To appear in: *The Constitution of Consciousness*, Volume 2, S. Miller, ed. John Benjamins Publishing Company, 2012

The Status of Consciousness in Nature

Berit Brogaard
Departments of Philosophy and Psychology, Center for Neurodynamics, University of Missouri, St. Louis
brogaardb@gmail.com
March 30, 2012

Abstract

The most central metaphysical question about phenomenal consciousness is that of what constitutes phenomenal consciousness, whereas the most central epistemic question about consciousness is that of whether science can eventually provide an explanation of phenomenal consciousness. Many philosophers have argued that science doesn't have the means to answer the question of what consciousness is (the explanatory gap) but that consciousness nonetheless is fully determined by the physical facts underlying it (no metaphysical gap). Others have argued that the explanatory gap in the sciences entails a metaphysical gap. The explanatory gap exists, they say, because there are two fundamental properties in the world that do not reduce to one another: Phenomenal and physical. This position is also known as 'property dualism'. A famous argument, formulated and defended at great length by David Chalmers, uses conceptual tools to argue for a metaphysical gap. When we just look at what the notion of phenomenal consciousness implies, we will find that it doesn't rule out that there could be entities functionally and physically identical to us but without phenomenal consciousness. A couple of further argumentative steps can get us from here to the conclusion that laying down the physical facts of our world does not necessitate phenomenal consciousness. I argue that this argument is compelling but that accepting the conclusion doesn't have the implication that science cannot discover what consciousness is. I begin by outlining and assessing a number of different positions philosophers and scientists have recently defended regarding the link between neurological systems and consciousness, I then argue that even if property dualism is true, that doesn't necessarily prevent the sciences from discovering what constitutes consciousness. That is, there may be no explanatory gap even if there is a metaphysical gap.

Contents

1. The Hard Problem of Consciousness

One of the most puzzling features of consciousness is the problem David Chalmers has named ‘the hard problem’. In “Facing Up to the Problem of Consciousness”, Chalmers describes the hard problem as follows:

> The really hard problem of consciousness is the problem of *experience*. When we think and perceive, there is a whir of information-processing, but there is also a subjective aspect. As Nagel (1974) has put it, there is *something it is like* to be a conscious organism. This subjective aspect is experience. When we see, for example, we *experience* visual sensations: the felt quality of redness, the experience of dark and light, the quality of depth in a visual field. Other experiences go along with perception in different modalities: the sound of a clarinet, the smell of mothballs. Then there are bodily sensations, from pains to orgasms; mental images that are conjured up internally; the felt quality of emotion, and the experience of a stream of conscious thought. What unites all of these states is that there is something it is like to be in them. All of them are states of experience.
>
> It is undeniable that some organisms are subjects of experience. But the question of how it is that these systems are subjects of experience is perplexing. Why is it that when our cognitive systems engage in visual and auditory information-processing, we have visual or auditory experience: the quality of deep blue, the sensation of middle C? How can we explain why there is something it is like to entertain a mental image, or to experience an emotion? It is widely agreed that experience arises from a physical basis, but we have no good explanation of why and how it so arises. Why should physical processing give rise to a rich inner life at all? It seems objectively unreasonable that it should, and yet it does. (Chalmers 1995)

The hard problem of consciousness is not to specify the neural correlates of consciousness but to explain ‘why and how’ phenomenal consciousness arises. Even though we know a lot about the brain, we don’t yet know how neural or other physical states lead to the rich inner life that we are all familiar with. The causal chain from neural or other physical states to conscious states remains a mystery.

Some physicalists hold that physical theory ultimately will be able to explain the process by which physical states lead to conscious states. Other physicalists propose that there is no way that physics or any other field of study could ever provide an explanation of how conscious states arise but that the physical still determines the mental. Conscious states somehow supervene on physical states in a non-causal but constitutive fashion. Dualists like Chalmers himself think that we cannot provide an explanation of consciousness in purely physical terms.

He furthermore holds that the physical isn't constitutive of consciousness. Rather, consciousness is an extra property of brain states that emerges from physical processes.[1]

I agree with Chalmers that consciousness probably isn't purely physical, and that it therefore isn't possible to explain consciousness in terms of purely physical states. But recognizing as much does not rule out that science can explain consciousness. I look at two candidates to be scientific theories that can help close the explanatory gap at different levels of explanation: String theory and non-reductive neuroscience.

2. A Priori Physicalism

According to Joseph Levine (1983), the explanatory gap between the mental and the physical refers to our incomplete understanding of how consciousness depends on the physical. There is an explanatory gap between the mental and the physical just in case we cannot rule out on a priori grounds alone that they can come part. The ontological gap, on the other hand, refers to a state of the universe in which the physical does not fully determine the mental. If there is an ontological gap, it is possible for there to be two physically and behaviorally identical systems such that the first system is conscious, whereas the other system is not.

A priori physicalism, or what Chalmers calls 'type A materialism, is the view that if you knew all the physical facts, you would need no further information in order to infer all the mental facts. A priori physicalism contrasts with a posteriori physicalism, or what Chalmers calls 'type B materialism'. On this view, you cannot infer the mental facts, even if you are given all the physical facts. This is because there is a gap between mental and physical facts but the physical facts nonetheless fully determine the mental facts.

