Eliminative materialism[1]

What are you? I mean, what is it to be you? What kind of thing is a human being? What makes a person a person? People have given surprisingly different answers to these questions. You might think, in light of evolutionary theory, that the answer is that we are animals. But sometimes, when someone is an irrecoverable coma or brain-dead, we say that they no longer exist, that they've 'gone'. But the body lying there is still the same animal. So our minds seem particularly important to who or what we are. Without a mind, I am not a person at all and I'm not 'me'.

But what is the mind? Should we talk about 'the mind' at all? A central question in metaphysics of mind is 'is the mind a substance?' Can your mind exist on its own, independently, or is it dependent on something else in order to exist? In particular, is your mind dependent on your body, perhaps especially your brain, in order to exist at all? If 'the mind' isn't a thing, then perhaps it is better to talk about 'mental properties' – properties of thought and consciousness. We can then ask about the relation between these properties and the brain.

Eliminative materialism (also known as eliminativism) argues that future scientific developments, especially in neuroscience, will show that the way we think and talk about the mind is fundamentally flawed, at least in some very important respects. At least some of our mental concepts are so mistaken that they refer to things that neuroscience will show don't exist. Central to our normal understanding of the mind are phenomenal properties and Intentionality. Eliminative materialism argues that neuroscience will revolutionise our understanding of each so that we may question whether they exist at all as we think of them now. As neuroscience proceeds, it will show that at least some of our central psychological concepts don't refer to anything – nothing exists that corresponds to some mental terms, e.g. 'belief', 'desire' or 'pain' or even 'Intentionality' and 'consciousness'.

PATRICIA CHURCHLAND ON REDUCTION AND ELIMINATION

Reductive explanation

In the simplest terms, an 'ontological reduction' involves the claim that the things in one domain (e.g. mental things) are identical with some of the things in another domain (e.g. physical things). For example, we can argue that heat is just mean molecular kinetic energy. They are the same thing. Or again, although they seem different, electricity and magnetism are the same force, electromagnetism. A similar claim could be made concerning mental properties: 'Every mental property is a certain physical property'. The identity claim is a reduction because we have 'reduced' mental properties – which we might have thought were a different kind

[1] This handout is based on material from Lacewing, M. (2017) *Philosophy for A Level: Metaphysics of God and Metaphysics of Mind* (London: Routledge), Ch. 3, pp. 223-32

of thing – to physical properties. I.e. there is *nothing more* to mental properties than being a certain kind of physical property. (This claim is discussed in the handout 'Mind-brain type identity theory'.) One reason to make such a claim is Ockham's razor. If mental properties are physical properties, then fewer things exist – the metaphysics is simpler.

But how does science ever come to make such a claim? In *Brainwise*, Patricia Churchland argues that what is involved is not just a simpler metaphysics (the claim that fewer things exist) but a more powerful explanatory theory. Inference to the best explanation goes beyond Ockham's razor. If identifying two properties enables you to explain something that you can't otherwise explain, that is the best reason for thinking they are the same thing. Otherwise, we can't move beyond the claim that they are merely correlated.

Ontological reduction is part of reductive causal explanation:

> a reduction has been achieved when the causal powers of the macrophenomenon are explained as a function of the physical structure and causal powers of the microphenomenon. That is, the macro-properties are discovered to be the entirely natural outcome of the nature of the elements at the microlevel, together with their dynamics and interactions.

For example, we can explain everything about water – why it is liquid at certain temperatures, why it is transparent, why cars skid on it, why we can't breathe in it but fish can, etc. – in terms of the nature of molecules of H_2O, how they are structured and how they interact with each other and other things (such as car tyres or our lungs). To 'reduce' water to H_2O is just to be able to explain all the causal powers of water – the effects it has on other things and the effects other things have on it – in terms of the causal powers of H_2O molecules.

The identity claim doesn't mean that the *concepts* of the macro-theory mean the same as those referring to the micro-properties. WATER doesn't mean H_2O, and THOUGHT doesn't mean 'neurophysiological firing pattern x'. However, when one theory offers a reductive explanation of things in another theory, it often happens that the meanings of the concepts *change* in light of new empirical discoveries. For example, the term ATOM meant 'indivisible fundamental particle', but then physicists became able to split the atom. So the meaning of 'atom' changed to mean 'the smallest existing part of an element consisting of a dense nucleus of protons and neutrons surrounded by moving electrons'.

