David Chalmers on Mind and Consciousness
Richard Brown
Forthcoming in Andrew Bailey (ed) *Philosophy of Mind: The Key Thinkers*.
Continuum Press

David Chalmers is perhaps best known for his argument against physicalism in the philosophy of mind (Chalmers 1996). But this is not his only contribution. He is a highly systematic philosopher who has offered important theories and insights in many areas of philosophy.

In what follows we will separate Chalmers' views into two broad projects. On the one hand we will look at what we might call the negative project where he aims to show that physicalism about consciousness cannot work. On the other hand we will look at what we can call his positive project, which consists in giving a theoretical account of consciousness and mind that is consistent with his anti-physicalism.

I. The Case against Physicalism

Chalmers starts with his distinction between what he calls the easy and hard problems of consciousness. The easy problems of consciousness all involve those things that straightforwardly involve functioning. This includes such things as discriminating, categorizing and reacting to the environment, the integration of information, reporting on our mental states, accessing our own internal states, focusing attention, controlling behavior in a deliberate way, and discovering the neural basis of sleep and wakefulness (Chalmers 2010 p 4). The hard problem of consciousness involves explaining why it is that any of that stuff gives rise to consciousness.

The sense of consciousness invoked in the hard problem is that of phenomenal consciousness or there being something that it is like for one to undergo various mental processes and instantiating various mental states. Thus the hard problem is that of explaining why it is that any of the functioning mentioned above feels the way it does from the inside. In particular it seems that any of the functioning could have occurred in the absence of any conscious experience whatsoever. So take, for instance, my seeing a red tomato while hearing someone say 'that's a tomato'. It is one thing to explain how it is that light reflected from the tomato affects my eye and how that produces activity in the brain, which results in various utterances and the grouping of that physical object with fire trucks and oxygenated blood. But none of that seems to explain what it is like for me to see red, or why there should be anything that it is like for me to see red in the first place. So too, it is one thing to explain how sound waves lead to certain kinds of brain activity but that doesn't seem to explain how I consciously hear the words

or understand their meaning. Why does any of that information processing happen consciously? Why isn't it all done in the dark?

One intuitive response at this point is to insist that at some point in the future we will be able to see how the hard problem is not really all that hard after all. Sometimes people use examples like life itself. It was once thought that we could not explain life in purely physical terms and that we would need to posit some mysterious non-physical essence to account for the difference between living and non-living things. We now think this isn't true, and that we can give a physical account of what it means to be alive. So, too, perhaps, the same will be true for consciousness in 1,000 years. But this response is not promising according to Chalmers (2010 p. 16). In the case of life it does in fact seem that what needs explaining are functional ideas. What is so striking about consciousness, from his perspective, is that it is totally unique in that it seems to be something more than function. When one has a conscious experience of pain, or seeing blue, or listening to jazz, or thinking about the philosophy of mind, one is in various states that differ from each other in what it is like for one to have them. Explaining why this is the case doesn't seem to be a matter of merely explaining functions. This is what makes the hard problem so hard.

So far this is just a puzzle. Given that we know we have conscious experience how do we explain how it arises from structure and function? It is not yet an argument. But an argument immediately presents itself. Following Chalmers (Chalmers 2010 p 106) we can call this the explanatory argument.

1. Physical accounts explain at most structure and function
2. Explaining structure and function does not suffice to explain consciousness
3. No physical account can explain consciousness

This argument captures the case against physicalism in its essence. It is intuitively plausible that if no physical account can explain consciousness then there is more in the world than what our physical sciences tell us. But the argument leaves much to be desired. One could insist that explaining structure and function is enough to explain consciousness, thereby denying that there is a hard problem of consciousness, or one could insist that failure to explain does not mean that there is more to our world than the physical.

Chalmers argues that when one looks at the various a priori arguments against physicalism one can see a pattern emerge (2010 p107, 193–203). These arguments start from an epistemic or conceptual gap between consciousness and physical properties and move to there being an ontological gap. That is exactly what was going on in the above explanatory argument. An explanatory gap, as Joe Levine (1983) calls it, leads us to infer that there is more in the world that what is merely physical.

To take another example, we can look at the Knowledge Argument advanced originally by Frank Jackson (1986). We imagine a super scientist who has been kept in a black and white room but who nonetheless has access to all of the physical theory about the world. Mary, as she is called, has never seen colors until one day she is released and sees a red ripe apple. She says 'ah! *that's* what it is like for one to see red!' Given that Mary cannot know what it is like to see red in her black and white room (that is, that there is an epistemic gap), but yet she knows all of the facts that can be captured by structure and function, the argument concludes that knowing what it is like to see red is knowing something beyond the physical facts. We cannot deduce one from the other.

