

The 'philosophical zombies' argument¹

This handout follows the handout on 'Property dualism'. You should read that handout first.

Property dualism is the view that, although there is just one kind of substance, physical substance, at least some mental properties are not physical properties (as type identity theory claims) nor functional properties (as functionalism claims), nor are they behavioural dispositions (as philosophical behaviourism claims). While mental properties are possessed by physical substances, they are a fundamentally different kind of property from physical properties.

Physicalism is the view that everything that exists - every substance, every property that substances have, every event that occurs - is either physical or completely depends upon something that is physical. 'Physical' means something that comes under the laws and investigations of physics, and whose essential properties are identified and described by physics. According to physicalism, the physical properties of the world determine all the properties of the world, not just causally but metaphysically. Property dualism rejects this. It claims that some mental properties are properties that are not determined by physical properties in the way that physicalism claims. They are something new, something in addition to physical properties.

POSSIBLE WORLDS

To understand the 'zombie' argument for property dualism, we first need to understand the idea of a possible world. And to do this, we need to understand physical, logical and metaphysical possibility.

Physical possibility: Call the world we live in, as it is in fact, the 'actual world'. This world has particular laws of nature, such as the law of gravity and $e = mc^2$, and physical constants, such as the speed of light. These laws and their application to physical objects define what is physically possible. For instance, it is not (physically) possible for human beings to fly unaided (on Earth), because the upward thrust they can generate using their bodies cannot exceed the force of gravity. What is physically possible is what is possible given the laws of nature as they are in the actual world.

Logical possibility: Logical possibility is easiest to understand by relating it to analytic and synthetic propositions. All meaningful synthetic propositions describe what is logically possible. True analytic propositions describe what is logically necessary (what must be the case). False analytic propositions describe what is

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. 299-308

logically impossible (what cannot be the case). For example, it is logically impossible for there to be a square with three sides. The phrase 'a square with three sides' is conceptually incoherent, i.e. the meanings of the terms contradict each other, and so no such thing can exist. Anything that is not logically impossible is logically possible (or logically necessary).

So, we can think of logical possibility as conceptual possibility - what our concepts allow as making sense. We can argue that this is the same as what is conceivable - what we can imagine without self-contradiction.

The laws of nature seem contingent, i.e. it seems possible that they could have been otherwise. Light could have gone faster or slower; the ratio of mass to energy could have been $e = mc$, and so on. Of course, these things aren't physically possible. But they are, it seems, logically possible. Nothing in the concept of light entails that it must travel at 299,792 kilometres per second. Or again, it isn't logically impossible that human beings can fly unaided, just physically impossible.

Everything that is physically possible is logically possible (unless our concepts are terribly muddled!), but not everything that is logically possible is physically possible.

Metaphysical possibility: Some philosophers want to stop there, with two types of possibility - physical and logical. But debates in metaphysics, including the metaphysics of mind, over the last 40 years have led many philosophers to argue that there is a third type of possibility, metaphysical possibility. The reason is that analytic truths and necessary truths may come apart.

For example, 'WATER' and 'H₂O' are different concepts, and before the discovery of hydrogen and oxygen, people knew about water. They had the concept of WATER, but not the concepts of HYDROGEN and OXYGEN, and so not the concept of H₂O. And so they didn't know that water is H₂O. Even after hydrogen and oxygen were discovered, someone may have thought 'I wonder whether water is made of hydrogen and oxygen or something else'. So 'water is H₂O' is not analytically true. On this understanding, it is conceivable, or logically possible, that water is not H₂O.

But water and H₂O are one and the same thing - the two concepts refer to just one thing in the world. Water is identical to H₂O. Now, nothing can be what it is not. So if the property of being water and the property of being H₂O are one and the same property, you can't have 'one' without 'the other'. If A is the same thing B, then A and B can't be separated - there is just one thing here. So while we have two concepts - WATER and H₂O - there is only one property that they both pick out in the world.

What this is means is that, although it is logically possible for water not to be H₂O, it is metaphysically impossible for water to be anything other than H₂O. It seems that not everything that is logically possible is metaphysically possible.

Why don't we just say that it is physically impossible for water to be anything other than H₂O? This claim is certainly true, but it isn't strong enough. If the laws

of nature are contingent, then perhaps they could be different. Light could still be light but travel at a different speed, couldn't it? The claim with water and H₂O is stronger. Water wouldn't be water if it wasn't H₂O. If the laws of nature changed, so that hydrogen and oxygen never bonded and there was no such thing as H₂O, then there would be no such thing as water. There couldn't be water, but with a different chemical composition.

We now have a sense of what metaphysical possibility is, and how it is different from both physical possibility and logical possibility. What is metaphysically possible is constrained by the real nature or identity of things.

To summarise: what is physically possible is what is possible given the laws of nature as they are in the actual world; what is logically possible is whatever is not conceptually incoherent or self-contradictory; and we didn't give a definition of metaphysical possibility, but we said it was constrained by the real nature or identity of things. We discussed it in terms of necessary truths that are not analytic, such as 'water is H₂O'. We can understand metaphysical possibility better by talking about 'possible worlds'.

Let's start by talking about true and false propositions. Propositions describe 'states of affairs'. Propositions can be true or false. A proposition that is true describes the actual world, the way things are, a true state of affairs. A proposition that is false describes the way things are not, a false state of affairs. However, false propositions can be necessarily false or just contingently false. A proposition that is necessarily false cannot be true - it is impossible for it to be true (either logically or metaphysically). A proposition that is only contingently false describes a state of affairs that is possible, but false, given how the world actually is. For example, 'I was born in Kenya' is false, but could have been true.

