Responses to the zombie argument[1]

This handout follows the handout on 'The 'philosophical zombies' argument'. You should read that handout first.

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. According to the zombie argument,

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.

In this handout, we consider three possible responses physicalists may make to the argument. First, they may argue that what is being proposed – in this case, a possible world that contains zombies – is not conceivable (P1 is false). Second, they may argue that although zombies are conceivable, they are not metaphysically possible (P2 – and therefore C1 – is false). Third, we may argue that even though zombies are metaphysically possible, this doesn't tell us what consciousness is, and its relation to physical properties, in the actual world (P3 is false).

A PHILOSOPHICAL ZOMBIE (OR ZOMBIE WORLD) IS NOT CONCEIVABLE

The first premise of the zombie argument claims that we can conceive of beings that have the same physical properties as us but without consciousness. Why think this is conceivable? Because when we think of physical properties, this doesn't determine what we must think of consciousness. By contrast, when we think of the answer to 3 × 4, we must – if we are thinking clearly – think of 12. It is inconceivable that 3 × 4 is anything other than 12. Or to use an example from Descartes, it is inconceivable that the internal angles of a triangle could add up to anything other than 180 degrees. By contrast, it does not seem inconceivable that there could be a being with identical physical properties to you, but without consciousness.

[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. 309-19

The first objection to the argument is that, despite appearances, zombies are not conceivable. If we think they are conceivable, we are not thinking clearly or we lack some relevant information. It is difficult to recognise that we are not thinking clearly. But we can spell out where we are going wrong in more detail.

First, if physicalism is true, we should note that something's physical properties determine its functional properties. So a physical duplicate of you is also a functional duplicate of you. (If physicalism is not true, then something's functional properties could depend on its non-physical properties as well. But we cannot assume that physicalism is false, since that is what the zombie argument is trying to prove. To assume physicalism is false is to beg the question.)

Second, we need to revisit the arguments that phenomenal consciousness can be analysed in terms of physical and functional properties; there are no qualia. If we are not persuaded by this claim, it is probably because our analysis of consciousness is still underdeveloped. But if we had a complete analysis, we would see that consciousness can be completely explained in these terms. In that case, a physical, functional duplicate of you would also have consciousness.

So, once we are clear on a being's physical properties, we can, in principle, deduce how it functions, and from this, with a complete analysis of consciousness, we can deduce whether or not it is conscious. So to imagine a being with identical physical properties to you but without consciousness is confused. It is like accepting the premises of a deductive argument but rejecting the conclusion. In conceiving of a 'zombie' as having identical physical properties, you conceive of it as having identical functions. But to function in certain (highly complex) ways just is to be conscious. So zombies – physically identical, but non-conscious beings – are inconceivable. (As Descartes might put it, the ideas of a zombie and of consciousness are not clear and distinct. When we make them clear and distinct, we see the contradiction in thinking that zombies are possible.)

P1. A zombie is a physical duplicate of a person with phenomenal consciousness, but without phenomenal consciousness.
P2. (If physicalism is true,) A physical duplicate is a functional duplicate.
C1. Therefore, a zombie is a physical and functional duplicate of a person, but without phenomenal consciousness.
P3. (If physicalism is true,) Phenomenal properties are physical properties realising particular functional roles.
C2. Therefore, a physical and functional duplicate of a person with phenomenal consciousness has phenomenal consciousness.
P4. A physical and functional duplicate of a person with consciousness cannot both have and lack phenomenal consciousness.
C3. Therefore, (if physicalism is true,) zombies are inconceivable.

This objection to the zombie argument depends on there being a complete physical and functional analysis of consciousness. I have inserted the phrase 'if physicalism is true' into two premises, (P2) and (P3), which the property dualist will contest. If there is no such analysis, because an analysis of consciousness in terms of its physical and functional properties doesn't provide an analysis of what

it is like to experience something, then it seems that this response to the zombie argument fails.

Should we accept (P2) and (P3)? To do so, we need to good reasons to think that phenomenal properties are can be understood or explained either in terms of physical structure or in terms of functions. But there is nothing in our phenomenal concept and experience of consciousness that supports the claim. And so we can conceive of that same physical thing either with or without phenomenal consciousness.

The debate looks like a stalemate. On the one hand, the zombie argument mustn't assume that physicalism is false, since it is trying to show that physicalism is false. On the other hand, the response seems to assume that physicalism can give a complete analysis of our concept of consciousness.