<u>Beyond reduction to elimination: the example of heat</u>
However, sometimes the empirical discoveries indicate that rather than changing the meaning of the concept, we should give up on that concept and what it refers to completely. In other words, the concept should be eliminated because nothing exists in the way it supposes.

A good example is given by the history of the science of heat. What do you think heat is, just from everyday experience? Well, hot things have more of it than cold things. Heat passes from hot things to cold things. Hot things 'give off' heat. So how about this suggestion: heat is a kind of fluid that makes things hot and can be

passed from one thing to another. This was the theory of heat in the late 18th century, and the fluid was called 'caloric' (as in 'calories').

Ok, so if hot things have more caloric fluid than cold things, they should weigh more. So, when you heat something up, it should get heavier. Scientists tested this. Heating something up doesn't increase its weight. Ok, so maybe caloric is a fluid that doesn't have any weight? A rather unusual physical substance…

Here's another puzzle: you can make two cold things hotter by rubbing them together, i.e. friction generates heat. How? Where does the caloric fluid come from? Well, perhaps caloric can be trapped between atoms and rubbing something releases its caloric, so it is now hot? Ok, but if that's true, then there will only be a finite amount of caloric fluid trapped between the atoms, so eventually it will run out and the thing you rub will no longer get hot. Scientists tested this. It's not true – friction never stops generating heat. So is caloric fluid not only weightless but also infinite? A very, very unusual physical substance…

In 1798, Benjamin Thompson, Count Rumford, suggested a different theory: heat is the motion of micro-particles (molecules, atoms). Over time, with other scientific developments on the movement of molecules and atoms, this became accepted. Heat is the kinetic energy of molecules that can be passed from one thing to another.

The theory of caloric fluid turned out be very mistaken. Heat isn't a kind of fluid at all, but something quite different. So we shouldn't say that actually, we have reduced caloric fluid to kinetic molecular energy, just changing the meaning of CALORIC FLUID along the way. Instead, we have eliminated caloric fluid – there is no such thing – and explained the phenomena of heat in different terms. Churchland says, 'the nonexistence of something [e.g. caloric fluid] is established as highly probably… through the acceptance of an explanatorily powerful framework that has no place for it'.

<u>Complexity: Genes and mental states</u>

The reduction of heat to kinetic energy and the elimination of caloric fluid was fairly straightforward. It is unlikely that finding a neuroscientific explanation of thought or consciousness will be anything like as simple. It is worth bearing in mind, then, that scientific reductions can be very messy. A reductive explanation doesn't have to identify one macrolevel thing with one microlevel thing to succeed.

Genes provide a good example. Genes are the fundamental 'units of heredity' that give rise to the observable characteristics of living things. Biologists talked about genes before they knew about DNA. But now we are all told that our genes are 'in' our DNA. However, a gene is not necessarily single stretch of DNA (although genes are often misleadingly thought of this way). What we think of as a 'single' gene, relating to a characteristic that is inherited from one generation to the next, can involve many distinct segments of DNA (called 'exons'). It even turns out that the same DNA segment can contribute to different observable characteristics, depending on the stage of development and the environment of the cell in which the DNA is located. Should we say that a gene, or part of a gene, 'for' one trait

can simultaneously be the gene 'for' a different trait or that the same strand of DNA is part of two different genes? Despite all this complexity, biologists accept that DNA, its structure and how it interacts with other things, provides a *reductive* explanation of genes. We can trace a line of cause and effect from DNA sequences to bodily traits, and no one thinks of genes are something *in addition* to DNA.

Understanding reductive explanations in science helps us understand what reductive explanations of mental properties in terms of neurological properties may involve. They may be very messy and complex. There may not be just *one* physical property that we can identify with a particular mental property, but this doesn't mean we can't reduce the mental property. The important point is that we can explain mental properties, such as Intentionality and consciousness, in terms of physical properties.

ELIMINATION AND MENTAL PROPERTIES

Unlike genes, but like caloric fluid, not all mental properties may survive the process of reductive explanation, claims Patricia Churchland. The way we think about the mind now may be completely changed as neuroscience progresses. But we can't predict how. Reduction can threaten elimination.

