

The knowledge argument

PROPERTY DUALISM

Property dualism is the view that, although there is just one kind of substance, physical substance, there are two fundamentally different kinds of property, mental and physical. Mental properties are possessed by physical substances; but at least some mental properties do not depend on physical properties in the way physicalism claims. The form of property dualism that is most discussed defends property dualism for phenomenal properties of consciousness. Property dualism claims that these properties, such as pain, the smell of coffee, the visual experience of a red rose, the feeling of joy, and so on, can't be reduced to physical, behavioural or functional properties. These properties, at least, are a completely new, irreducible type of property.

Property dualism rejects physicalism. First, it argues that the properties identified by physics do not form the *complete* fundamental nature of the universe, because in addition, there are properties of consciousness. Physics misses something fundamental. When all the physical properties of the world are finalized, there is still work to be done - properties of consciousness have not been fixed. Property dualists are happy to allow that there may be correlations, even natural (though not physical) laws that connect particular physical and mental properties. But, they argue, mental properties are nevertheless distinct - an entirely new kind of property in the world. Second, some property dualists argue that these mental properties have their own causal powers, which can affect physical events. This threatens physicalism's claim that non-physical causes do not contribute to the way the physical world changes over time.

Chalmers distinguishes between what he calls the 'easy' and the 'hard' problems of consciousness. The 'easy' problem involves analyzing and explaining the functions of consciousness, e.g. the facts that we can consciously control our behaviour, report on our mental states, and focus our attention. Chalmers thinks that understanding how the brain works will eventually provide the solutions. So this doesn't threaten physicalism. The 'hard' problem relates to the phenomenal properties of consciousness, what it is like to undergo conscious experiences. How and why are certain physical processes in the brain associated with such experiences?

The physicalist says that these conscious experiences *just are* certain physical processes or certain physical states playing a particular functional role. But, Chalmers argues, a physical account of something can only explain its physical structure and function - how something is constituted and how it works. And this, he objects, is not enough to explain phenomenal consciousness. Such explanations miss out how experiences 'feel', what it is like to undergo them, their subjective or first-personal aspect. There is more to phenomenal consciousness than structure and function.

Frank Jackson defends property dualism on the basis of his 'knowledge argument'. He describes the following scenario. Suppose there is a neuroscientist, Mary, who has lived all her life in a room in which everything is black and white. She has never seen any colour other than black, white and shades of grey. However, she has specialized in the science of vision, and through textbooks and black-and-white TV, she has come to know every physical fact there is to know about colour vision - everything about the properties of light, everything about the eye, everything about the nerves and the brain related to vision. So, Mary knows all the physical information there is to know about what happens when we see a ripe tomato. She is then let out of the black-and-white room, and comes to see something red for the first time. Does she learn something new?

Jackson claims that 'it seems just obvious' that she will. She will learn about what it is like to see the colour red. And so she learns something new about our visual experience of the world. However, we said that she knew all the physical facts while she was in the room. So not all the facts are physical facts. It is possible to know all about the physical properties of the brain involved in having an experience and yet not know about the qualia.

1. Mary knows all the physical facts about seeing colours before being released from her black-and-white room.
2. On being released, she learns new facts about seeing colours.
3. Therefore, not all facts are physical facts.
4. Therefore, phenomenal properties are non-physical and physicalism is false.

By 'all the physical facts', Jackson means not only what we already know about physics and neurophysiology. Mary knows all the physical facts as discovered by a *completed* physics and neuroscience. Furthermore, she has worked out all the causal and functional facts that are entailed by these facts. Because physicalism claims that the world is entirely physical (if we include causal and functional properties), it must claim that to have complete physical knowledge is to have complete knowledge. But no amount of physical information can enable Mary to know what it is like to see a ripe tomato.

In 'What Mary Didn't Know', Jackson puts the argument in another form, which may help to clarify it. Let's generalise Mary's knowledge to everything physical.

1. Mary (before her release) knows everything physical there is to know about other people.
2. Mary (before her release) does not know everything there is to know about other people (because she *learns* something about them on her release).
3. Therefore, there are truths about other people (and herself) which escape the physicalist story.