One motivation for a posteriori physicalism is that it provides a good way to explain the intuitive pull of cases like Frank Jackson's knowledge argument. The knowledge argument runs as follows (Jackson 1982, 1986). Mary is an excellent neuroscientist confined to a black-and-white room with black-and-white television screens hooked up to external cameras and access to everything that has ever been written about colors and color perception. After years of studying in her cell Mary comes to know every physical fact about colors and color perception. But she is still not satisfied. There is something books and black-and-white television cannot teach her: What it is like to perceive in colors.

The story about Mary was originally meant to undermine physicalism of any kind. If Mary knows every physical fact about colors and color perception but is still able to learn something new about phenomenal properties upon her release,[2] then phenomenal truths are not physical. So, physicalism is false.

As it stands, however, there is an obvious worry about the argument. If Mary already knows every physical fact about the colors and color perception prior to her release, and facts about

[1] Chalmers is sympathetic to a number of different forms of dualism as well as panpsychism.

[2] Following Chalmers (2004a) I take phenomenal properties ("what it's like") to be properties of a mental state, a brain state or an individual.

phenomenal properties are physical, then she already knows every fact about phenomenal properties. So, for it to be true that Mary learns something new upon her release, it must be implicitly assumed that phenomenal properties are not physical. But this begs the question against the physicalist. To avoid begging the question it is better to say that, prior to her release, Mary knows all the lower-level physical truths (e.g., the truths of ideal physics, chemistry, and biology) (Chalmers 2004b, 2006).

Even when modified in this way, however, the knowledge argument does not threaten to undermine just any kind of physicalism. It is open to argue that Mary cannot come to know all phenomenal truths in her black-and-white cell because the phenomenal truths are not a priori necessitated by the lower-level physical truths and therefore are not deducible from the lower-level physical truths. The knowledge argument is thus best construed as an argument against a priori physicalism, the position that the higher-level phenomenal truths are a priori necessitated by the lower-level physical truths. To refute a posteriori physicalism, the position that phenomenal truths are necessitated but not a priori necessitated by the lower-level physical truths, additional premises are needed.

The success of the knowledge argument stands and falls with the learning claim, the claim that upon her release Mary learns a new fact about color perception, a fact which she would have known prior to her release, had a priori physicalism been true. To refute the knowledge argument a priori physicalists must explain away the appeal of the learning claim.

One way to do this is to argue that while Mary acquires new knowledge of what it's like to perceive in colors upon her release, the knowledge she acquires is not propositional. Several thinkers have taken this route. Lawrence Nemirow (1980, 1990), David Lewis (1983, 1988), and others, have argued that to acquire new knowledge of what it's like to perceive in colors just is to acquire new abilities to imagine, recognize and memorize color experiences, and Terry Horgan (1984), John Bigelow and Robert Pargetter (1990) and Earl Conee (1985, 1994) have argued that knowledge of what it's like to perceive in colors reduces to direct acquaintance with phenomenal color properties.

The ability and acquaintance replies, as originally formulated, must be distinguished from closely related replies to the effect that pre-release Mary already knows what it's like to perceive in colors yet acquires new knowledge or new skills nonetheless, knowledge or skills which are distinct from knowledge of what it's like to perceive in colors. Whereas the first kind of reply attempts to reduce knowledge of what it's like to perceive in colors to nonpropositional knowledge, the latter kind of reply merely attempts to explain away the intuition that Mary comes to know what it's like to perceive in colors by appealing to other kinds of knowledge, skills or concepts which Mary acquires upon her release.

The first sort of reply is considerably easier to refute than the latter. If 'knowing what it's like' ascriptions attribute propositional knowledge, then it is simply false that Mary both learns what it's like to perceive in colors and fails to gain propositional knowledge. It may be that our intuitions about what Mary learns upon her release are unreliable and that Mary doesn't really

learn what we think she learns, but if ‘knowledge what it’s like’ ascriptions attribute propositional knowledge, and Mary comes to know what it’s like to perceive in colors upon her release, then Mary gains knowledge of a new fact, a fact which it would seem she should already have known if a priori physicalism had been true.

Do ‘knowledge what it’s like’ ascriptions attribute propositional knowledge? On a common analysis of knowledge-wh, Mary knows what it is like for her to see blue at t iff there is an x such that Mary knows that x is what it is like for her to see blue at t (Brogaard 2009, Brogaard 2011). ‘What it’s like for Mary to see blue at t’ is equivalent to ‘what seeing blue is like for Mary at t’. This, in turn, is equivalent to the predicate nominal ‘the x: seeing blue is like x for Mary at t’, where ‘is like x’ plays the same semantic role as ‘has the property x’. When embedded in attitude contexts, predicate nominals denote properties. So, ‘what it’s like for Mary to see blue at t’ denotes a property.

The learning intuition as formulated above was expressed in the form of a knowledge-wh claim. But one can also express it in the form of a know-how claim, viz. the claim that upon her release Mary comes to know how it feels to see blue. This version of the knowledge argument is, in my view, much more forceful than the original. While it is somewhat plausible that pre-release Mary already knows what it’s like to see blue, it is highly implausible that pre-release Mary already knows how it feels to see blue. But knowledge-how is just a species of knowledge-wh. So, if Mary comes to know how it feels to see blue at t, then there is a phenomenal property or state Q such that Mary knows that Q is how it feels to see blue at t. So, Mary gains new propositional knowledge upon her release.