A contingently false proposition describes a way things could be, if they were different. We can say that in some other 'possible world', a contingently false proposition is true, the state of affairs it describes is part of the way that world is. In some other possible world, I was born in Kenya. A possible world is a way of talking about how things could be.

Possible worlds are distinct from one another depending on what we are supposing to be true in that world. So the possible world in which I was born in Kenya is different from the possible world in which I was born in Argentina which is different from the possible world in which I don't exist at all.

These examples - of where I was born or even not existing - describe possible worlds that are physically possible as well as metaphysically and logically possible. There is nothing physically, metaphysically or logically impossible about the state of affairs of my being born in Kenya. We can imagine much bigger differences from the actual world without leaving physical possibility, e.g. a world in which the Earth never formed or in which evolution never gave rise to human beings.

But we can also talk about possible worlds that are physically impossible, worlds in which the laws of nature are different, e.g. in which light travels at a different speed, or that contain physically impossible things, perhaps things such as angels

and ghosts. If these are genuine possible worlds - ways that a world could be - then they are worlds which are physically impossible but metaphysically possible.

As we will see, philosophers disagree on which worlds are possible worlds. It is not always easy to tell. For example, is there a possible world in which water is not H₂O? In our previous discussion of metaphysical possibility, we said that 'water is H₂O' is not an analytic truth, so it is conceivable (logically possible) that water is not H₂O. But water and H₂O are identical - just one thing. So it is not metaphysically possible for water to exist without being H₂O. There is no possible world in which water exists, but is something other than H₂O.

We can say that what is metaphysically possible is what can exist or occur as part of a possible world. Metaphysical possibility is narrower than logical possibility, since it turns out that not all conceptually coherent propositions describe how things could exist. So physical possibility concerns how things can be given the actual laws of nature; metaphysical possibility concerns how things can be in any possible world; and logical possibility concerns whether a proposition is conceptually coherent.

CHALMERS' ZOMBIE ARGUMENT

Property dualism claims that phenomenal properties (which many property dualists claim are qualia) are not physical properties, nor do they supervene on physical properties. It rejects physicalism. But how can a property dualist show this? In 'Consciousness and its place in nature', David Chalmers uses the idea of possible worlds to make the argument.

According to physicalism, everything that exists is either physical or depends on what is physical. So if physicalism is true, a possible world that is an exact physical duplicate of our world (the actual world) will be an exact duplicate of our world *in all respects*. This is just the claim of supervenience, but at the level of the world. Consider: a painting that is an exact physical duplicate of another painting has all the same aesthetic properties as that painting. So a whole world that is an exact physical duplicate of another world also has all the same aesthetic properties. But what goes for aesthetic properties goes for all properties, according to physicalism. There can be no difference in, say, mental properties without a difference in physical properties. In other words, it is *metaphysically impossible*, says physicalism, for two worlds to have the same physical properties and different mental properties, because the physical properties determine the mental properties.

Therefore, if there is a possible world that is an exact physical duplicate of our world but is different in any way, e.g. it has different phenomenal properties, then physicalism is false. If two physically identical worlds have different properties of consciousness, those properties of consciousness don't depend on physical properties. This is what Chalmers tries to show with the idea of a philosophical zombie.

What is a philosophical zombie?

A 'zombie', in the philosophical sense, is an exact physical duplicate of a person - you, for instance - but without any conscious subjective quality of experience. It therefore has identical physical properties to you, but different mental properties - it has no phenomenal consciousness.

Of course, zombies are not possible in the actual world. They are not physically possible, i.e. given the laws of our universe, we have every reason to believe that any being that has identical physical properties to you will also have consciousness.

What we are thinking about when thinking about zombies is a different possible world - a world which has all the physical properties of our world but without consciousness. We are describing a world that may be metaphysically possible.

But is such a world really (metaphysically) possible? To argue that a world with zombies is possible is to argue for property dualism. How does the argument work?

The argument

First, it seems that zombies are at least conceivable. I've just described them, and there isn't an obvious contradiction in the idea. Second, given their conceivability, we may argue that zombies are therefore metaphysically possible. There is a possible world which has all the same physical properties as the actual world, but has no properties of consciousness.

Now, if consciousness were *identical* with physical properties, it would be impossible for a creature to have the same physical properties as you but not have consciousness. This is Leibniz's principle of the indiscernibility of identicals. As we saw with water and H₂O, if *A* is identical to *B* - if *A* is *B* - then you can't have *A* without *B* or vice versa; they are the same thing. So if zombies are possible - if a creature could be physically identical to you but not have consciousness - then consciousness is *not* identical to any physical properties. So, if zombies are metaphysically possible, then consciousness is not identical to any physical properties. Furthermore, if zombies are metaphysically possible, consciousness doesn't supervene on physical properties either, because you and your zombie 'twin' have identical physical properties, but different phenomenal properties. And so property dualism is true: phenomenal properties are neither reducible to nor supervenient upon physical properties.

- P1. It is conceivable that there are zombies.
- P2. If it is conceivable that there are zombies, it is metaphysically possible that there are zombies.
- C1. Therefore, it is metaphysically possible that there are zombies.
- P3. If it is metaphysically possible that there are zombies, then phenomenal properties of consciousness are neither physical properties nor supervene on physical properties.
- C2. Therefore, phenomenal properties of consciousness are neither physical properties nor supervene on physical properties.
- C3. Therefore, physicalism is false and property dualism is true.