Or, finally, take Kripke's (1980) well-known modal argument. In its basic form the modal argument goes as follows. Identity statements are necessarily true, if true at all. Those who think otherwise, like the early identity theorists (see chapter 6) are actually confused. When we think we can imagine lightning, say, without electrical discharge, what we really imagine is something that superficially presents the same appearance as electrical discharge (i.e. lightning) does to us, but which is not lightning. It is essentially fool's lightning, or something which resembles our lightning but is not. So if mind/brain identity statements were true then it would be impossible to have one without the other. But in the case of, say, 'pain is identical to some neural activity' it seems easy to imagine the pain without the neural activity and the neural activity without the pain. But if so then how can we explain the fact that the identity is necessary? It seems we cannot appeal to the previous strategy since it doesn't seem to make sense to say that there is something that presents itself in the same way that our pain does but is not really pain.

Though the arguments against physicalism all have a similar structure the one that is most well worked out, and which makes all of the issues maximally clear, is the two-dimensional conceivability argument (2010 p 142), which has the following form:

1. P&~Q is conceivable
2. If P&~Q is conceivable then it is metaphysically possible
3. If P&~Q is metaphysically possible then physicalism is false

The 'P' above is a placeholder that can be filled in with whatever physical theory you like. It is meant to be a total description of the world in physical terms. The 'Q' is a claim about phenomenal consciousness, like that I see blue or that someone feels pain. Each premise needs to be clarified and defended.

Premise one says that we can conceive of all of the physical facts holding without any consciousness at all. This is what Chalmers calls a zombie world. This premise relies on what Chalmers calls *ideal negative* conceivability. The notion of ideal conceivability is meant to capture what an ideal reasoner would be able to conceive under ideal conditions as opposed to what merely seems conceivable

under certain impoverished reasoning conditions to a less than ideal reasoner. Something is negatively conceivable just when it cannot be ruled out on a priori grounds. Something is positively conceivable when we are able to form some positive conception of the thing in question. It goes beyond merely not noticing any contradictions. In addition we are able to form some conception of how the thing in question could be true. Given this premise one says that it is ideally negatively conceivable that there be a zombie world. This in turn means that this world is not ruled out a priori or that there is nothing contradictory that follows from this description of the world. A common way to make the point is to say that there are no 'conceptual hooks' that allow us to move from talking about physical things to talking about phenomenal consciousness.

Premise two makes the claim that the right kind of conceivability is a reliable guide to what is metaphysically possible. These terms get used in many different ways and Chalmers is careful to distinguish various kinds of conceivability and the corresponding notions of possibility. We have already seen that he is interested, in the most part, in ideal negative conceivability but we must introduce the two-dimensional framework to make the rest of the distinctions.

The fundamental idea of two-dimensional semantics is that there are two different aspects of the meaning of statements, which roughly captures something like a Fregean distinction between meaning and reference. Chalmers calls these primary and secondary intensions. These intensions can be thought of as the contents of statements, where what it means for something to be a content is roughly that it divides up the space of possible worlds in a particular way. So corresponding to the two kinds of intensions will be two different ways of carving up the space of possible worlds, which he calls primary and secondary possibility.

It is perhaps easiest to start with the notion of possibility. We can think of the space of possible worlds as containing every coherent description of the way things could be. We can, if we like, metaphorically think of it as knowing all the ways that God could make the world if He so chose. Among that vast set of worlds will be one that describes the world that we actually live in, the real world. Once we know which one that is we then think of all of the other worlds as 'counter-factual' worlds. That is, we think of them as describing what could have been the case. But while it is the case that one of those descriptions corresponds or captures the way the world actually is we can in principle see that any of those descriptions could be the actual world and so we can then think about this space in two different ways. One way of thinking about these worlds is, so to speak, from their point of view. To do this we think 'what if this world were the actual world? What would be true then?' Another way of thinking about these worlds is, again so to speak, from our point of view. To do this we ask 'given that we know that our world is the actual world, and that as a result this, that and the other facts hold true, what could have been the case?' The first is the notion of primary possibility, the second the notion of secondary possibility.