There are various sophisticated ways of explaining away the intuition that post-release Mary comes to know what it’s like or how it feels to see blue. These ways involve denying the legitimacy of the learning intuition, viz. the intuition that Mary comes to know what it’s like or how it feels to see blue. For example, even if ‘knowledge what it’s like’ and ‘knowledge how it feels’ ascriptions attribute propositional knowledge, it is still open to argue that pre-release Mary already knows what it’s like or how it feels to perceive in colors and that post-release Mary simply acquires new abilities without thereby gaining new propositional knowledge.

The ability reply, however, is unsuccessful. As I have argued elsewhere, if a person acquires new abilities, she also acquires new propositional knowledge (Brogaard 2011). Abilities are states which are constituted by bodily capacities and procedures that have been internalized by the agent and which are therefore essentially mind-involving. The new knowledge you acquire when you acquire a new ability is a kind of implicit knowledge. An ability to ride a bike involves the skills it takes to write a bike but also an implicit strategy for achieving the results in the right sort of environment. An ability to recognize colors requires an internalized categorization system and a strategy for correctly classifying new stimuli.

Though there may be other possible ways to explain away the learning intuition, explaining it away is not child’s play. It is obvious that some new mental event takes place when Mary first

sees a red flower. I believe that regardless of the nature of that new event, this new event is unexplainable by third-person physical fact.

Note that the knowledge argument is not exclusively an argument against physicalism. Suppose dualism is true, and that Mary learns everything that one can learn in a black and white room about the lower-order physical and mental facts. Mary still learns something new the first time she sees a red flower. She learns what experiencing red is like from her first person perspective. What the knowledge argument shows is that because of the special first person acquaintance that it takes for someone to be conscious, third-person scientific theory cannot fully explain consciousness. Let us call this explanatory gap 'the first-person explanatory gap'. There is a further potential explanatory gap about which the knowledge argument does not have anything to say. This explanatory gap arises if not all third-person mental facts follow a priori from physical facts. Let us call this explanatory gap 'the third-person explanatory gap'.

3. A Posteriori Physicalism

A posteriori physicalism does not posit that the mental facts follow a priori from the lower-level physical facts. So exponents of a posteriori physicalism can take the learning intuition in Jackson's argument at face value. They can admit that Mary learns something new upon her release. Even though Mary couldn't figure out everything in her black and white room, the lower-level physical facts fully determine all mental facts.

This position strikes some people as odd. If the lower-level physical facts fully determine the mental facts, then why do the mental facts not follow a priori from the physical facts? Well, here is one way to understand it. It doesn't follow a priori from 'Allen Stewart Konigsberg was born on December 1, 1935' and 'Woody Allen directed Annie Hall' that Allen Stewart Konigsberg directed Annie Hall. But the former facts completely necessitate the latter.

In this case, however, we cannot infer on a priori grounds that Allen Stewart Konigsberg directed Annie Hall, because the starting facts were incomplete. If we were told that Woody Allen is identical to Allen Stewart Konigsberg, then we would have been able to infer the latter fact on a priori grounds. It is difficult to see how we could have a *complete* collection of lower-level physical facts that necessitate all the mental facts but nonetheless does not a priori imply them.

One way this can happen is if the information we need to get from the physical facts to the mental facts is not strictly about the physical world. Let us call the extra information we need to infer the mental from the physical on a priori grounds 'bridge laws.'

However, saying that inferring the mental facts from the physical facts requires bridge laws does not solve the problem. Even if Mary learns all lower-level physical facts *and* the bridge laws in the black and white room, and she has great deductive powers, she still learns something new when she is released from her captivity and is first exposed to colors. So we are back where we started.

One strategy that has been adopted by a posteriori physicalists to avoid the knowledge argument is the phenomenal concepts strategy (Carruthers 2000, 2005, Tye 2000, Stoljar 2005, Chalmers 2006). Phenomenal concepts are concepts that refer to our experiences. They are unlike all other concepts and cannot be explained by or reduced to other concepts. They are conceptually isolated and to acquire them, a particular kind of acquaintance with how the experience in question feels is required. What Mary learns upon her release from captivity is a phenomenal concept, a new way to refer to color experience. The explanatory gap arises, the proponents say, because a physical explanation of phenomenal concepts would require using only theoretical concepts. However, the explanatory gap does not entail an ontological gap because all mental facts supervene on the physical. Mary's possession of the phenomenal concept, for example, can be explicated in physical terms.

The explanatory gap that the proponents refer to is what I called 'the first-person explanatory gap' above. Phenomenal concepts motivate the existence of a first-person explanatory gap but it is far from obvious that they motivate the existence of a third-person explanatory gap. I return to this question at the end of the next section.

4. The Zombie Argument

There are other options for those unconvinced by the physicalist strategies for accommodating our feeling that physical and mental properties are distinct. One option is to take the appearance at face value. This view is also known as 'property dualism'. Chalmers is famous for having advanced the so-called zombie argument in favor of some kind of property dualism. The first premise of the argument concerns conceivability, or what is consistent with what we know on a priori grounds. The premise says that we cannot rule out on a priori grounds, or by reflection alone, that there are creatures that are physically and functionally identical to us but that don't have phenomenal consciousness. Chalmers calls these creatures 'zombies'. The second premise states that if we cannot rule out on a priori grounds that there are philosophical zombies, then it is possible that there are zombies. It follows that there are creatures physically and functionally identical to us but without phenomenal consciousness. But this conclusion is inconsistent with physicalism, which holds that the physical *necessitates* the mental.[3] Or to put it differently: Physicalism holds that it is necessary that our physically and functionally identical twins are conscious.

