

# What is Consciousness?

## Philosophy behind the mind

Any parent can vouch for just how complicated things become when a child starts reaching the age of hard questions. We've all been there:

*"Why is the sky blue?"*

*"Because, honey, the molecules in the air scatter blue light more than other colors."*

*"Oh. Where do babies come from?"*

*"Sperm and egg come together to form a zygote, which develops into a fetus and then a child."*

*"Hmm. Is there a God?"*

*"Maybe."*

*"What about Santa?"*

*"Definitely."*

*"Why does his breath always smell funny?"*

*"He uses a special mouthwash from the North Pole."*

*"Daddy, what's consciousness?"*

*"Um..."*

If someone asked you to explain consciousness, could you do it? Very, very smart people have spent their entire careers trying to understand the answer to that question. It is surprising that something we all experience is so hard to explain. The difficulty comes in describing the "what it's likeness" that characterizes consciousness. There's something it's like to experience the color red, to taste chocolate, to feel happy or sad. Philosophers call this phenomenology. Unlike other worldly stuff, it isn't something we can point to or hold in our hand. It's not something we've been able to calculate. And we've yet to find a rigorous method of measuring it.

In 1994 David Chalmers published a paper explaining why consciousness is such a challenging phenomenon to understand. Although he wasn't the first to discuss these challenges, he was the first to categorize them into two types of problems: "easy" problems and the "hard" problem. Easy problems involve the explanation of how the mind integrates information, focuses attention and allows us to report on mental states. Though not a piece of cake, such problems are easy because solving them only requires that we determine the mechanisms that explain these behaviors. Easy problems are physical by nature, falling within the empirical domains of psychology, cognitive science and neuroscience. Given the current trend in science of the mind, we're confident that one day we will solve these problems.

The hard problem, by contrast, may never be solved. Specifically, the hard problem is determining why or how consciousness occurs given the right arrangement of brain matter. What makes it hard is that we cannot just point to some physical mechanism to solve it, for that would be the solution to the easy problem. Instead, our goal is to explain why certain physical mechanism gives rise to consciousness instead of something else or nothing at all. Consider an analogy from physics: knowing every equation predicting how

# What is Consciousness?

## Philosophy behind the mind

mass and gravity interact does not tell us why they interact *in the way they do*. To understand why mass and gravity interact, we must appeal to highly esoteric explanations involving relativity, quantum mechanics or string theory.

But while theoretical physicists have produced some pretty specific models that are ready to be tested with the likes of the Large Hadron Collider, consciousness lacks the sort of general consensus that would allow us to move on and test our theories. And for good reason—the hard problem is tricky.

Some argue that the hard problem simply is unsolvable. The argument for this view can take two different forms. The first argument is that our puny brains aren't capable of coming up with a solution, for our brains do not have the ability to process the complicated information that would lead to an understanding of consciousness. The second argument is that a solution to a problem requires that you aren't a part of the problem. What does this mean? To solve a problem, or so goes the argument, you must have a bird's eye view of all the facts. But since we are all conscious, we can never have such a view. We simply cannot solve the hard problem because we don't have access to the level of information necessary to piece everything together.

I think this argument fails, for proponents of this view don't explain why we cannot come to understand such high concepts through induction. Inductive reasoning is the "bottom-up" logic we often use to construct general belief from individual examples.

One way we use induction every day is when choosing what to wear. We choose a particular type of outfit to wear based on past weather patterns surrounding that morning. If we know the outdoor temperature has been 25 degrees Fahrenheit for the last 10 days, we can assume it will be cold outside, thus it is appropriate to put on winter clothes. Of course, you could fail to arrive at a true proposition through induction. For example, if you live in St. Louis sometimes you might find that a 65-degree day follows a 25-degree day. In that case the belief you arrived at through induction is false. You instead should choose spring clothing. But induction seems to work in many cases, especially in the physical and mathematical realms. So it's unclear why we could not use inductive reasoning to solve the hard problem.

This inductive approach is indeed what many philosophers and cognitive scientists have tried to take. Based on what is known about phenomenal states along with the brains that possess them, many theories of consciousness have emerged, leading to huge debates in philosophy and the sciences. Perhaps the greatest debate has focused the distinction between *dualism* and *physicalism*.

Physicalism holds that consciousness is entirely physical. One type of physicalism denies there is a problem at all. For folk in this philosophical camp, called *identity theorists*, the mental just *is* the physical, that is, the mental is nothing else than certain physical situation that obtains given certain arrangements of atoms. Why would one want to hold this view? First off, its simplicity is attractive. We need not appeal to "spooky" concepts

# What is Consciousness?

## Philosophy behind the mind

like emergence to explain what's going on when we're conscious. Second, it gives us a really good reason to think that computers can become conscious. If brains are nothing but biological implementations of computers running a certain program, it's possible that a silicon chip could run the same software as us.