We can add:

4. Therefore, phenomenal properties are non-physical and physicalism is false.

OBJECTIONS

Mary gains ability knowledge, not new propositional knowledge

A first objection points out that there is more than one meaning of 'to know', more than one kind of knowledge. We can and should accept that Mary gains new knowledge when she sees red for the first time. But this doesn't mean that she

gains knowledge of some new *fact*. We can argue that instead of gaining knowledge of a proposition (e.g. 'that red looks like this'), Mary gains *know-how* - the knowledge involved in certain abilities. For instance, to see red for the first time is to gain the ability to know how to imagine or recognize red.

Suppose that seeing red gives us these abilities. Are such abilities *all* that is involved in knowing what it is like to see red, Jackson asks in 'What Mary Didn't Know'? Suppose Mary wonders whether what it is like for others to see red is the same as what it is like for her. She isn't wondering about her abilities to imagine and recognize red. She is wondering about the truth of a proposition. So when Mary first learns what it is like to see red, she *does* gain knowledge of a new fact.

Is the objection even right to think that knowing what it is like to see red involves knowing how to imagine red? Suppose there is someone who (for whatever reason) has no ability to imagine seeing red. Now suppose this person looks attentively at something red. While they look at red, they know what it is like to see red. And yet they cannot imagine seeing red. This shows that the ability to imagine is not necessary for knowing what it is like to see red. Now suppose someone else has the most amazing ability to imagine seeing colours. They are told that there is a shade of red, e.g. burgundy, that is between plum red and tomato red. They are now *able* to imagine burgundy, but as long as they don't *actually* imagine burgundy, they still don't know what it is like to see burgundy. This shows that the ability to imagine a colour is not sufficient to know what it is like to see it. (We can make similar arguments for recognizing colours.)

If the ability to imagine seeing red is neither necessary nor sufficient for knowing what it is like to see red, then when Mary comes to know what it is like to see red, she learns more than simply knowing how to imagine seeing red. The objection fails to show that Mary does not learn a new fact. It fails to show that the knowledge argument is mistaken.

Mary gains acquaintance knowledge, not new propositional knowledge

A second objection argues that Mary gains a different kind of knowledge again, not propositional knowledge (knowing that), but not ability knowledge (knowing how) either. Instead, she gains 'acquaintance knowledge' - a direct awareness of the thing. To see red is a *direct* apprehension of red, as contrasted with descriptions of seeing red. How does the objection work?

Suppose that what it is like to see red is a physical property of the visual experience, which itself is a physical process. In other words, the phenomenal property of what it is like to see red is some property of the brain (type identity). Mary can then know all about this physical property, about what it is, when it occurs, and so, before she leaves the room. However, she is not *acquainted* with the property - she doesn't have direct knowledge of it because *her brain has never*

itself had this property. When she sees red, this property occurs in her brain and she becomes acquainted with it. She gains new knowledge, but she hasn't learned any new fact. She already knew all about this property before she left the room.

There are two possible responses to this. First, we can argue that acquaintance knowledge involves propositional knowledge. What it is to be acquainted with red is to know *that seeing red is like this* (having the experience). Becoming acquainted with red involves learning some new fact. So Mary does learn a new, and therefore non-physical, fact when she becomes acquainted with red. So what it is like to experience red can't simply be a physical property of the brain.

Jackson gives a different response in 'What Mary Didn't Know'. The objection misunderstands the argument. He agrees that, of course, Mary doesn't have acquaintance knowledge of what it is like to see red. As the objection claims, to have such acquaintance knowledge, she would need to have direct experience of seeing red. But the knowledge argument isn't about *Mary's* experience. The argument is that Mary didn't know everything about *other people's* experiences before she left the room, even though she knew everything physical about their experiences. Mary doesn't know what it is like for anyone to experience red. This is a *fact* about experiences that Mary doesn't know. When Mary leaves the room, she realises how impoverished her conception of people's colour experiences has been. So there are facts about other people's experiences of seeing red that Mary learns.