Though some thinkers have denied that zombies are conceivable, the most frequently challenged premise in the zombie argument is the second, which says that if it is conceivable that there are zombies, then it is possible that there are zombies.

[3] Chalmers admits that Russellian monism is compatible with the conclusion of his argument. Russellian monism is the view that physical theory deals only with observable properties and therefore cannot get at the inner substrate of things. When we conceive of the zombie scenario, we keep the physical facts fixed. But this does not rule out a change in the inner substrate of things. If the inner substrate of things is indeed physical, the mental may still supervene on the physical.

Chalmers' main argument for this premise is that conceivability and possibility come apart only for a posteriori necessities. A posteriori necessities are necessary truths which we can only figure out are true via empirical discovery. For example, as 'Hesperus' and 'Phosphorus' both refer to Venus, and it is necessary that Venus is self-identical, it is necessary that Hesperus is identical to Phosphorus. But the truth that both names refer to Venus is an empirical discovery, not something we can discover through reflection alone. We cannot rule out on a priori grounds, or by reflection alone, that Hesperus is not identical to Phosphorus. So a scenario in which Hesperus is not identical to Phosphorus is conceivable, despite being impossible. This could be a scenario in which Hesperus, the brightest object in the evening star, is Mars, and Phosphorus, the brightest object in the morning sky, is Jupiter.

Here is another case: It is necessary that chemically pure water is H_2O, but it took empirical investigations to determine the molecular structure of water, we could not figure it out by reflection alone. It is consistent with what we know on a priori grounds that water is not H_2O but a different compound, call it 'XYZ'. But as 'water' refers to H_2O, and H_2O is self-identical, it is not possible that chemically pure water is not H_2O.

There are also necessary truths with an a priori status. Good examples are mathematical, logical and analytic truths. We don't need empirical discovery to figure out that 2 + 2 is 4, that Modus ponens is valid or that no bachelor is married. When we introduce names by description, we create various a priori contingent truths. For example, when the police introduced the name Jack the Ripper as a name of the murderer of the London prostitutes, it become a priori true that Jack the Ripper murdered prostitutes. But Jack the Ripper could have failed to murder prostitutes. So 'Jack the Ripper murdered prostitutes' is not a necessary truth, even though we can figure out by reflection alone that it is true.

A posteriori necessary truths contain proper names and common nouns, such as 'Hesperus' and 'water'. This class of expressions are rigid designators, that is, they refer to the object they actually refer to in every possible world in which that object exists. Returning to the zombie scenario, this scenario can be described in terms that do not involve any rigid designators. As only descriptions that contain rigid designators can block the route from conceivability to possibility, the conceivability of the zombie scenario entails the possibility of the zombie scenario. But the possibility of the zombie scenario is squarely at odds with physicalism, which holds that the physical necessitates the mental.

If sound, Chalmers' argument shows that it is not necessary that the physical determines the mental. So the physical and the mental are ontologically distinct, which is to say that there is an ontological gap between the physical and the mental.

Proponents of the phenomenal concept strategy have argued that their strategy can also be used to explain the explanatory gap in the Zombie scenario (Tye 2000, Balog 2012). Let us consider a simple case: Red-Zombie is physically and functionally identical to me, he even has most of the mental states I have. But he has no awareness of the color red. Proponents of the phenomenal concept strategy accept the first premise that zombie scenarios cannot be ruled

out on a priori ground but they deny the premise that the explanatory gap entails an ontological gap. Their strategy in the case of red-Zombie would be to say that when we conceive of red-Zombie, what we really conceive of is someone who lacks the phenomenal concept that refers to red.

There is a forceful objection to this strategy for blocking the zombie argument.[4] Either phenomenal concepts can be explicated in physical terms or they cannot. If they can, then we can derive a contradiction. If we cannot rule out on a priori grounds that my zombie twin doesn't possess a phenomenal concept that refers to red, then there is an explanatory gap between the physical and phenomenal concepts. This entails that phenomenal concepts cannot be explicated in physical terms. But this contradicts our earlier assumption.

Suppose then that phenomenal concepts cannot be explicated in physical terms. Then it cannot be ruled out on a priori grounds that red-Zombie doesn't have phenomenal concept that refers to red. But the red-zombie scenario can be described in a neutral language that doesn't contain types of words that can generate a posteriori necessities and a priori contingencies, such as nouns and names. Since conceivability and possibility can come apart only in a non-neutral language, it is possible that the physical doesn't give rise to phenomenal concepts. It follows that there is an ontological gap between the physical and phenomenal concepts. So, physicalism is false.

The zombie argument, if sound, refutes physicalism.[5] What to put in its place? Chalmers does not propose a single candidate view to replace physicalism. But he does make several suggestions. One is that the mental emerges from the physical in a strong sense of emergence. In this strong sense, the physical doesn't determine the mental but nonetheless gives rise to it.