Many philosophers have concluded that we should grant that zombies are conceivable, and focus the discussion on whether they are metaphysically possible. That takes us to our second response.

WHAT IS CONCEIVABLE MAY NOT BE METAPHYSICALLY POSSIBLE

The second response targets the second premise of the zombie argument. Although zombies are conceivable, they aren't in fact metaphysically possible. What we are able to conceive is not always a reliable guide to what is possible.

Identity and metaphysical possibility

To understand this, let us return for once again the example of water and H_2O. As we saw, the two concepts WATER and H_2O are distinct, and it is not an analytic truth that water is H_2O. So it is conceivable (even if false) that water is not H_2O.

Given this, it is easy to think that water could have been different, i.e. in some possible world, water is not H_2O. However, given that water is H_2O, it's not metaphysically possible that water isn't H_2O. This was an important claim about identity first made by Saul Kripke in *Naming and Necessity*. It's not possible for A to be B and for it not to be B. So if A is identical to B – if A is B – then A is B in every possible world. Because water is H_2O, it is H_2O in every possible world.

It is possible that the water in the oceans could have been fresh, not salty. Or in other words, in another possible world, the water in the oceans is fresh, not salty. The fact that oceans are salty is a contingent property of water in our world. It isn't what makes water what it is. Or again, the fact that water falls as rain is a contingent property of water. If it never rained, this wouldn't change what water is. So in another possible world, water never falls as rain.

But turn now to the question of what makes water what it is? What is the essential property of water? The answer: its chemical composition, H_2O. Now, what makes water what it is is not a property that water can lack in some possible world. A world without H_2O is a world without water, because water just is H_2O.

Suppose there is another possible world in which a transparent, odourless liquid falls as rain, fills the oceans, freezes and evaporates, etc. but isn't H_2O. Is this liquid water? No, says Kripke. It is something just like water, in that it has many of the contingent properties of water. But it isn't water, because it isn't H_2O.

Kripke concluded that identity claims – 'A is identical to B' – are necessarily true, if true at all. They are true in all possible worlds.

We said that we can conceive of water not being H_2O. But we have argued that it isn't possible that water is not H_2O. This shows that we cannot always infer metaphysical possibility from conceivability.

The response to the zombie argument

We can now apply the point to zombies. The fact that we can conceive of zombies doesn't show that zombies are metaphysically possible. If phenomenal properties just are certain physical and/or functional properties, then it isn't possible for zombies to exist (even if they are conceivable). Given the physical properties we have, if physicalism is true, it just isn't possible for a being with the same physical properties not to have consciousness as well. If physicalism is true, then when we think of phenomenal consciousness and, say, certain neurological or functional properties, we are thinking of one and the same property in two different ways, using two different concepts.

This response doesn't have to claim that phenomenal properties are physical properties, that physicalism is true. It only has to claim that the zombie argument cannot show that physicalism is false. The premise that zombies are metaphysically possible cannot be defended without assuming that phenomenal properties are not, unknown to us, physical properties.

A disanalogy?

This second objection to the zombie argument relies on an analogy between phenomenal consciousness and scientific identities, such as water and H_2O or life and chemical processes. Property dualists can argue that this analogy doesn't work.

Something isn't water if it isn't H_2O, because H_2O is the 'essence' of water. The concept WATER is a concept of something that has a particular structure and causal role, which science can then discover. Water is precisely the kind of thing that could be – and is! – identical with a chemical property. This is why you can't have water without H_2O or H_2O without water.

By contrast, say property dualists, the essence of phenomenal properties is what it is like to experience them. The essence of pain – what makes pain pain – is how pain feels. Its essence isn't some physical or functional property. The essence of a physical property is its physical structure or composition; the essence of a functional property is what causes it and what it causes. In arguing that neuroscience can tell us what consciousness 'really is', physicalists are assuming that the essence of consciousness, like the essence of water, is something physical. But this is a mistake. Consciousness is essentially first-personal, i.e. what it is like for the person. The concept of consciousness is not the concept of

something that has a particular physical structure or set of causal relations that science can then discover. So consciousness is not essentially a set of brain properties described by the neuroscientist.

If this is correct, then the correlation between brain properties and consciousness in the actual world is contingent. As it happens, certain brain processes give rise to consciousness. But you could have the brain processes without consciousness and vice-versa. It is only the essential properties of something that can't change in different possible worlds, the contingent properties can. The same physical processes that are correlated with consciousness in the actual world may not be correlated with consciousness in another possible world. Because phenomenal properties have a different essence from physical and functional properties, each can exist without the other. So zombies are possible.

