Functionalism and qualia[1]

Functionalism is the view that every mental state consists of a disposition to behave in particular ways and to have certain other mental states, given certain inputs from the senses and certain other mental states. In other words, we can give an analysis of what mental states are in terms of their 'inputs' and 'outputs'. The inputs are inputs from the senses and other mental states; the outputs are behaviour and other mental states. Most functionalists understand the relations between inputs, mental states and outputs causally. Any functional state can be described in terms of what typically causes it, and what it typically causes in turn. But some functionalists understand 'function' computationally, in terms of the software instructions for a machine. (For more on functionalism, see the handout 'Functionalism'.)

Objections to functionalism take the form of arguments that we cannot reduce mental properties to functional properties. Objections can focus on either of the two features of mental states, Intentionality or consciousness. In this handout, we look at two objections concerning consciousness.

THE CHALLENGE TO FUNCTIONALISM

Phenomenal properties are properties which give an experience its distinctive experiential quality, 'what it is like' to undergo that experience. According to some philosophers, phenomenal properties are intrinsic and non-Intentional. If they are, then they are 'qualia'. (For more on qualia, see the handout 'What do we mean by mind?'.)

The objection to functionalism is this: if phenomenal properties are qualia, then they cannot be completely understood in terms of their causal roles (or inputs and outputs on a machine table), because these are relational properties, not intrinsic properties. It is not what causes them and what they cause in turn that makes pain or the smell of coffee or the visual sensation of red what it is. What it is like to experience these mental states – how pain feels, how red looks, how coffee smells – can't be analysed in terms of functions. Yes, of course, how pain feels is important to what it causes, e.g. it causes you to cry out or withdraw your hand from the fire. But the feeling of the pain isn't just these causal relations. So functionalism can't explain phenomenal properties.

P1. Qualia, by definition, are intrinsic, non-Intentional properties of conscious mental states.
P2. Intrinsic, non-Intentional properties cannot, by definition, be completely analysed in terms of their causal roles (or as machine table states).

[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. 274-83

C1. Therefore, if qualia exist, some mental properties cannot be analysed in terms of their causal roles (or as machine table states).
P3. Functionalism claims that all mental properties are functional properties which can be completely analysed in terms of their causal roles (or as machine table states).
C2. Therefore, if qualia exist, functionalism is false.
P4. Qualia exist.
C3. Therefore, functionalism is false.

The controversial premise is (P4). In what follows, we will look at a thought experiment that tries to establish this premise.

Thought experiments are designed to test a hypothesis or philosophical claim through imagining a hypothetical situation, and coming to a judgment. In philosophy of mind, the most common kind of thought experiment is one in which we are asked to judge whether the hypothetical situation is possible or not. Thought experiments are used when actual experiments are either practically or physically impossible, or when the judgments concern matters that cannot be investigated by empirical means, e.g. the correct application of concepts or metaphysical questions of identity.

THE POSSIBILITY OF A FUNCTIONAL DUPLICATE WITH DIFFERENT QUALIA (INVERTED QUALIA)

We can show that phenomenal properties cannot be understood just in terms of their functions if we can show that it is possible for two people to have states with identical functions but different phenomenal properties. The most popular version of this objection is known as the case of 'inverted qualia'.

Suppose that you and I are looking at ripe tomatoes and fresh grass. Because we have grown up in the same linguistic community, we have learned to use the word 'red' to describe the tomatoes and 'green' to describe the grass. So we both say that the tomatoes are red, the grass is green. But the particular way that tomatoes seem to me is the way that grass looks to you, and vice versa. Functionally, we are identical, and yet we have different colour experiences. 'The way grass looks to you' and 'the way grass looks to me' are functionally identical; both are caused by the same inputs (grass) and cause the same outputs (e.g. saying 'grass is green'). But they are not identical in terms of their intrinsic properties. They refer to different qualia.

Of course, we might not know whether this is true or not. But that is irrelevant. The objection is that inverted qualia are possible. If functionalism were true, inverted qualia would be impossible. So functionalism is false.

PATRICIA CHURCHLAND ON INVERTED QUALIA

The functionalist can reply that in the case described, you and I are not, in fact, functionally identical. There are going to be small, but very important, differences, because the causal relations of phenomenal properties are very

complex. In *Brainwise*, Patricia Churchland argues that we have no good reason to think that qualia can be inverted in the way the thought experiment describes.

In presenting her argument, Churchland is defending eliminative materialism rather than functionalism. But because she argues that phenomenal properties are not intrinsic properties, her defence works for both theories.

She starts by making the objection from inverted qualia clearer. The main claim in the objection is that you and I – or our brains – could function in exactly the same way, but we would have different qualia. This is not being proposed as an empirical hypothesis, e.g. that you and I really do see red and green differently. Why not? Because empirical hypotheses are tested against the evidence. First, we have no evidence from neuroscience that identical brain functioning gives rise to different conscious experiences in different people. Second, as an empirical hypothesis, it is poor, since it proposes that there could be empirical differences (in our conscious experience) that are undetectable, since they make no functional difference. But science does not proceed by supposing undetectable facts! So if the inverted qualia objection were empirical, then it is either false or bad science.

So in saying that you and I could function in exactly the same way, but have different qualia, 'could' must mean 'it is conceivable'. So the argument is like other arguments from conceivability, and faces the same objections as other arguments from conceivability. We can question whether it is conceivable; whether, even if it is conceivable, it is possible; and whether its possibility tells us anything about whether there actually are qualia (i.e. whether phenomenal properties are intrinsic and non-Intentional).

