Mental causation[1]

SUBSTANCE DUALISM

Substance dualism holds that there are two fundamentally different types of substances. In traditional dualism, these two types of substances are physical substances ('bodies', physical objects) and mental substances (minds). Minds are distinct from bodies – they are not bodies, they are not parts of bodies, and because they are substances, they are not properties of bodies either.

In the *Meditations*, Descartes claimed that mind and body causally interact with one another. Walking, talking and other bodily movements, are caused by thoughts, decisions and feelings, and we feel pain from physical causes and acquire beliefs from our sense experience. Indeed, nothing seems more obvious than that the mind and the body interact with each other, e.g. I decide to phone a friend and move my body to do so.

But how is it that a mental substance, which, according to Descartes, is not in space and has no physical force, can affect a physical substance, which is in space and moved by physical forces?

P1. The movement of a physical object is only initiated by some physical force, exerted at some point in space.
P2. If dualism is true, then the mind is not in space and cannot exert any physical force.
C1. Therefore, if dualism is true, the mind cannot cause any physical object to move.
C2. Therefore, either dualism is false or the mind cannot cause (any part of) the body to move.
P3. The mind can cause the body to move.
C3. Therefore, dualism is false.

The challenge is just as daunting when thinking about how physical objects could cause changes in the mind. How can something which is not thought or consciousness bring about changes in a substance that is entirely thought and consciousness? Physical causation operates, as we said, through the exertion of forces at particular points in space. But it seems impossible to exert a physical force on a mental substance which has no spatial location.

PHILOSOPHICAL BEHAVIOURISM

Philosophical behaviourism is a family of theories that claim that we *can* analyse mental concepts in terms of concepts that relate to the body, and in particular,

[1] Lacewing, M. (2017) 'Mental causation'

the concept of ‘behaviour’. Philosophical behaviourism claims that *what we are talking about* when we are talking about the mind and mental states is behaviour – what people do and how they react. On this view, the mind is not a ‘thing’. Rather, we can talk about organisms ‘having minds’, or better, having mental states, on the basis of how they behave.

In *The Concept of Mind*, Gilbert Ryle argues that to talk of mental states and processes is to talk not only of actual behaviour, but also of ‘dispositions’ to behave in certain ways. A disposition, in its simplest form, is simply how something will or is likely to behave under certain circumstances. For instance, sugar is soluble. Solubility is the disposition to dissolve when placed in water. When we talk of someone having a certain mental state, like being proud or believing that the earth is round, we are talking of what they would do, could do, or are liable to do, in particular situations or under particular conditions, including conditions that they are not in at the moment.

On Ryle’s analysis, dispositions are not causes. A disposition isn’t something that brings something else about. A cause is something that occurs, something expressed in a categorical statement (how things *are*), rather than referring to a disposition. So ‘He made lunch because *he was hungry*’ shouldn’t be understood along the lines of ‘the glass broke because *a stone hit it*’, but along the lines of ‘the glass broke when the stone hit it because *it was brittle*’. Hunger and brittleness are both dispositions; a stone hitting glass is an event. So when we explain an action by referring to someone’s mental state, such as hunger, we aren’t referring to a non-physical cause, we are situating the action in relation to a number of hypothetical statements. Making lunch is just the kind of thing someone who is hungry would do, in the right circumstances.

On this understanding, there is no mental causation. Mental states aren’t causes. This isn’t to adopt epiphenomenalism. Rather, mental states aren’t the *right kind of thing* to be causes (or fail to be causes). Talk of mental causation is a category mistake. Instead, we explain how people behave in relation to the dispositions they have.

We may object that this is not right, because there is mental causation. First, while dispositions may not be causes in the same sense as particular events, we can argue that they are part of the ‘causal story’. For example, the stone won’t break the glass if the glass *isn’t* brittle. So dispositions make a contribution to causal chains, and citing a disposition can be a causal explanation. The same is true of beliefs, desires and other mental states, we may claim. When I say ‘I went to the party because I thought you would be there’, we are citing something that is causally relevant to my action.