Now we can introduce the primary and secondary intensions of a statement. Let's take as our statement the old standby 'water is H_2O'. The primary intension of the statement is a function from a description of a possible world, considered as the actual world, to a truth-value. The secondary intension of the statement is a function from a description of a possible world, considered as a counter-factual world, to a truth-value. To make this concrete consider the possible world made famous from debates about meaning and reference of terms like 'water' (Putnam 1973). This is the famous Twin Earth, which is just like Earth except that there water is not H_2O but is rather some other chemical substance, dubbed XYZ. XYZ acts and looks in every way like water, and the people on Twin Earth even call it 'water'! Twin Earth is strange but nothing about it is contradictory. It seems entirely coherent that our world could be that way. But given that our world contains in fact H_2O it doesn't seem like it could have been the case that it was otherwise.

So, 'water is H_2O' is true when we consider any possible world as counter factual. It has a secondary intension that is necessary, which we can call '2-necessity. This is because 'water is H_2O' is an a posteriori necessity, as Kripke pointed out. There are no worlds, considered as counter-factual, where water isn't H_2O. What this means is that when we consider Twin Earth as a counter-factual world it turns out that there is no water on Twin Earth and that it is still true that water is H_2O. This is because it is 2-necessary that water is H_2O. This is exactly what Kripke argued. On Twin Earth there is only fool's water. It is something that looks the way H_2O does to us but since it is not H_2O it is not water.

But even so, Twin Earth can be coherently described. If we consider Twin Earth as actual instead of counter factual, 'water is H_2O' comes out false. On Twin Earth 'water is XYZ' is true and so, when we consider Twin Earth as actual, 'water is H_2O' is false. That is to say that if Twin Earth were the actual world 'water is not H_2O' would be true (because 'water is XYZ' is true there). Whether or not our world could have been one where water was not H_2O, Twin Earth is possible in some sense.

Now, Chalmers continues, the zombie argument relies only on primary conceivability, not secondary conceivability. We are to think of the zombie world as if that world were the actual world. But if the zombie world were the actual world then consciousness is not physical, since everything physical is there but without consciousness. This can be put a bit more technically by saying that if the zombie world were the actual world then any proposed physicalist theory of consciousness would be false. But since we can (ideally and negatively) conceive of a zombie world it is a way our world could be. So, the way our world is physically is not enough for consciousness. Which is just to say that the thesis of physicalism is false.

II. Categorizing the Responses

How should one respond to this argument? Chalmers categorizes the responses according to the way in which one reacts to the conceivability argument.

One way to react is to deny the first premise. Those who take this route Chalmers calls type-A physicalists. According to the type-A camp we can see now that zombies are inconceivable. Or to put it another way they deny that there is an epistemic gap or a hard problem of consciousness. Chalmers sees many physicalists' views falling into this category. He cites analytic functionalism of the kind David Lewis held (see chapter 7) and eliminativism, which includes theories like those of Ryle (see chapter 5) and Dennett (see chapter 10), as examples (Chalmers 2010 p. 111). According to analytic functionalism it is a priori that mental states are connected to functioning and so anything that was functionally like us would have consciousness. According to the eliminativist consciousness doesn't exist in the way that gives rise to the hard problem.

The problem with the type-A view, for Chalmers, is that it seems not to take the data seriously. This marks a fundamental divide in the philosophy of mind. There are those who take it as a sort of starting point that there is more to explain than functioning. There is that to explain, of course; those problems make up the so-called easy problems of consciousness. But certainly it is the case we are conscious and it seems to be a further question as to why any of the functioning is done consciously.

A second way to react is to deny the second premise, which Chalmers labels the type-B response. The type-B camp accepts that there is an epistemic gap but then goes on to deny that this amounts to an ontological gap, thereby denying the link between conceivability and possibility. Many philosophers have defended the type-B approach. It is widely accepted that conceivability is in general not a good guide to what is metaphysically possible and this is precisely the type-B strategy. On these kinds of views it is metaphysically necessary that pain, say, is identical to a certain kind of physical state even though it is conceivable that you have that physical state without it being a pain state (this is the zombie world).

Putting this into the two-dimensional framework the type-B response is that 'consciousness' has a necessary primary and secondary intension, which Chalmers calls a 'strong necessity'. It has a necessary primary intension because there are no possible worlds that falsify it. So, take the identity between pain and some brain state or other. According to the type-B camp we can imagine a world with that brain state and no pain but there is no corresponding metaphysically possible world. Since the space of metaphysically possible worlds does not have one where the identity is false it is necessarily true, or true in all possible worlds. Chalmers goes on to argue that these strong necessities are deeply strange and inelegant.

