is someone of whom the conditionals are clearly false – he simply doesn't have any behavioural dispositions. Yet he is given to thought, emotion, imagination, and so on. According to behaviourists, this ought to be impossible. So much the worse for behaviourism.

I said earlier that there is a second advantage to the mind-brain identity theory. The second advantage is simply that bodiless brains are not a counterexample to the mind-brain identity theory. If a mental state is a physical state of the brain, then a disembodied brain can still have thoughts and emotions. The fact that Fred's brain is not affiliated to any body is simply irrelevant.

What we need, clearly, is a position that is not threatened either by multiple realisability, or by bodiless brains. It's at this point that functionalism comes to the rescue.

Functionalism is a theory of what mental states are. Functionalists don't regard mental states as states of an immaterial substance, or as physical states of the brain, or even as dispositions to behave in a certain way; rather, mental states are said to be functional states.

This is not much use until we have answers to two questions. First, we need to know what a 'functional state' is, in general. Second, we need to know which particular functional states the various mental states are to be identified with. I'll take these questions in turn. Along the way we'll see how it is that functionalism keeps the advantages of both behaviourism and the mind-brain identity theory but avoids their shortcomings.

The general notion of a functional state is perhaps best introduced using non-mental examples, since there are plenty of functional states that have nothing much to do with the mind. Two examples are: the state of being a heart; and the state of being a post-box.

Being a heart is not the same thing as being a lump of muscle-tissue in the centre of a ribcage. To be a heart is to serve a specific function, the function of pumping blood in an organism. This function could conceivably be performed by all sorts of things other than the things that, as it happens, beat away inside our ribcages. For example, something made of titanium strapped to a person's hip could conceivably serve as a heart…could conceivably be

a heart. The state of being a heart is said to be a functional state, the state of functioning to pump blood around the relevant body. The thing that does the pumping can be made of anything whatsoever: it will still be a heart just as long as the pumping gets done.

The second, non-mental, example of a functional state is that of being a post-box. Post-boxes, in the UK at least, are metallic, they're red, and they have a royal crest on them. But being red, metallic, and so on is not what being a post-box is. Post-boxes could have been made of blue plastic, with no royal crest. Had they been designed like this, they would still have been post-boxes. Conversely, something could be red, metallic, and royal crested without being a post-box. It could be a fake post-box, designed and installed by postal thieves. So, what is it to be a post box? The answer seems to be that to be a post box is to have a certain role or function – roughly, the function of being a thing within which items can be left for collection by the postal service. Red metallic boxes fill this function in actual fact, but other kinds of thing might easily have filled it, and perhaps will do, sooner or later.

In general, functional states are states of doing something, of fulfilling some role, of serving some function in a wider system. The functions in our examples are of pumping blood or of being a thing in which items can be left for postal collection. Functionalists claim that mental states are like this too. Mental states, they say, are states that play a particular role; they serve a particular function in our mental activity.

Typically, something quite different from the thing that actually serves a particular function could have served the function equally well. The things that pump our blood could have been made, not of muscle tissue, but of titanium. The things we deposit our letters in could have been made, not of red metal, but of blue plastic. Equally, the things that enable us to think, imagine, dream, hope, and so on, happen to be made of carbon-based material; but they didn't have to be so.

Already we can see why functionalism is immune to the multiple-realisability objection. It's a characteristic of functional states that they can be realised in many different ways. There are many different ways for something to be in the functional state of being a post-box, for example. So, too, there are many different ways a belief that tape cassettes are double sided could be realised, so long as we grant that belief states are functional states. The carbon-based, human way is one way, but the silicon-based, space-alien way could be just as viable. Being made of different physical stuff does not have to mean having different mental states, so long as mental states are functional states. This accords well with our broad-mindedness about what sorts of stuff intelligent things could be made of.

I want to turn now to how functionalism differs from behaviourism, from dualism, and from the mind-brain identity theory. Recall what behaviourists claim about thirst, namely that to be thirsty is to have a behavioural disposition – something like the disposition to reach for water. Functionalists want to agree with behaviourists that thirst must bear some relation to behavioural dispositions of this sort. But they disagree about what the relation is. Functionalists will say that thirst is an inner state that typically causes this behavioural disposition. Behaviourists think this is mistaken. They are keen to avoid appeals to inner states, because inner states would be hidden from scientific investigation. Behaviourists say that being in a state of thirst is having the behavioural disposition; there is no need to treat the state of thirst as an inner cause of the disposition.

There's an assumption implicit in behaviourists' hostility towards inner causes, and functionalists reject it. Functionalists deny that the only way to find out about something in science is to observe it directly. Scientists routinely make claims about states of affairs that they can't directly observe. Think about the very plausible claim that dinosaurs once roamed the earth. This is not something we can directly observe. All we have are the fossils. But we are not tempted to say that what it means to claim that dinosaurs once roamed the earth is that present-day investigators are disposed to uncover skeleton-like objects in rock strata. Behaviourists may have come up with the first scientific theory of the mind, but they had a very warped idea of what science must be like.

In allowing mental states to be regarded as hidden inner states that cause behavioural dispositions, functionalists distinguish themselves from behaviourists. But doesn't this just turn them into dualists?