A famous argument from Donald Davidson in ‘Actions, Reasons, and Causes’ supports this. Suppose I have two reasons to do something, but I only act on one of those reasons. For example, I want to see you and believe you’ll be at the party, and I believe the party will be fun and I want to have fun. Suppose I go to the party because I want to see you, not because I want to have fun. How can this be true? What makes it the case that the first reason, and not the second, is the reason I act on? ‘Central to the relation between a reason and an action it explains

is the idea that the agent performed the action *because* he had the reason'. This 'because' must be a causal 'because'. What makes it true that I act on the first reason, and not the second, is that the first reason causes my action. So behaviourism is wrong to think that there is no mental causation.

Second, there are mental occurrences that may operate as causes like particular events do. For instance, in working through a problem, one thought may 'lead to' the next and this 'leading to' should be understood causally. The whole mental process is a causal chain, with each stage causing the next stage.

TYPE IDENTITY THEORY

Type identity theory claims to solve the problem of mental causation. All mental properties are identical with brain properties. Mental occurrences are identical with neurons firing. Mental states that involve behavioural dispositions are neurological connections. And so all mental causation just becomes a form of physical causation. For my desire for food to cause my searching for food is just for certain physical properties of my brain to cause that behaviour. To say my decision to watch TV is the cause of my picking up the remote control is just to say that some particular event in my brain with certain physical properties is the cause of my picking up the remote control. Mental states and processes cause actions because they are physical states and processes.

NON-REDUCTIVE PHYSICALISM AND FUNCTIONALISM

Type identity theory is a reductive form of physicalism. Non-reductive forms of physicalism, such as physicalist forms of functionalism, claim that mental properties are 'dependent but distinct'. Mental properties are not physical properties, but they supervene upon physical properties. According to physicalism, physical properties 'fix' *all* the other properties in such a way that it is *not possible* for the other properties to change without changing the physical properties. For example, the functional properties of an object depend on its physical properties, e.g. whether an eye can see (whether it can convert light waves into neural signals) depends on its physical properties.

Suppose that I make a decision to get some tea, stand up and walk to the kitchen. Did my decision cause my actions? According to causal role functionalism, the decision will need to be realised by some inner state (or states), which neuroscience indicates are states of my brain. By playing a particular function, that brain state or event has the property of being my decision. We might say, therefore, that if the state or event in my brain causes my action, then my decision causes my actions.

(There are other forms of functionalism that understand functions non-causally, in a very similar way to how Ryle understands dispositions, for example Daniel Dennett's 'teleological functionalism' in *The Intentional Stance*. On this view, whether someone has a particular functional property is *just* a matter of whether certain hypothetical statements are true about them or not, and whether we can explain and predict their behaviour using such statements. These forms of

functionalism do not usually make any claims about 'inner' states that correspond to the functions we identify with our talk of mental states.)

However, philosophers have objected that this is too quick and simple. Mental causation needs more. In saying that my decision caused my actions, we want to be able to say that it was because the event was my decision - a mental property – that it caused my actions, not because it was some brain event – a physical property. In other words, we can ask 'was it because of its physical properties or because of its mental properties that the event caused its effects?'. Do I go make tea because I want tea and decided to get some, or do I go make tea because certain neurones fire in my brain? How can we defend the common-sense view that my desire and the decision are causally relevant?

Causal closure

The 'causal closure of the physical', a commitment of physicalism, is the view that every physical event has a sufficient physical cause. This entails that the physical properties of the mental event must be sufficient to bring about its effects. If type identity theory is true, mental events can be causes without violating physicalism, since mental properties just are physical properties. But if non-reductive physicalism is true, e.g. if mental properties are functional properties and these are distinct from physical properties, this is less obvious.