Ultimately the dispute here is a local instance of the more general dispute about rationalism and empiricism. In particular we can see the dispute as an instance of the general debate about the principle of sufficient reason. Roughly put, this principle states that every positive fact must have an explanation. Thus if one accepts that it is a positive fact that pain, say, is identical to some physical state then we should be able to give an explanation for why that identity claim holds true. On Chalmers' view identities are in principle knowable a priori in the sense that an ideal reasoner who knew the relevant facts could come to know that the identity is true. So, in the case of water and H_2O, if an ideal reasoner knew all of the facts about H_2O and about the way that 'water' is used, they would be in a position to know that water was H_2O. But the type-B response is to deny that this is true for 'consciousness', which would seem to make it very different from other concepts like 'water' and 'gold'.

There are a couple of ways that type-B folks have responded. One way has been championed by Ned Block (Block & Stalnaker 1999, Block 2007, Block forthcoming), a well-known type-B physicalist. Identities in general, and between mind and brain in particular, are brute facts about the world. Identities, on his view, do not get explained. Rather, they get stipulated in order to license greater explanatory power. In the case of water and H_2O, stipulating that they are identical allows us to explain the way water behaves in terms of the way H_2O behaves. We can, for instance, explain why water freezes when it does in terms of the way H_2O behaves. But even this identity is not knowable a priori on Block's account.

Another type-B strategy is appeal to the special nature of phenomenal concepts to try to explain why there is a hard problem of consciousness but not of water in a way that is consistent with physicalism (Balog 2012). This has come to be known as the 'phenomenal concepts strategy'. Very roughly put, the idea is that we know about our own conscious experience in a unique way. Echoing Russell (1912) they say that we are acquainted with our own experience whereas we know everything else in a secondary kind of way. Some spell this out in terms of appeals to indexicals like 'I' and 'now' or demonstratives like 'this' and 'that'. Others argue that the phenomenal experiences actually constitute part of our beliefs about them.

As we will see later in this chapter Chalmers holds a version of this kind of view as well, but he argues against the physicalist appealing to it by developing what he calls his Master argument (Chalmers 2010 p 312). The basic idea behind this mater argument is that any kind of explanation that is going to be given by someone who wants to invoke the phenomenal concept strategy has to be tested by conceivability. In particular the claim is that we need to know whether it is conceivable that we have a physical duplicate that has the feature in question or not but that lacks consciousness. If this is conceivable then the concept is not explainable in physical terms and so is a form of dualism, or it is not conceivable

in which case we have not succeeded in explaining the actual relationship that we have with our conscious experience. The reason for this second claim is roughly that if it isn't conceivable then that means that zombies could have this property, but zombies are stipulated not to share our epistemic situation (that is, they are not conscious).

In general there are only two alternatives, on Chalmers' view. We either start off with a conception of consciousness that builds in special epistemic relations, like acquaintance, or we don't. If we do then we have the problems from conceivability arguments all over again. If we don't then we haven't captured the way our consciousness is to us. In the sections that follow we will look at the way that Chalmers develops his account of acquaintance.

The type-C response holds that there is a prima facie epistemic gap but that this gap will eventually be closed. With respect to zombies the claim is that they are prima facie conceivable, or conceivable given what we know now, but that they are not ideally conceivable. Paul Churchland (see chapter 12) is often cited as a type-C physicalist. Thomas Nagel (1974) has used the analogy of a contemporary physicist trying to explain to Socrates $e=mc^2$. Socrates just doesn't have the concepts to understand it. So, too, Nagel suggests, we may be like Socrates with respect to consciousness. Chalmers responds to this move by constructing a dilemma. Either we will discover more structure and function or we will expand science to go beyond structure and function. If we do the first then it seems we haven't answered the argument. We can still ask why *that* structure and function result in consciousness. If we take the second then it looks like we have admitted that dualism of some sort is true. So, let's look at the various dualist responses.

The type-D response is the traditional interactive substance dualism familiar from Descartes (See chapter 2). It is often claimed that this kind of dualism is at odds with science but Chalmers argues that this is not the case (Chalmers 2010 p 126–130).

Type-E responses are the epiphenomenal property dualist responses. Epiphenomenalism is the view that physical states cause or produce mental effects, but these mental effects are themselves unable to cause anything in the physical world. Pain, on this view, is a non-physical property of the brain, which is produced by the workings of the brain, but which has no effect on the way the brain functions. Epiphenomenalism has the theoretical cost of denying that conscious experiences are casually involved in the production of action. This is a severe theoretical cost but it is not a knock down argument against the view.