Actually, functionalism is compatible with dualism. If it turned out that the states that cause behavioural dispositions belong to an immaterial realm, then both dualism and functionalism would be correct about what mental states are. What distinguishes functionalists from dualists is that functionalists don't have to believe that the inner states that give rise to our behavioural dispositions are states of an immaterial mind. Indeed, because of familiar anxieties about mind-body causation, hardly any functionalists combine their functionalism with dualism. Instead, they say that the states that actually causally underpin our behavioural dispositions are inner physical states of the brain.

But now it sounds like functionalism is indistinguishable from the mind-brain identity theory. After all, saying that mental states, the states that cause behavioural dispositions, are physical states of the brain is just what mind-brain identity theorists say.

The difference is that functionalists allow that something other than a physical state of a carbon-based brain could play the same functional role. Carbon-based physical brain states are merely a common local realisation of mental states. Mind-brain identity theorists, by contrast, are committed to claiming that mental states are identical with carbon-based brain states; they are hence committed to claiming that nothing is or could be in a mental state unless it was in this brain state. Functionalists will often agree about what kinds of thing actually do the causing in us. But they insist that what makes an instance of a mental state the mental state that it is, is its causal role and not its physical make up.

Functionalists say that mental states are functional states, states of an individual that serve a particular causal function in their mental activity. The difficult task will be to try to say what kind of functional states particular mental states might be. We've already heard some initial suggestions Thirst, it was said, is a state that normally causes certain behavioural dispositions. We might add that thirst is normally caused by a lack of water in one's body, and normally causes one to think that one is deprived of water. In other words, the state of thirst is characterised by the typical causal links it has to other states, including both states of the external world and other mental states.

This pattern generalises. Mental states are identified with a characteristic causal role, involving causal connections to physical states as well as to other mental states.

Functionalists don't try to pretend that the causal roles with which mental states can be identified are simple or uniform. On the contrary, they admit straightaway that the causal roles will be hugely complex and varied, and unsurprisingly so. Minds, and the brains that realise them, are complex, flexible things. An example will bring this out.

Consider the mental state of believing that a kettle has just boiled. What is the role of such a belief? What function does it serve in the life of someone who holds it? One important function of our beliefs is to guide our actions. But the way beliefs guide actions is not at all simple. You might think that a belief that the kettle has just boiled would guide the believer into pouring the kettle's contents into the waiting teapot. But although the belief may typically cause this action, it will generally do so only if other background conditions hold. If the believer actually wants cocoa and not tea, or if she thinks the kettle contains vodka rather than water, then her belief that the kettle has just boiled is unlikely to cause her to pour its contents into the teapot.

Another important function of beliefs is to be caused by appropriate evidence. But the same complexity arises here, too. A belief that the kettle has just boiled will normally have been caused by the sight of steam, the sound of whistling, and so on. But imagine someone who is ignorant of the fact that steam is associated with boiling water, and who thinks that the sound of whistling is just a trick. The sights and sounds of the kettle wouldn't cause this person to believe that the kettle has just boiled.

This example shows just how difficult it will be to give plausible statements of the causal roles that characterise mental states. Functionalists are not put off by this. They still maintain that mental states are functional states. But they seek to describe the broad structure or architecture of the mind rather than the nuances that we may never be able to pin down exactly.

I've not said how functionalism escapes from the bodiless-brain objection to behaviourism. In fact it's not immediately clear how it can escape the objection. After all, the causal roles that characterise mental states often require links with bodily behaviour; and Fred's brain, being bodiless, could have no such links. This seems to commit functionalists to the implausible view that bodiless brains could not have mental states.

Functionalists have a reply to this kind of case. They can say that so long as a state would normally play a particular functional role, that's enough to make it the mental state that it is. It's just that being a brain in a vat is not normal. An analogy should help. Chairs are things whose normal function is to be sat on. This doesn't mean that a chair, spinning in outer space where it couldn't be sat on, wouldn't be a chair. Something is a chair so long as, were it in a normal context, were it not spinning in outer space for example, its function would be to be sat on. Likewise, the physical states of a bodiless brain could not in that context serve a functional role defined in terms of causal links to behaviour. But these physical states still count as mental states by virtue of the fact that they could fulfil such a role; that is, they could if they were housed in a skull on a human body and not in a vat of fluid.

I'll finish this brief introduction to functionalism with a quick summary. Functionalism developed out of two earlier alternatives to dualism. Unlike behaviourism and the mind-brain identity theory, functionalism is said to be immune to both the multiple-realisability objection, and to what I have called the bodiless-brain objection. It is also possible, and usual, for functionalists to be thoroughgoing materialists, and so to avoid the puzzles faced by dualists over mind-body causation.

What functionalists claim is that something is a mental state by virtue of its serving a particular causal function. What actually serves this function in us is almost certainly a state of our carbon-based brain. But there could be mental beings for whom this is not the case.

The functional roles that functionalists have in mind are extremely varied and complex. That's to be expected given the rich diversity among our mental states. What mental states do all have in common, and this if anything is the key insight, is that being in them doesn't require that one be made out of human flesh.