There is more than one way of knowing the same physical fact

A third objection to Jackson's argument distinguishes between two ways we might talk about 'facts'. Suppose I know that there is water in the glass. Is that the same as knowing that there is H₂O in the glass? No - because someone may know one of these truths without knowing the other. Someone can have the concept of 'water' without having the concept of 'H₂O'. Or again, someone may have both concepts, but not know that water and H₂O are the same thing. So we can say that to know that there is water in the glass and to know that there is H₂O in the glass is to know two different facts. In this sense of 'fact', we count facts in terms of concepts.

However, in another sense of 'fact', the fact that there is water in the glass *just is* the fact that there is H₂O in the glass. Both of these claims are made true by just one state of affairs in the world. In this sense of 'fact', we count facts in terms of how the world is, not how we think about it.

We can now apply this to the knowledge argument. Before leaving the room, Mary has a concept of red in physical terms - wavelengths of light, neurons firing, and so on. Call this the 'physical' or again a 'theoretical' concept of red. We can contrast this with a 'phenomenal' concept of red. A phenomenal concept of something is the concept by which you recognize something when you experience or perceive it. So we gain the phenomenal concept of red by seeing red.

When Mary comes out the room and sees red, she acquires the phenomenal concept of red for the first time. She is now able to think about red in a new way, in terms of what it is like to see red. She couldn't know what it is like to see red before because she didn't have the phenomenal concept. But, we can claim, the phenomenal concept of red is a concept of *the same thing* that her physical concept is a concept of - they are two different concepts of a physical property

of the brain (like water and H₂O are two concepts of the same physical substance). Mary gains knowledge of a new fact in one sense (because she gains a new concept) but not in the other sense (since she already knew about the property).

Let us accept that the knowledge argument shows that there are different ways of thinking about physical things, some of which depend on experiencing, rather than describing. To know what it is like to see red, you need to have the phenomenal concept of red, and this you can only gain from experience. So Mary gains knowledge of a new fact, in the sense of fact that relates to concepts.

However, and this is the objection, physicalism and property dualism are claims about what exists. They are claims about *properties*, not about concepts. The knowledge argument does *not* show that Mary gains knowledge of a new property. It doesn't show that Mary gains knowledge of a new fact in the sense of learning about something in the world she didn't know about before. It doesn't show that what it is like to see red cannot be a physical property. So the argument fails to show that there are any non-physical properties. So it fails to show that physicalism is false.

All physical knowledge would include knowledge of qualia

The objections above all accept the claim that Mary learns something when first leaving her room. But a fourth objection denies this. If Mary really did know *everything* about seeing red, she would not learn anything when she first sees red. The experience of seeing red is nothing more than highly detailed knowledge of what it is to see red, and Mary already has this highly detailed knowledge.

This objection is, however, counter-intuitive. It requires that Mary is able to work out what it is like to experience a colour without ever having seen one. But, we might argue, we cannot describe such experiences (seeing red) so fully as to know what it is like to experience them without actually doing so. No one can know what it is like to see red without actually ever having seen something red.

In 'What Mary Didn't Know', Jackson points out that we can even allow that Mary could *imagine* what it is like to see red before she leaves her room. Imagining something is not the same as knowing it. You only have to imagine what something is like if you *don't already know*. If Mary knows all the physical facts, and these were all the facts there are, then she *would* know what it is like to see red, so she wouldn't have to imagine it. So even if Mary can imagine what it is like to see red, she still doesn't *know* what it is like.

Is this right, though? This is our intuition, but is there an argument to support it? We don't really know what knowing *all* the physical facts about seeing colours would involve. Perhaps Mary will be entirely unsurprised at seeing red - she already knew what it would be like. Is it impossible for any amount of information describing the experience to convey it? This objection claims that there is, in principle, a complete analysis of phenomenal properties in physical and functional terms.

JACKSON'S CHANGE OF MIND

In 1998, Jackson accepted this last objection and so concluded that the knowledge argument doesn't work after all.

His change of mind is as controversial as the original argument. Many philosophers prefer the solution that there is more than one way of knowing the same physical fact. That solution claims that it is impossible to work out what an experience is like from theoretical knowledge about it, because knowing what an experience is like requires a phenomenal concept, and that requires actually having the experience. But Jackson maintains that it is possible, from complete knowledge of the physical facts, to work out what experiences are like.