Why believe in the causal closure of the physical? Because if we do not, we suppose that some physical events have no complete physical explanation. We could argue that events that need a non-physical explanation are miracles, from the point of view of physics. Furthermore, wherever science has so far found the cause of a physical event, it has found a sufficient physical cause. Denying the causal closure principle commits us to saying that science cannot and will not find the causes of certain physical events (in the brain, most likely).

The options

The problem of mental causation, as it is faced by non-reductive physicalism, brings us to four options:

1. Epiphenomenalism: mental properties have no causal effects. This is very counter-intuitive.
2. Overdetermination: both the physical properties and the mental properties of a mental event are each sufficient to bring about the effect. I decide to go to the kitchen to make tea. We can't say that if the neurones in my brain didn't fire, the decision to do so would have been enough to cause my walking. To walk I need muscular contractions, and these in turn require nervous stimuli from the brain. The mental properties, if relevant at all, aren't sufficient – they rely on the physical properties occurring as well. But, according to physicalism, the physical properties are sufficient on their own. So given the primacy of the physical properties, it seems that it is not in virtue of being mental that these states and events cause other events. This takes us back to epiphenomenalism.
3. Solving the two challenges above.
4. Abandoning causal closure, and therefore, rejecting physicalism.

In the next section, we discuss these last two options.

NON-REDUCTIVE PHYSICALISM AND PROPERTY DUALISM

Many non-reductive physicalists have analysed mental causation in terms of counterfactuals. Let's vary the example. Suppose, on having a headache, I reach for the aspirin. Is it the pain of the headache that causes this movement, or the neurons firing? To say that it was because the headache hurt that I reached for the aspirin implies that if the headache hadn't hurt, I wouldn't have acted as I did. If we want to say that it was just the neurones firing that caused me to reach for the aspirin, we have to say that if the neurones had fired, but the headache hadn't hurt, I would still have reached for the aspirin. But this seems false. Can we suppose that the neurones would have fired anyway, without the headache hurting? If the neurones firing are one particular physical realization of the pain of a headache, how could they have fired without the headache hurting?

There are a wide variety of theories about the supervenience relationship between the pain and the neurones, but they all attempt to secure the idea that without the pain, I wouldn't have done what I did. For example, functionalism argues that pain can be realized by many different physical states, so mental and physical properties can't be the same. However, mental states, on any particular occasion, are realized by physical states. So the causal role of the mental state is, on any particular occasion, filled by a physical state. The physical properties fill the causal role of pain on this occasion. In this way, we can say both that it was the pain that caused me to reach for the aspirin and that it was the physical properties.

But non-reductive physicalists face this problem: according to physicalism, the physical event of reaching for the aspirin has a sufficient physical cause, so the firing of the neurones on its own must be sufficient. If it is sufficient on its own, then the pain of the headache is not necessary. Now, of course, if the neurones firing is a realization of the pain, then the pain will always accompany the firing; but it is still unclear whether the pain plays any part in bringing about my reaching for the aspirin, or whether it is epiphenomenal. If we get a complete causal explanation without mentioning the pain, then the fact that the physical properties realize pain appears to be irrelevant.

Some property dualists argue that to secure the causal relevance of mental properties, we must reject physicalism. On this view, we should say that the neurones firing is sufficient for me to reach for the aspirin only given the presence of a headache, i.e. there must be a mental cause present as well. So the neurones firing is not sufficient on its own. Furthermore, it is not true that if the neurones hadn't fired, I wouldn't have reached for the aspirin, because for me to have a headache, some other neurones would have fired. So these particular neurones firing is not necessary, as long as some neurones fired in a suitable way. Since the physical properties of the event are neither necessary nor sufficient to explain the effect, they are not a complete physical cause.

On this account, our causal explanation must mention the mental property, the pain. It is true that the pain isn't sufficient, because some neurones must fire for any effect to follow. However, the pain is necessary, and this secures its causal

relevance. The conclusion is that physicalism is false, because the physical cause of my reaching for the aspirin is not sufficient. This account of mental causation is emergentism, and a form of property dualism, because mental properties have irreducible causal powers.