WHAT IS METAPHYSICALLY POSSIBLE TELLS US NOTHING ABOUT THE ACTUAL WORLD

A third objection to the zombie argument targets the inference from the claim that zombies are possible to the conclusion that property dualism is true. The zombie argument shows, at best, that in another possible world, physical properties and phenomenal properties are distinct. But why does this entail in the actual world that they are distinct? Couldn't it be the case that physicalism is true in the actual world, but property dualism is true in a different possible world? Or in other words, the zombie argument only shows that property dualism is possible; it doesn't show that property dualism is true.

We can reply that this objection makes two mistakes. First, the objection misunderstands identity. It suggests that phenomenal properties could be physical properties in this world but not in another possible world. But this isn't possible. Nothing can be something else. I can't not be me in another possible world. If 'I' were not me, but you, say, then that person is not me. In any possible world, the only person I can be is me. Likewise, water can't be something other than water. Since water is H_2O, it can't be something else in another possible world.

The same goes for phenomenal properties. If phenomenal properties are physical properties in this world, then they are physical properties in every possible world. And if they are not physical properties in another possible world, then they are not physical properties in any possible world, including the actual world. When it comes to identity, possibility does tell us about reality.

Second, if the objection is intended to defend physicalism, it misunderstands what physicalism claims. Physicalism claims that what exists is either physical or supervenes upon what is physical. We need to be clear about supervenience. For example, we want to say that if the physical properties of two paintings are identical, then the aesthetic properties cannot be different. It is not strong enough to simply say that they aren't different, since that would allow that the physical properties don't 'fix' the aesthetic properties.

What we said about the aesthetic properties applies to properties of consciousness as well, and what applies to paintings is true of whole worlds. According to

physicalism, once the physical properties of a world are finalised, then there is no further work to be done to 'add' consciousness. It is already part of the world. Phenomenal properties cannot differ independently of physical properties. So physicalism is a claim about what is metaphysically possible.

The zombie argument attacks this claim. It argues that there can be two worlds that are physically identical but with different phenomenal properties. Once the physical properties of a world are finalised, then there is still further work to be done to 'add' consciousness. Thinking about possibility does, in this case, tell us about reality.

PATRICIA CHURCHLAND ON THOUGHT EXPERIMENTS

In *Brainwise*, Patricia Churchland is sceptical about the use of appealing to conceivability or metaphysical possibility to discover the nature of the world. In imagining a zombie, we are imagining a being with a brain just like ours. But in imagining that it doesn't have phenomenal consciousness, we are imagining that if we knew everything about neuroscience, we still wouldn't have explained or understood consciousness. But all this imagining is really a reflection of our own epistemic limitations and the fact that neuroscience just isn't very developed yet. The thought experiment of zombies don't tell us anything significant about the nature of consciousness. Property dualists are mistaken in trying to get conclusions about what exists out of epistemic premises.

Suppose someone (perhaps 200 years ago) said 'I just can't imagine how living things could really be composed of dead molecules – how can life arise out of the interactions of things that are not alive?'. Or again, suppose someone proposed the thought experiment of 'deadbies', creatures who are physically identical to us, but aren't alive. They claim to be able to imagine such creatures. None of this would persuade us to think again about vitalism and the existence of a special, non-reducible 'life force'. On the current biological theory of what it is to be alive, deadbies are impossible and life really is just the highly complex interactions of molecules.

Similarly, we shouldn't be persuaded by the property dualists' appeal to zombies. First, from what is conceivable we cannot infer anything about the nature of how things are. Our grammar – our concepts as they are now – are not necessarily a good guide to how things really are. We change our concepts as we discover more about the world. Second, the same goes for thought experiments about what is metaphysically possible. What things really are is what they are in the actual world. We discovered what water is through scientific investigation. Similarly, the right way to think about consciousness is through scientific investigation, and we shouldn't let considerations about concepts determine in advance what scientific investigation may or may not discover. For example, contemporary biology argues that genes are DNA. Should we object this reductive explanation because in another possible world, genes – understood as the units of heredity – might not be DNA? No. What genes are is what genes are in the actual world. And all we need for to make this claim is an empirical identity, supported by scientific explanation. The same goes for consciousness.

Churchland is arguing that philosophy simply can't do metaphysics in this way, using thought experiments and possible worlds to discover what something is. Philosophical 'speculation' must give way to experimental science.