Another theory, *functionalism*, holds that mental states are constituted by the function or role they play in a given system. Under this view, mental states exist as causal relations to other mental states. Functionalism is especially popular among computationalists, those who believe the brain is just a biological implementation of a computer. According to computationalists the brain is one system physically able to realize mental states, other systems such as computers could also realize these mental states. However functionalism does have its weaknesses. Many philosophers have argued that the theory is insufficient to account for consciousness because the role a mental state plays doesn't explain why the state must be one that is conscious. It's not clear why all our mental states wouldn't just be processed unconsciously.

Fundamentally different than physicalism, dualism is the theory that consciousness somehow falls outside the domain of the physical. To be a dualist, you need not believe that consciousness is a totally non-physical entity floating about the tops of our physical brains, you simply must believe that the hard problem is not solvable merely through sole appeal to the physical.

There are many types of dualism purported to best solve different aspects of the hard problem. For example, *Cartesian dualism*, one of the oldest forms of substance dualism, holds that there are both physical and non-physical substances and that consciousness is located within non-physical substance. Another theory, *property dualism*, holds that consciousness is a non-physical property that emerges from the same things that give rise to physical properties. Property dualists believe that neural activity has both physical and non-physical properties. Physical properties include things like electromagnetic potential while non-physical properties include things like consciousness.

Property dualists can further be distinguished into fundamental, emergence and neutral monist groups. Proposed by Chalmers, *fundamental* property dualism holds that conscious properties are basic properties of the universe similar to physical properties like electromagnetic charge. According to Chalmers, these properties can interact with other properties such as physical properties, but like fundamental physical properties, conscious properties are their own distinct fundamental entities. Consciousness works like electrical charge or other physical properties do: it may cause physical matter to transition among physical states and these physical states in turn may affect consciousness.

A closely related theory, *panpsychism*, holds that all aspects of reality have some "psychological" properties apart from their physical properties. This type of property dualism suggests that the universe has consciousness at its base. While this theory

# What is Consciousness?

## Philosophy behind the mind

certainly is elegant, it is thought by some to carry metaphysical baggage. One complaint has been that, if this theory is true, then all matter would have a certain element of consciousness to it. Because consciousness is inherently connected with the phenomenal, this is a peculiar result, for it's hard to imagine how there could be something it's "like" to be an electron, table, chair, tire or other inanimate object.

Furthermore, consciousness seems to have boundaries. There is something that it's like to be me, to be you and to be someone else. Panpsychism has trouble explaining how phenomenology has a boundary. If consciousness is a fundamental property of matter, it appears as though all matter and collections of matter have a conscious aspect. Collective consciousness entails that, not only do individual electrons have consciousness, so do neurons and collections of neurons. It's hard to see the level at which *you* would exist as a conscious being. How many neurons constitute *you*? Is it one of them, a few of them, *all* of them? Of course, one might respond that, compared to other problems of consciousness, the boundary problem is rather small. In that case, the elegance of the theory might outweigh its weaknesses.

Like fundamental property dualism, emergent property dualism holds that consciousness is a property that emerges from particular types of physical arrangements of matter. But it differs in that consciousness is a property that emerges *over and above* what could be predicted given the arrangements of the matter's physical properties. Yet another type of property dualism, *neutral monist property dualism* holds that physical and conscious properties are *both* dependent on some more basic level of reality.

Why would one be motivated to hold one of the above dualist views? What is the need to postulate such spooky entities as non-physical conscious properties? Physicalists have trouble explaining several aspects of consciousness in a way that is consistent with our observations of how physical properties interact.

Frank Jackson's thought experiment gets at one of the problems of the physicalist approach:

Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room via a black and white television monitor. She specializes in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like 'red', 'blue', and so on. She discovers, for example, just which wavelength combinations from the sky stimulate the retina, and exactly how this produces via the central nervous system the contraction of the vocal cords and expulsion of air from the lungs that results in the uttering of the sentence 'The sky is blue'. [...] What will happen when Mary is released from her black and white room or is given a color television monitor? Will she learn anything or not?

Jackson argues that Mary in fact will learn something new: what it's like to see blue. Even though she knows everything about the science of color, she has never experienced

# What is Consciousness?

## Philosophy behind the mind

color. Jackson later discusses another person, Frank, who experiences a color that no other human has ever seen. It seems that no matter how much information we have about the neural processes behind Frank's experience of the color, we will never know what it's like to have Frank's experience. According to Jackson, that there is something learned only through phenomenal experience shows that experience is not reducible to the purely physical.