5. The Beginning of a Solution: The String Consciousness Model

We have looked at three positions concerning consciousness. A priori physicalists hold that there is neither an explanatory gap nor an ontological gap. A posteriori physicalists hold that there is an explanatory gap but they deny that that gap entails an ontological gap. Property dualists, as we have seen, hold that there is an explanatory gap and that this entails an ontological gap. What about the fourth possibility? Could there be an ontological gap but no explanatory gap? Well, that depends on what we mean by 'explanatory gap'. As mentioned earlier, there are two different kinds of explanatory gap. One is between physics and first-person conscious experience. The other is between physical facts and third-person conscious

[4]A variant of this argument can be found in Chalmers 2006.

[5]Chalmers (2009) adds that the zombie argument does not undermine a position known as 'Russellian monism.' Russellian monism holds that elementary particles have an intrinsic nature that can vary between this world and the zombie scenario, even if my zombie twin and I share all (relational) physical and functional properties.

experience. I don't think there is a way to close the first-person explanatory gap. The first-person perspective cannot be accounted for by any physical theory. However, I believe there can be an ontological gap without a third-person explanatory gap. This sort of scenario could happen if the fundamental theory of the universe posited at least two fundamental properties, which together could explain how consciousness arises.

One theory that may have the potential to do this is string theory, also known as M-theory. In this section I will briefly present string theory and outline some ways in which it might help to explain consciousness. It is not my goal to defend the theory here.

According to string theory, the fundamental entities of the universe are tiny, massless, one-dimensional open or closed strings that can vibrate with different frequencies, just like a rubber band or a guitar string. Different vibrations correspond to forces, such gravity, electromagnetism and radiation, and the subatomic particles. Because the strings are so small, no longer than the Planck length or 1.6×10^{-33} cm, they look like particles to us. As we cannot observe anything smaller than the Planck length, we cannot see that strings have extension. Every string has a space-time history that is described by functions Xm(s,t). The functions specify how the strings' two-dimensional "world sheet", represented by coordinates (s,t), is mapped onto space-time Xm.

Given string theory, particles and forces are manifestations of the vibrations of strings. To visualize what a manifestation is, think about how differently the edge of a table will look if you put it under a microscope. To the naked eye, it looks straight and smooth. Under a microscope it looks rough and curvy. Just like the straight and smooth look in some sense is a manifestation of an entity that is rough and curvy, so the particles and forces are manifestations of vibrating strings.

There are four elementary forces and 12 elementary "matter" particles. The four elementary forces are gravity, electromagnetism and the weak and strong nuclear force. The 12 elementary particles include six quarks and six leptons. The quarks are components of neutrons and protons. The leptons include the electron, two siblings of the electron and three neutrinos. Forces are made up of elementary "force" particles, which all exhibit wave-particle duality. For example, light is made up of photons. The strong and weak nuclear forces are made up of gluons and W and Z bosons. No one knows exactly what constitutes gravity. Each of the elementary particles appears point-like to us. If string theory is correct, however, all fundamental particles are vibrating strings of energy.

One of the most controversial aspects of string theory is that it postulates that our universe has 11 dimensions, as opposed to the four spatial and temporal dimensions. The reason for this is that strings must occupy 11 dimensions if they are to account for the manifestation of all particles and forces. The seven extra spatial dimensions are thought to be too small for us to see, possibly as small as the Planck length. The thought is that they are curled up into tiny geometrical spaces all around us.

String theory may provide an opportunity to explain consciousness. If we cannot explain consciousness in well known physical terms, maybe this is because consciousness is constituted by a separate kind of elementary force or particle. Given string theory, elementary forces and particles are vibrating strings. Consciousness, too, could be a manifestation of vibrating strings. Let's call this idea the 'string consciousness model'.

On the string consciousness model, there is no explanatory gap. Science can in principle give a full explanation of how manifestations of consciousness arise from certain vibrational modes of strings. But the model does not close the ontological gap. Manifestations of consciousness and manifestations of forces and particles arise from different types of vibrations. There is a range of fundamental vibrations, which do not reduce to one another.

Simply saying that consciously is composed of vibrating strings may not seem to help with the hard problem of consciousness. At least one challenge remains, namely that of explaining how vibrating strings combine to form macrolevel subjective experience. This challenge is an instance of what is sometimes called 'the combination problem'. William James formulated the problem as follows:

> Take a sentence of a dozen words, and take twelve men and tell to each one word. Then stand the men in a row or jam them in a bunch, and let each think of his word as intently as he will; nowhere will there be a consciousness of the whole sentence … Where the elemental units are supposed to be feelings, the case is in no wise altered. Take a hundred of them, shuffle them and pack them as close together as you can (whatever that might mean); still each remains the same feeling it always was, shut in its own skin, windowless, ignorant of what the other feelings are and mean. There would be a hundred-and-first feeling there, if, when a group or series of such feeling were set up, a consciousness *belonging to the group as such* should emerge. And this 101st feeling would be a totally new fact; the 100 original feelings might, by a curious physical law, be a signal for its *creation*, when they came together; but they would have no substantial identity with it, nor it with them, and one could never deduce the one from the others, or (in any intelligible sense) say that they *evolved* it (1890/1950: 160)

The combination problem challenges theories that posit micro-experiences or micro-subjects as fundamental to explain how these micro-experiences or micro-subjects could add up to macroscopic subjective experiences. As far as James is concerned, it is ridiculous to suppose that once we pile up enough microscopic experiences, it all adds up to the familiar macroscopic experience.