Type-F monism is the view that there are phenomenal or at least protophenomenal properties that underlie physical properties like mass and charge. This is a version of panpsychism. One way of getting to this kind of theory is by way of the zombie argument. We have so far been assuming that the

primary and secondary intension for terms like ‘mass,’ ‘charge’, and other terms that appear in physical theory, are the same. But it is possible that they come apart. This is often called ‘Russellian Monism’ since Russell (1927) suggested it at one point. This is the view that science as we know it only describes the relational properties of reality. Mass, for instance, is defined in terms of its causes and effects. We are not told what it is that has mass, or what the fundamental nature of mass is. If this is so, then it may be the case that the fundamental stuff is consciousness.

Chalmers has remained in principle neutral on these dualist positions, and claims that any of them could turn out to be true; but he seems most attracted to type-F views. This is because it seems to have the best of all worlds. It preserves the spirit of physicalism in that the fundamental phenomenal properties can be thought of as an extension of fundamental posits of physical theory. Zombie worlds are conceivable, because those worlds do not have the fundamental natures of our world—they lack consciousness. But we can also say that consciousness is causally efficacious. If it is the fundamental base of mass and charge then it has a fundamental role to play in the causal structure of our world.[1]

Another way to respond is by denying the whole apparatus that sets these arguments up, which Chalmers labels the type-Q response. In the spirit of Quine this response denies that there is any sense in modal talk (for a defense of this view see Mandik & Weisberg 2008). One way to defend this view is by developing the claim that what is conceivable depends on what theories one (tacitly) holds. If this were the case then finding zombies conceivable might be evidence that one (tacitly) holds a dualist theory or some kind of identity theory, rather than showing us what is possible.

Another challenge that cuts across these distinctions has come from the appeal to the conceivability of physical creatures that have consciousness, but no non-physical properties (Brown 2010, Frankish 2007). Brown has called these creatures ‘shombies,’ Frankish calls them ‘anti-zombies’. Assuming we are committed to the two-dimensional framework, it cannot be the case that both zombies and shombies are ideally conceivable. This could be used to defend a type-A position, or a type-B position, or a type-Q position, depending on how one proceeds. In its most general form the claim is simply that it is conceivable that consciousness be physical. Even if we have no clue how it could be physical there does not seem to be anything contradictory in the idea. Chalmers himself has expressed sympathy with the claim that shombies are at least prima facie negatively conceivable (Chalmers 2010 p 180), though he denies that this is enough to ground a premise in an anti-dualist argument.

---

[1] There are also a couple of other responses that he lays out. One could accept causal overdetermination, which he labels the type-O response, or one could be an idealist that he labels the type-I response.

We can now turn to examining Chalmers' positive project.

III. The Science of Consciousness

It would be a mistake to conclude from the foregoing discussion that Chalmers is not optimistic about the chances for a science of consciousness. He is very much in favor of a science of consciousness but, in his view, it must be one that transcends physical science as we know it now. In particular, as we have seen, he thinks it must include facts about consciousness as basic irreducible facets of reality. Doing this will broaden one's conception of what counts as science, rather than precluding a science of consciousness. So what does the science of consciousness look like according to Chalmers?

He begins by identifying two sources of data for a science of consciousness. The first is the third-person data that we derive in the pursuit of answering the easy questions of consciousness. The other is first-person data, which are the experiences that one has. The science of consciousness then consists in gathering both sorts of data and finding out the ways in which they are linked. Since Chalmers is convinced that first-person data is irreducible to third person data, the science of consciousness will be in the business of finding the fundamental laws which link these two sorts of data.

Chalmers has speculated about what some of these fundamental laws might look like. One of these he calls the 'principle of structural coherence,' the other 'the principle of organizational invariance'. The principle of structural coherence tells us that there is a lawful correlation between the structure of our experience and the structure of what he calls 'awareness,' which is one of the easy problems of consciousness. This principle boils down to the idea that we can see law-like regularities between brain activity and conscious experience. Chalmers can then happily accept that psychological theories give us important insights into the structure of conscious experience on the basis of facts about awareness, without positing a reduction.

The principle of organizational invariance tells us that it is the functional organization of the brain that matters for consciousness and mind rather than the specific material of the brain. This brings out the role that computation plays in the science of mind and consciousness for Chalmers. He has defended the claim that reality is in principle computable and that there are a set of computations that suffice for having a mind. This is because mind and consciousness are examples of what Chalmers has called 'organizational invariants'. Something is organizationally invariant when, roughly, a simulation of that thing counts as the real thing. So, being a hurricane doesn't count because a simulation of it is not the real thing. A very good example of something that is organizationally invariant is being a computer. A simulation of a computer is indeed a computer. His claim is that the consciousness and mind are also organizational invariants.