Churchland then points out that the thought experiment is much too simple. First, every colour that we can discriminate has unique similarity and dissimilarity relations to all surrounding colours. For instance, red is more similar to orange than green is, while green is more similar to blue than red is. So we can't simply switch red and green without messing this up. If you and I saw red and green 'switched', then we wouldn't agree on whether red was more similar to orange or blue. And this is a functional difference.

One response would be to change the thought experiment – it is not just red and green that are inverted, but the whole spectrum. This could keep all the similarity relations as well. Someone who sees red as green also sees orange as blue, and so they say that red is similar to orange, but what they see is what I see when I look at green and blue.

But this meets another problem. Human beings can make much finer discriminations in green, yellow and orange than we can in blue. If we inverted everything, this would be apparent from behaviour, as whoever sees the inverted colours would be able to make finer discriminations among blue than the rest of us can, and fewer discriminations in green, yellow and orange. And so it is empirically impossible for someone to have inverted qualia without functional differences.

But isn't it still conceivable that we could correct for this as well? The person with inverted qualia has the same objective discrimination abilities as the rest of us,

even if subjectively, they are discriminating between blue qualia in ways that we can't. This is conceivable, but, asks Churchland, should we say that this person really sees the same colours as us? The colours that they see bear new similarity relations to every other colour, and they have different powers of discrimination. What is essential to a colour being the colour it is?

The qualia theorist argues that phenomenal properties are intrinsic – they are essentially what it is like to experience them. In other words, it is conscious introspection that identifies whether two colours are the same or not. But is this right?

We can offer an explanation of our experience of colour in terms of our physical constitution and how it functions. For example, why does colour have the three dimensions of red, green and blue? The answer has to do with types of colour-sensitive cells, 'cones', in our retinas and the way they are wired up to the brain. If we say that our experience of qualia is the way the brain processes this information, we can explain our colour experience – similarities and dissimilarities, what colours we can see and what we can't etc. But if qualia are something distinct, we have no explanation for why we see colour as we do. What colour experience is should be decided by who has the best explanatory theory of colour experience, and not by thought experiments.

BLOCK ON THE POSSIBILITY OF A FUNCTIONAL DUPLICATE WITH NO QUALIA

The second objection to a functional analysis of consciousness tries to show that phenomenal properties cannot be understood just in terms of their functions because it is possible for two systems to have states with identical functions but one system (you, say) has phenomenal properties, but the other does not. The most popular version of this objection is Ned Block's China thought experiment.

In 'Troubles with functionalism', Block frames his argument in terms of machine-table functionalism. Suppose we have a complete functional description of your mental states. For each and every one of your mental states, we have an input–output analysis, giving us a machine table for your mind. Now imagine that a human body, like yours, is connected up via its sensory and motor nerves not to a brain but, through electronic transmitters, to the whole population of China. The Chinese are linked up to each other by two-way radios, and some of these are linked up to the input and output nerves of the body. (Block picks China because the population of China is 1 billion, which may be enough to fulfil the functions that comprise your mental states). Then, for a short time, the Chinese population realizes the same machine table that describes the functions of your mental states.

According to functionalism, this should create a mind; but even if we could accept that this set-up could have Intentional mental states such as beliefs or thoughts, it is especially difficult to believe that there would be a 'Chinese consciousness'. If the Chinese system replicated the functioning of my brain when I feel pain, would something be in pain? What? Is there something it is like to be this system? The objection is that the Chinese system, although it duplicates your functioning, can't

duplicate your mind, because some mental states are qualia, and the system can't have qualia because they are not functional states.

Functionalists can reply that the Chinese system won't be functionally identical to you. For instance, it could be disrupted by things that your mind isn't disrupted by, e.g. the radios running out of batteries or the system being disrupted by bad weather.

True, but irrelevant, says Block. First, although this *could* happen, if it doesn't, then we have functional duplication, and the functionalist must say that the Chinese system is conscious. Second, these disruptions don't count as inputs or outputs, any more than having a brain tumour counts as an 'input' to our mental states. It is not part of their functioning – that's why they are disruptions.

Functionalists can object that the Chinese system is much slower than our brains. But, replies Block, why should this matter for whether it has mental states? Couldn't there be much slower minds than ours? In any case, this is just an objection about what is physically possible. A Chinese system that operated as fast as our brains is still metaphysically possible.

A PHYSICALIST RESPONSE

If Block's objection works, then not *everything* about the mind can be explained in terms of functions. But perhaps we can combine functionalism and the type identity theory to argue as follows.

If the Chinese system can have Intentional mental states, then functionalism provides an accurate account of all mental states except for consciousness of phenomenal properties, which involves qualia. Why should this be? We could argue that the intrinsic properties of qualia depend on the specific *physical* properties of the system that realises the functional states. We saw this type of argument in Patricia Churchland's response to the objection from inverted qualia above. How colours look to us isn't (just) a matter of what causes the colour experience and what effects it has, it also depends on our physiology – the types of cones we have and the way our brains are wired. So what mental states something has depend on its functional properties *and* its intrinsic physical properties. Mental states are still nothing more than physical states playing a functional role. A physical, functional duplicate of a person with consciousness will have the same conscious states.