James' challenge is taking. However, it is not a knock-down objection to theories that posit consciousness at the fundamental level of reality. The combination problem, as stated, arises only if we assume that the hard problem of consciousness is to explain subjectivity. However, subjectivity is not a problem of consciousness. Nor is it a physical problem. Subjectivity is not a problem of consciousness because subjectivity arises even without consciousness. For

example, the dorsal-stream representations required for vision for action represent mostly egocentric/perspectival properties (e.g., the location of the computer mouse relative to my hand). Attention without consciousness presumably also involves subjectivity or a first-person perspective.

To see why subjectivity is not a physical problem, consider again the case of philosophical zombies (for our purposes here we can assume that they are sophisticated biological android without consciousness). To be able to reach to and grasp objects or otherwise interact with the world Zombies need a first-person perspective. But consider now the case of Zombie Mary. Zombie Mary has been locked in a room without any circular-shaped things for all of her life. She has had access to any information she would want through computer monitors, except specific information about how circular-shaped objects are presented to us when tilted. When Zombie Mary finally is released and for the first time sees a tilted coin, she is unlikely to assume that the coin is circular-shaped and hence will do a bad job when attempting to collect it. Through trial and error she will learn that tilted circular-shaped things in her environments are visually presented to her as oval-shaped.

The reason Zombie Mary does not have the ability to interact proficiently with circular-shaped objects when she is first released is that she doesn't yet possess the right concept. An analogy: Given externalism about concepts, you cannot possess the concept of water without having been acquainted with (or relevantly causally connected with) a watery environment. Likewise, seeing a tilted coin or bracelet as circular-shaped requires interaction with tilted circular-shaped objects in your environment (walking around these objects might suffice). Thus, for Zombie Mary to possess the concept *circular-shaped*, she must have interacted with circular-shaped objects in the environment outside her prison.

Of course, we could feed imprisoned Zombie Mary the right sort of data about the visual environment outside her prison. This would suffice for her coming to possess the concept. She is a zombie after all. But note that having a scientist poke regular imprisoned Mary's brain to introduce red phosphenes would also be sufficient for Mary coming to learn what it is like to see red.

The argument just presented assumes an externalist account of concepts. However, we can ascribe mental states and concepts to zombies only if we presuppose externalism. Unconscious mental states can be said to have content only if they give rise to certain kinds of behavior. You can be an internalist about conscious states but not about unconscious states. Without an external element, unconscious states would be pure potentiations or neuron firings without content.

An analogy: Consider the difference between stored memory and memory retrieval in one of us. Memory storage is just a pure potential of a group of neurons to interact. Memory retrieval is the generation of a mental state with a phenomenal character. When we talk about "stored" memories, we are referring to the potential of the brain to general conscious mental states during memory retrieval. If memory retrieval didn't give rise to conscious mental states, we

would have to interpret “stored” memories in terms of the behaviors they would give rise to during retrieval.

Something similar applies to zombies. We can attribute mental states and concepts to them only in relation to the behaviors to which these states or concepts give rise. Mary can be said to possess the concept of circular-shaped, only if she would be able to reach to and grasp both tilted and non-tilted circular-shaped objects. But this ability requires having interacted with tilted circular-shaped objects.

The upshot is that while consciousness necessarily involves subjectivity, perspectivality is not exclusive to conscious experience. The hard problem is not that of explaining perspectivality but rather that of explaining the qualitative aspect of experience.

Once we realize that subjectivity isn’t a problem of consciousness, there is no combination problem, as originally stated. Of course, there is still a problem of explaining how small particles or strings add up to qualitative character. But the idea of a new thing (weakly) emerging out of many things put together is not completely outlandish (Strawson 2006). Put together enough H_2O molecules under the right conditions and a liquid emerges. Put together enough minerals under the right conditions and a rock emerges. In these cases, the appearance of a novel entity can be explained by the formation of chemical bonds among molecules.

But, it may be said, if consciousness can somehow emerge from many small vibrating strings the way that a liquid can emerge from water molecules, is there any reason to postulate its existence at the fundamental level? There is. The reason, I maintain, is that the particular modes of vibration are required for the right sort of bonds to form among strings. Just like gravity cannot emerge from strings vibrating in the electromagnetic frequency, so consciousness cannot emerge from strings vibrating in the ways that we observe as elementary forces and particles.

6. The Quantum Consciousness Model

Though I am in no way committed to the string consciousness model, I do want to point to some advantages of this type of model compared to other equal-spirited models.

A fascinating model of what consciousness is is the quantum consciousness model of the brain proposed by Stuart Hameroff and Sir Roger Penrose. (1996a, 1996b, Hameroff 1998) On the quantum model, consciousness originates in the microtubules of the neurons. Inside microtubules quantum computations that take place during superpositional coherences convert pre-conscious possibilities into conscious episodes. Each conscious episode corresponds to a quantum state reduction, or collapse of the wave function. Hameroff provides the following illustration. If your friend Liz suddenly shows up in front of you, the visual stimulus gives rise to quantum computations inside the microtubles. The quantum computations lay out the different possibilities of who this person could be: {Liz, Alice, Amy}. These computations are not

conscious to you. Consciousness arises only once the wave function collapses. At this conscious moment, a choice has been made.