Jackson's argument is very brief.

1. Assume that interactionist dualism is false.
2. Therefore, what causes our conscious experiences is purely physical (either physicalism is true or epiphenomenalist dualism is true).
3. Therefore, when Mary comes to learn what it is like to see red, this process has a purely physical causal explanation.
4. Therefore, *what* Mary learns is also something physical - it can be understood and explained in purely physical (and functional) terms. We shouldn't think that what she learns is something that doesn't feature in the explanation of how she comes to learn it.
5. Therefore, we should reject epiphenomenalist dualism and accept physicalism.

An experience, Jackson claims, is simply highly complex functional information. It doesn't seem like this, which is why we *think* that what Mary learns is non-physical. Usually, to acquire complex functional information, e.g. about the effects of a certain drug, we have to investigate and bring information together from different sources, make inferences, do tests and so on. But sensory experience gives us this kind of information in a highly unusual way - remarkably quickly and easily. So it doesn't seem like functional information, but knowledge of some intrinsic property of the experience - qualia. However, appearances are misleading, and this is knowledge of a physical and functional property.

Having argued that, in principle, complete physical knowledge would give someone knowledge of what experiences are like, Jackson doubts whether it is possible for us to gain such complete physical knowledge. There may be too many things about the world that we will never be able to understand.

QUALIA (AS DEFINED) DON'T EXIST

This section follows the handout 'Functionalism: Objections'; you should read that handout first.

The knowledge argument claims that phenomenal properties can't be understood in terms of physical and functional properties. Instead, it understands phenomenal properties as qualia - intrinsic and non-representational properties of experience. The last objection - the claim that all physical knowledge would include knowledge of 'qualia' - rejects this understanding of phenomenal properties. There are no qualia. Phenomenal properties are, instead, just physical, functional properties.

To defend this view, we must find reasons to reject the arguments in favour of qualia. For example, is the idea of an inverted spectrum really coherent, once we factor in very detailed functions? Is the idea of a Chinese consciousness coherent? (Or perhaps - what's wrong with accepting that there will be a Chinese

consciousness?) When we think through carefully just how much knowledge Mary must have in order to know *everything* about colour experience, is it coherent to say that she won't know what it is like to see red before she leaves her room?

We have a good reason to reject the claim that phenomenal properties are qualia, namely Ockham's razor. Our explanation of the mind, and what exists, is simpler if we can explain phenomenal properties in terms of physical and functional properties. We need a really good reason to think that *everything else* in the world can be explained in physical and functional terms *except consciousness*. Are the arguments that qualia exist really strong enough?

Logical behaviourism and qualia

We can press the objection another way. Do the arguments that qualia exist even understand consciousness correctly? Ryle and other logical behaviourists argue that the concept of qualia misunderstands our talk of sensations, feelings, images, and so on. These are not each a 'something' that has peculiar properties of 'what it is like'. The whole metaphysical picture here is wrong. When we express our experiences, we use words that derive their meaning from describing physical objects. To say 'what it is like' to see red is simply to describe what we see when attending to the colour of a red object, or if it is not in front of us, we give a report of our memory of seeing it. The redness that we experience is the redness *of the rose*, not a property of our experience of it.

People don't normally talk about 'sensations' or 'what it is like' in the sense of qualia in everyday language, before being exposed to some theory. If you ask someone 'what it is like' to see a rose, they will usually respond evaluatively, e.g. 'it's wonderful' or 'it's calming'. Of course, experiences differ from each other. But this isn't because what each experience 'is like' differs. We can express the difference in terms of what the experience *is of*, and how we evaluate it, e.g. whether we enjoy it or find it boring. The property dualist has misunderstood our mental concepts.

What would a logical behaviourist say about Mary? Perhaps this. In knowing all the physical facts, Mary can't yet understand our normal way of talking about experiences. She has no experiences of coloured objects that she can express and report, and as a result, she has only a limited understanding of our discussions of them. But none of this has to do with knowledge of facts, either facts about some 'inner' conscious experience or facts about the brain. To think otherwise is a category mistake.