This underscores the central place of the notion of computation in Chalmers' thinking about consciousness and mind. A computation, for Chalmers, is specified relative to some formal system (Chalmers 2011). So, take the classic notion of Turing Machine. A computation for his kind of system involves a reader detecting symbols on a tape and moving the tape and printing new symbols as a result. A physical system implements a computation when the transitions between the states of the physical system reflect or mirror the abstract computation in question. Chalmers has argued that there is a set of computations which, when implemented, are necessary and sufficient for the existence of consciousness (that is, holding the relevant laws in place). Consciousness and mind are organizational invariants on his view.

The argument for this comes from a priori reflection on cases of what Chalmers calls fading and dancing qualia. The basic idea is to imagine having one's brain replaced bit by bit, say by replacing individual neurons one by one with a functional duplicate. Either one's conscious experience would change over the course of doing this or it would not. If it changes we seem committed to the claim that one cannot notice or report this change, since one is functionally exactly the same from moment to moment. To make this vivid we can imagine that you are eating your favorite food as the process is being carried out. You will go on saying that you don't notice any change and that the ice cream is delicious, etc. If this were the case then we would be radically out of touch with our own conscious experience. This gives us good reason to reject the claim that our conscious experience will change as this is happening. If so then the basic laws which relate the physical to the phenomenal depend on functional organization rather than the material of the brain. Thus if we were to build a robot that had a functional analog of availability it would have consciousness. And if we were to somehow upload the computational structure that our brain implements we would have uploaded our mind into a computer.

It is important to be clear that Chalmers is not endorsing a functionalist account of consciousness. He has rejected physicalism and all physicalist theories and is thinking of these computations as involving fundamental non-physical phenomenal properties.

IV. Acquaintance with Consciousness

Concepts, on Chalmers' account, are mental entities and are the constituents of thought. So when we think about our own experience we do so by employing some kind of phenomenal concept. Chalmers distinguishes at least two types of phenomenal concepts. There are concepts that pick phenomenal properties out by some relation, and those that pick out phenomenal properties directly. The relational concepts are the concepts we use in our public language community, the concept that I have individually, and indexical concepts. In each case the

referent of the term is picked out by some relation. So, in the case of the public language concept, ‘red’ picks out whatever in my community typically causes red experiences. This may be the same or different from what ‘red’ picks out for me individually. In the same way when I am having a visual experience of green and I think ‘this experience is pleasant,’ the concept ‘this experience’ picks out the phenomenal experience I am currently having and so its content is determined relationally.

The other kind of concept picks out phenomenal properties directly via their intrinsic natures. These he calls ‘pure phenomenal concepts’. One natural way to see the difference here is to think about pre-release Mary in her black and white room. She will be able to use the word ‘red’, as she will know that people often talk about red outside the room. So pre-release Mary has the public language community concept. The reference of this term is determined by whatever it is that actually produces red experiences.

When Mary is let out of her room and sees red for the first time she will be able to form a concept that picks red out as such. She can think ‘ah *this* is what they meant when they called things ‘red’’. She can also form the thought ‘this very experience is red’ where she is, so to speak, pointing to the experience and noting that it is red. ‘This very experience’ is the indexical concept. When Mary has the thought ‘this very experience has a red quality’ the indexical concept picks out her red experience. The other side of the thought is occupied by the pure phenomenal concept.

Pure phenomenal concepts can become part of our phenomenal beliefs. So when I believe that I am seeing red the phenomenal property of redness is taken up and becomes part of the belief itself. To see this Chalmers invokes a thought experiment about inverted Mary. Inverted Mary sees what we would call green in response to the things that we both call red. So Invert Mary will see fire trucks as green and tomatoes as green even though she will call them ‘red’. When invert Mary is let out of her black and white room and sees a tomato for the first time she can think ‘this experience is what it is like to see red’. She will have a different thought than the one that non-inverted Mary has (because she is not having an experience that we would recognize as red). What this shows, for Chalmers, is that these kinds of beliefs cannot be reductively accounted for because they have as one of their parts an irreducible phenomenal quality. Pure phenomenal concepts are very close to what Bertrand Russell had in mind by knowledge from acquaintance.