It may seem initially implausible for macroscopic quantum states to occur inside biological organisms. They simply would not be sufficiently isolated from the biological surrounding to continue to exist. There seems to be too much interference from other molecules and ions. Hameroff and Penrose are aware of this issue. They propose that microtubules provide a suitable climate for quantum states to persist. The quantum events, they say, are contained in hollow microtubule cores or intra-protein hydrophobic pockets.

The quantum consciousness model has received its fair share of criticism. In a response paper, Mark Tegmark (2000) argues that the phenomena described by the quantum model aren't feasible. He admits that superpositional coherences required for quantum computation could exist inside neural microtubules but, he argues, the states would be too short to control information processing in the brain's neurons. A superpositional coherence in a microtobule would last 10^{-13} sec. Neuron firing takes several milliseconds.

Even if Tegmark is wrong, I believe there is a further problem with the quantum model. The correlates of consciousness are most likely not single neurons. The authors propose that quantum states can spread macroscopically throughout the brain via quantum tunneling at gap junctions, that is, the tight connections between neurons that allow molecules and ions to pass from one neuron to the next. Groups of neurons connected via gap junctions fire synchronously. This sort of synchronous firing is a feature of neurons that correlates with consciousness. (Crick and Coch 1996)

However, I believe a challenge remains. The brain contains lots of zombie-like areas that do not correlate with consciousness. Neurons in the dorsal stream, for example, never correlate with conscious experience. Some of dorsal stream areas are involved in calculating how best to perform actions. If quantum computations are involved in neural decision-making in areas of the brain that correlate with consciousness, then it seems that quantum computations should also be involved in areas that do not correlate with consciousness. But the question then arises why collapses of the wave function and quantum tunneling in the zombie brain regions do not lead to consciousness. There certainly are sufficient gap junctions in the dorsal stream to allow for the tight networks that might facilitate quantum tunneling.

Models that allow for elementary consciousness do not face the same problem. If the string consciousness model is true, the answer to why some brain regions correlate with consciousness and others don't is simply that the vibrational modes of the underlying strings differ between different brain regions.

7. Non-Reductive Neuroscience

Where does the story told so far leave our search for neural correlates? Well, let us first take a closer look at the definition of 'neural correlate'. Though we often think of a neural correlate as something that is required for consciousness, we cannot take a neural correlate of awareness of a particular feature to be whatever is required for awareness of that feature.[6] Lots of neurons in areas buried deep inside the brain's subcortical matter, in thalamus and the brain stem are no doubt required for consciousness of any feature whatsoever. Aware of these kinds of problems, Chalmers (2002) has proposed the following definition of a neural correlate of consciousness.

> Neural Correlate of Consciousness
> A neural correlate of consciousness is a minimal neural system N such that there is a mapping from states of N to states of consciousness, where a given state of N is (nomically) sufficient, under conditions C, for the corresponding state of consciousness.

Conditions C are to be understood as conditions involving 'normal brain functioning, allowing unusual inputs and limited brain stimulation, but not lesions or other changes in architecture'. It is important that 'sufficient' is understood as 'nomically sufficient'. A is nomically sufficient for B just in case, in any situation that obeys our laws of nature, B is present whenever A is. There could be a single neuron that just always fires when you see me, a freak coincidence. This is certainly not unthinkable. If 'neural state N is sufficient for state of consciousness S' were to be understood as a material conditional in which the given state of N is the antecedent and the state of consciousness is the consequent, then this kind of neuron would count as a neural correlate of your visual perception of me. However, the neuron wouldn't be nomically sufficient for the state of consciousness.

Chalmers' definition does not suggest that states of consciousness must have a neural basis or that there can't be neural systems not associated with consciousness. If the string consciousness model is correct, elementary consciousness presumably can add up to full-blown consciousness in non-neural systems. Likewise, there are neural correlates in which elementary consciousness does not add up to consciousness.

But if neural correlates are neither necessary nor sufficient for consciousness, the question may arise why we should be interested in searching for neural correlates of consciousness.

I think that there are many reasons why our continued search is a worthwhile enterprise. Identifying the neural areas that manifest consciousness in our world, just like studying evolutionary patterns that could have been very different, is of fundamental interest. In their paper "Is the Brain a Quantum Computer", Litt, et al. argue that 'explaining brain function by appeal to quantum mechanics is akin to explaining bird flight by appeal to atomic bonding characteristics' (Litt, et al. 2006). Their criticism is not meant to be a criticism of appeals to quantum processes in explaining consciousness per se. Rather, it is meant to be a criticism of

[6]Miller (2007) distinguishes between (nomic) constitution and correlation. Constitution is also what some thinkers call the 'true correlate of consciousness'. It is the notion of constitution or true correlate that I am referring to here.

appeals to microscopic events as an explanation of macroscopic phenomena. Though what counts as a satisfying explanation is context-dependent, macroscopic phenomena often are not adequately explained by appeal to microscopic phenomena. If I tell you my house burned down, and you ask me how the fire occurred, it would normally suffice to reply with 'There was a shortcircuit'. I don't need to provide a story about microphysical events. In fact, it would ordinarily be highly inappropriate if I did.

A non-reductive neuroscience can serve the purpose of providing adequate explanations of many of our questions about consciousness. It will not give us an explanation of what consciousness is, nor will it solve the hard problem. But it can nonetheless provide answers to interesting questions, such as 'Why do my visual experiences alternate in binocular rivalry?', 'Why don't blindsighters have conscious visual experience?' or 'Why don't human beings have conscious experiences of ultraviolet?'