Importantly for eliminative materialism, we aren't going to get reductive explanations of the mind just working from our everyday psychological concepts of 'belief', 'desire', 'emotion', and so on. These concepts are part of a theory about human behaviour (more on this below). Compared to the workings of the brain, human behaviour occurs over long periods of time (seconds, minutes, days), involves huge complexity and a far wider range of things in space, including other people. A theory of how the mind works can't reduce a theory of human behaviour to the very fast and tiny processes of neuroscience. An intermediate theory will be needed, e.g. how people process information, what happens when one desire conflicts with another, what processes are involved in a single emotion, how does imagination work, etc.

A good part of this intermediate theory will be developed by cognitive science. Before we can reduce mental properties to neurophysiological properties, we need a much better scientific theory of how the mind works. This will develop *side-by-side* with neuroscience. Only after cognitive psychology and neuroscience have 'co-evolved' will reductive explanations be possible. By this point, we can expect that our usual categories for thinking about how the mind works – beliefs, desires, emotions – will have changed and neuroscientific reduction will change them further. We have already rejected many psychological theories from the past, and we can expect this to continue.

For example, it turns out that some people are more easily addicted to substances (food, alcohol, smoking, drugs) than others. We might, common-sensically, say that they have less 'will-power'. But it turns out that they have different dopamine systems (dopamine is a neurochemical that relates to motivation and a sense of pleasure or 'reward' when you get what you want). So now what should we think about 'will-power'? What is it? Is there really anything such thing?

PAUL CHURCHLAND ON WHY 'FOLK PSYCHOLOGY' MIGHT BE FALSE

We mentioned above that we have a common-sense theory about why people behave as they do. For example, if someone is thirsty, they will – under normal conditions – look for something to drink. If someone believes it is raining outside, and doesn't want to get wet, they will – under normal conditions – pick up an umbrella or other covering to keep them dry. And so on. With claims like these, we are able to understand, explain and sometimes predict each other's behaviour very successfully. We do this by referring to each other's beliefs, desires, emotions, intentions and so on. Call this body of knowledge 'folk psychology'.

According to Paul and Patricia Churchland, folk psychology is an empirical theory. As such, it may turn out false, and the central concepts that it uses may, like 'caloric fluid', turn out not to refer to anything that exists. So far, however, the claim has only been that this *may* happen. In 'Eliminative materialism and the propositional attitudes', Paul Churchland argues that there are three good reasons to think that it *will* happen.

1. There are many aspects of mental life that folk psychology cannot explain, such as mental illness, the nature of intelligence, sleep, perception and learning. Explanations of these phenomena will need concepts that folk psychology lacks.
2. If we look at the history of folk psychology, it reveals no progress since the ancient Greek authors, 2,500 years ago. By contrast, neuroscientific explanations are constantly growing in scope and power.
3. We cannot make folk psychology coherent with other successful scientific theories. In particular, the central idea of Intentionality is highly problematic.

This third objection requires some unpacking. Why think that Intentionality can't be reduced by other scientific theories, but has to be eliminated? The reason is that it is very puzzling how anything physical could have Intentionality. Intentionality names the property in virtue of which thoughts are directed onto objects or propositions. Thoughts are 'about' something, objects or events in the world. For example, I might have a belief *about Paris*, a desire *for chocolate*, be angry *at the government*, or intend *to go to the pub*. In all these cases, my state of mind is 'directed' towards an 'object', the thing I'm thinking about (Paris, chocolate, the government, going to the pub). However, it seems that physical things are never 'about' anything. A particular molecular structure or physical process, described in physical terms, is not 'about' anything. For example, digestion is a chemical process, in which acids in your stomach break down food. What is that process 'about', what does it represent? Nothing – the question itself is puzzling! But the states and processes of your brain are just chemical states and processes, just like the states and processes of your stomach. So how could brain processes or states ever be about anything? So how could Intentional mental states be states of your brain?

Churchland concludes that folk psychology, with its explanations in terms of Intentional mental states like beliefs and desires, does not fit in with empirically robust theories, such as neuroscience, and so we have reason to abandon it.