They also serve as the basis of our knowledge about consciousness. A direct phenomenal belief is partially made up out of the phenomenal property it is about. The subject is acquainted with the phenomenal property. This allows Chalmers to respond to various objections to standard forms of property dualism. For instance it is sometimes held that if epiphenomenalism is true then we cannot know about our own consciousness. But on the acquaintance view we

can know about it, though not via a way that is causal or functional. We have a direct kind of knowledge of our own conscious experiences. It also allows him to give an account of how experiences can justify beliefs. They do so through the relation of acquaintance.

V. Non-Reductive Representationalism

So what are these phenomenal properties that form the basis of phenomenal concepts and which are fundamental, non-physical aspects of reality? Chalmers suggests that it is natural to think that they are representations. This is because we can assess them in terms of accuracy and inaccuracy (Chalmers 2010 p 345).

Representationalism is very popular in the philosophy of mind (See chapter 11) though most, unlike Chalmers, hold a reductive version of it. Chalmers defends what he calls an internal Fregean representationalism. This is the view that phenomenal properties are identical to a certain kind of representation and that there are at least two kinds of content to a visual experience. On the one hand is the representation of the way the world is. So if I am looking at a red circle then my perceptual experience consists in a representation of the world as containing a red circle. This is what he calls the Russellian content. Then we have the mode that representation is presented under, which is the Fregean content. Here Chalmers argues that the phenomenal quality is presented as being typically caused by some external objects. The representation also includes what he calls a 'manner of representation'. For instance representing red visually and representing it in belief involve two different manners of representation. This is distinct from the mode of presentation, which is an aspect of the intentional content of the manner of representation. So for Chalmers phenomenal properties are phenomenal manners of representation that have as their content a Russellian color property (in the case of color experience) which is presented in a particular way; in particular it is presented as being typically caused by the external physical property or object.

In virtue of having Russellian contents our experiences represent the world as instantiating certain phenomenal properties. The natural next question is whether objects in our world actually do have these kinds of properties or not. Chalmers calls worlds where objects do instantiate phenomenal properties an 'Edenic world'. In an Edenic world when an apple looks red it is because the apple itself has the property of being phenomenally red. Chalmers argues that our world is not an Edenic world, which means that we represent the world as having properties that it does not actually have.

The strongest reason for thinking this comes from thinking about color inverts. Suppose that we have you and your inverted twin looking at an apple. You experience it as what we would call red, while your invert twin experiences it as

what we would call ‘green’. We have no reason to think that either you or the invert is getting it right yet if both are veridical then we have to say that objects instantiate opposite properties. But this is absurd so Chalmers concludes that it is better to think that the world does not actually instantiate these properties.

If our world is not an Edenic world then what kind of properties do the objects in our world instantiate? Chalmers argues that the objects in our world must have properties that ‘match’ Edenic properties. To match, roughly speaking, is to play the same role. In Eden phenomenal properties cause in perceivers perfect experiences. The properties in our world then can be said to match that role in so far as these properties bring about in us experiences with phenomenal properties. Roughly speaking the properties of objects in our world can be said to be doing the same kind of work that Edenic properties do in Eden, which is just to cause in us experience with phenomenal properties.

One rival camp of representational theory is the higher-order thought theory of consciousness. This kind of theory is often defended by physicalists, but it is possible to hold a non-reductive non-physical version of it. Chalmers has developed a prima facie case against these kinds of theories, stemming from considerations about the unity of consciousness, and we will end this section by briefly looking at that argument.

Chalmers distinguishes several kinds of unity but the most philosophically relevant for the present purposes is what he calls ‘subsumptive phenomenal unity’. Intuitively this is the idea that all of our conscious mental states at any given time are necessarily bound into a unified whole. He is tentative about endorsing it, but he does say that it has a strong intuitive appeal to him. However, if it is true, there are many theories of consciousness that would not be able to account for it. Higher-Order Thought theories can account for the unity of consciousness but cannot account for the necessity of this unity. The same is true of several versions of reductive representationalism. If it really is a necessary fact about our experience that it is subsumptively unified, then that would be strong evidence against the higher-order thought theory of consciousness.

VI. Other Issues in the Philosophy of Mind

Up until this point we have focused on issues surrounding consciousness. We will conclude by briefly looking at how the ideas developed here can be applied to other issues in the philosophy of mind.