Chalmers poses a different problem for physicalism. He tells us to imagine a molecule-by-molecule replica of ourselves, exact down to each individual neuron and its firing state. He then asks to consider whether it's conceivable that this creature is a philosophical zombie, that is, a creature that is behaviorally indistinct from us but lacking consciousness. Even though on the outside a zombie appears conscious just like you and me, is it possible for it to be "dark" inside? Chalmers says yes. And if that's the case, he argues, then physicalism must be false. If we admit we can conceive of a world physically indistinct from ours yet lacking phenomenal consciousness, we cannot hold that the mental wholly is based on or reduces to the physical.

But can we really conceive of an exact copy of ourselves that lacks consciousness? On the face of it, most of us would say no. I simply cannot imagine what it would be like to be me without there being anything it's like to be me. But that's not what Chalmers is asking us to do. As part of his argument, he makes a distinction between positive and negative conceivability. Positive conceivability means that something is conceivable insofar as you have an imaginative picture of the situation that obtains if the conception were true. But negative conceivability means that something is conceivable insofar as you cannot rule it out *a priori*, that is, you cannot rule it out from reason alone.

This distinction is crucial for Chalmers' argument to work. Now all he needs is to get you to admit that reason alone does not provide you with motivation to reject the possibility that your exact physical copy could lack consciousness. Since physicalism holds the physical gives rise to the mental, and because we've provided an example of a situation wherein the physical does not give rise to the mental, we must either reject physicalism or reject our negative conception of zombies. Since it isn't possible to reject the latter, we must reject the former. So physicalism must be false.

Philosophers like Chalmers and Jackson argue that the only appropriate action is to reject physicalism and move in the direction of dualism. These arguments have gotten much attention in the literature from authors on both sides of the debate. And many objections have arisen.

In response to Jackson, Laurence Nemirow argues that knowing what it's like to have an experience is the same thing as knowing how to imagine having the experience. In Mary's case, she didn't learn something new, she just gained the ability to experience color. David Lewis makes a similar argument: Mary gained the ability to remember, imagine and recognize.

# What is Consciousness?

## Philosophy behind the mind

Earl Conee makes a slightly different argument: knowing what it's like does not require having an imaginative experience. He introduces Martha, who is able to visualize intermediate shades of colors she has not experienced that fall between pairs of shades that she has experienced. Martha is not familiar with the shade of "cherry red" but knows that cherry red is halfway between burgundy red and fire red, two shades she has experienced. According to Conee, Martha could know what it's like to experience cherry red, but so long as she never imagined it, she might never have experienced that color.

Interestingly, Jackson himself changed his mind several times as to the implications of the Mary argument. Although the argument appeared to offer a strong problem for physicalism, he also believed that all behavior is caused by physical forces. Given that the argument seemed to prove the existence of non-physical phenomenology, Jackson needed to find a way square it away with his conception of the physical. His strategy was to argue for *epiphenomenalism*, which holds that phenomenal states are caused by physical states, but phenomenal states do not affect the physical. Epiphenomenalism means that our phenomenal states are akin to movies that constantly play while unconscious brain processes direct all behavior.

Later on, Jackson decided that Mary's conscious recognition of color *did* have an effect on the physical: it made her say "wow." If a conscious recognition could be the cause for an expression requiring, which requires a change in mental states, then consciousness appears to play some role in guiding our behavior. So epiphenomenalism would be false. He argued that physicalism could in fact account for this "wow" experience. If the color experience is entirely contained within in the brain, it's possible that a new experience can cause further changes to the brain, resulting in the utterance of a "wow" statement. He used the analogy of akinetopsia, a deficit that causes the inability to perceive motion. He argued that someone cured of this condition would not discover anything new about the world (since she knows about motion). Instead, surprise would be the response of a brain now able to see motion.

Chalmers' zombie argument has been subject to scrutiny as well. Because it's logically valid, that is, because evaluation of the argument's premises do lead it its conclusion, one must attack the premises in order to undermine the argument. One objection has been that the zombie argument undermines itself. Because the argument builds a world that is defined entirely physically, the world necessarily would contain consciousness; therefore zombies cannot exist in it. Other philosophers attack negative conceivability. They argue that the mere possibility of something does not mean it actually exists. Although zombie worlds might be possible, it is still the case in this world that consciousness is entirely physical.

So which theory wins? Dualism or physicalism? It depends on who you ask. Many empirical researchers are hardcore physicalists, but not all are. The answer to this question will require more insight into the fundamental structure of our physical world. It might turn out that a really consistent theory of physics could lead us to understand

# What is Consciousness?

## Philosophy behind the mind

exactly what consciousness is. But it might not. Consciousness might forever remain a mystery.