Recall the Twin Earth thought experiment that we introduced earlier. Many philosophers have taken the moral of that story to be that natives of Twin Earth have thoughts about XYZ (not H_2O) and so the content of their mental states will be different than ours. When they believe that water is wet they believe

something about XYZ. When I believe it, I believe something about H_2O. If I were to be suddenly transported to Twin Earth and thought to myself "that water is wet," while looking at a glass of (what I didn't realize) was XYZ, my belief would be false. This suggests that the contents of thoughts are broad in the sense that we can have people with the same psychological make up who nonetheless have beliefs with differing contents. This is counter intuitive in that it seems that the content of a thought or belief should be narrow in the sense of being determined by mental things rather than environmental things. Given this we would expect that there should be something that is common to the beliefs that we and the Twin Earthers make.

Chalmers (2012) has argued that we can use the two-dimensional analysis to shed light on this debate. The people on Twin Earth have mental states that have different secondary intensions. But we share our primary intensions with them. If so, then primary intensions can be used to make sense of narrow content. Mental states have both kinds of content on Chalmers' account. Each belief or thought can be associated with both a primary intension that depends on the mental make up of the subject and a secondary intension that depends on the environment that the thinker is situated in.

This should not be taken to mean that he insists that all mental phenomena must be in the head. Chalmers (Clark & Chalmers 1998; see also chapter 13) has argued that the mind can extend outside of the skull and into the environment. The mind, for Chalmers, minus that part that involves consciousness, consists in the performance of certain functions like the ones listed at the beginning of this chapter as the easy problems. These functions are performed by the brain but there is no reason in principle that they couldn't be performed by suitable functional devices outside the head. So for instance, if I am currently thinking that 'I am in New York City' and this thought is realized by a certain computational process in my brain it follows from the principle of organizational invariance that we could replace the neurons performing that computation with functionally equivalent artificial neurons. This would not affect my belief or its content. Nor does it seem to matter whether this process occurs in the brain or not. As long as it is connected to the brain in the right way so as to allow normal functioning and unimpeded computation, the mind will be undisturbed. Thus if you are looking at some H_2O and believe that it is wet, and this belief is realized by a computer outside of your head (connected and functioning in the right way to your brain), and another, psychologically similar, person is on Twin Earth looking at a glass of XYZ but whose belief is realized by computations in the brain, you both have beliefs that have the same narrow content and different broad content. One of them has a belief that is extended beyond the skull, but that is the only difference between them.[2]

[2] Thanks to David Chalmers and Andrew Bailey for comments on an earlier version of this chapter

Work Cited:

Balog, K. (2012) “Acquaintance and the Mind-Body problem”, in *New Perspectives on Type Identity: The Mental and the Physical*, Hill and Gozzano (eds.), (pp. 16-43), Cambridge University Press

Block, N. & Stalnaker, R. (1999) “Conceptual Analysis, Dualism and the Explanatory Gap" *The Philosophical Review*

Block, N. (2007) “Max Black’s Objection to the Identity Theory” in Block, N (ed) *Consciousness, Function, And Representation* MIT Press

Block N. (forthcoming) “Functional Reduction”

Brown, R. (2010) “Deprioritizing the A Priori Arguments against Physicalism” *Journal of Consciousness Studies* 17(3-4):47-69

Clark, A. & Chalmers, D. (1998) “The Extended Mind” *Analysis* 58:10-23

Chalmers, D. (1996) *The Conscious Mind: In Search of a Fundamental Theory* Oxford University Press

Chalmers, D. (2010) *The Character of Consciousness* Oxford University Press

Chalmers, David J. (2011). “A computational foundation for the study of cognition”. *Journal of Cognitive Science* 12 (4):323-357.

Chalmers, D. (2012) *Constructing the World* Oxford University Press

Frankish, K. (2007) “The Anti-Zombie Argument” *Philosophical Quarterly* 57(229):650- 666

Jackson, F. (1986) “What Mary Didn't Know” *Journal of Philosophy* 83: 291-295.

Kripke, S. (1980) *Naming and Necessity* Harvard University Press

Levine, J. 1983. “Materialism and qualia: the explanatory gap”. *Pacific Philosophical Quarterly*, 64: 354-361.

Mandik, P. & Weisberg, J “Type-Q Materialism” In Chase Wrenn (ed.), Naturalism, Reference and Ontology: Essays in Honor of Roger F. Gibson. Peter Lang Publishing Group (2008)

Nagel, T. (1974) “What is it Like to be a Bat?” *The Philosophical Review* 83(4): 435-450

Putnam, H. (1973). "Meaning and Reference," *Journal of Philosophy* 70: 699-711

Russell, Bertrand. (1912). *The Problems of Philosophy.* Oxford University Press.

Russell, B. (1927) The Analysis of Matter