Our search for neural correlates may also be of interest from the point of view of intervention. If we know which neurons manifest different kinds of consciousness, we may be able to generate altered states of consciousness in controlled ways. In fact, we already do generate altered states of consciousness using techniques such as transcranial magnetic stimulation, or TMS. Using TMS we are in a position to alter visual consciousness, for example suppress it or enhance it. But we are in a position to do this only insofar as we have identified a particular region as the neural correlate of visual consciousness or at least a neural region that transmits to a neural correlate of consciousness.

References

Balog, K. 2012. "In Defense of the Phenomenal Concept Strategy", Philosophy and Phenomenological Research 84.
Bigelow, J. and Pargetter, R. (1990). "Acquaintance with Qualia", Theoria 61: 129-147.
Brogaard, B. (2009). "What Mary Did Yesterday: Reflections on Knowledge-wh", Philosophy and Phenomenological Research 78: 439-467.
Brogaard, B. (2011). "Knowledge-How: A Unified Account", Knowing How: Essays on Knowledge, Mind, and Action, J. Bengson and M. Moffett eds., Oxford: Oxford University Press (2011): 136-160.
Carruthers, P. (2000). *Phenomenal Consciousness*. Cambridge University Press.
Carruthers, P. (2005). *Consciousness: essays from a higher-order perspective*. Oxford University Press.
Chalmers, D. (1995). "Facing Up to the Problem of Consciousness", *Journal of Consciousness Studies* 2: 200-19
Chalmers, D.J. (1996). *The Conscious Mind*, Oxford: Oxford University Press
Chalmers, D. (2002) "What is a neural correlate of consciousness?" In Neural Correlates of Consciousness: Empirical and Conceptual Questions (Metzinger, T., ed.), MIT Press
Chalmers, D.J. (2004a). "The Representational Character of Experience", The Future for Philosophy, ed. B. Leiter, Oxford: Oxford University Press: 153-81.

Chalmers, D.J. (2004b). “Phenomenal Concepts and the Knowledge Argument”, in P. Ludlow, D.
Stoljar and Y. Nagasawa, ed. There’s Something about Mary: Essays on Phenomenal Consciousness and Frank Jackson’s Knowledge Argument, Cambridge: MIT Press, 2004.
Chalmers, D. (2006). “Phenomenal concepts and the explanatory gap”, in Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism, eds. T. Alter and S. Walter (Oxford University Press).
Chalmers, D.J. (2009). “The Two-Dimensional Argument against Materialism”, Walter, S., Beckermann, A., McLaughlin, B.P., *The Oxford Handbook of Philosophy of Mind*. Oxford: Oxford University Press.
Conee, E. 1985. “Physicalism and Phenomenal Properties”, Philosophical Quarterly 35: 296-302.
Conee, E. 1994. “Phenomenal Knowledge”, Australasian Journal of Philosophy 72: 136-150.
Crick, F. C.,& Koch, C. (1990). Towards a neurobiological theory of consciousness. Seminars in the neurosciences, 2, 263–275.
Hameroff, S. (1998). “Quantum Computation in Brain Microtubules? The Penrose-Hameroff “Orch OR” Model of Consciousness” Philosophical Transactions of the Royal Society of London A, 356, 1869–1896.
Hameroff, S.R., and Penrose, R., (1996a) “Orchestrated Reduction of Quantum Coherence in Brain Microtubules: A Model for Consciousness. In: Toward a Science of Consciousness - The First Tucson Discussions and Debates, S.R. Hameroff, A. Kaszniak and A.C. Scott (eds.), MIT Press, Cambridge, MA.
Hameroff, S.R., and Penrose, R. (1996b) Conscious events as orchestrated spacetime selections. *Journal of Consciousness Studies* 3: 36-53.
Hellie, B. 2007. “ ‘There’s something it’s like’ and the Structure of Consciousness”, Philosophical Review 116: 441-463.
Horgan, T. 1984. “Jackson on Physical Information and Qualia”, Philosophical Quarterly 32, 127-136.
Jackson, F. 1982. “Epiphenomenal Qualia”, Philosophical Quarterly 32: 127-136.
Jackson, F. 1986. “What Mary Didn’t Know”, Journal of Philosophy 83: 291-295.
James, W. (1890/1950). *The Principles of Psychology*, v. 1, New York: Henry Holt and Co.
Levine, J. (1983). “Materialism and Qualia : The Explanatory Gap," *Pacific Philosophical Quarterly* 64: 354-361.
Lewis, D., 1983, Postscript to “Mad Pain and Martian Pain”, In Philosophical Papers, Vol.1, Oxford: Oxford University Press.
Lewis, D. 1988. “What Experience Teaches”, Proceedings of the Russellian Society, University of Sydney.
Litt, A., Eliasmith, C. Kroon, F.W., Weinstein, S., Thagarda, P. (2006) Is the Brain a Quantum Computer? Cognitive Science 30: 593–603
Lormand, E. 2004. “The Explanatory Stopgap”, Philosophical Review 113: 303-57.
Lycan, W. 1996. Consciousness and Experience, Cambridge, MA: The MIT Press.
Miller, S. 2007. “On the Correlation/Constitution Distinction Problem (and Other Hard Problems) in the scientific Study of Consciousness”, Acta Neuropsychiatrica 19: 159–176

Nemirow, L. (1980). “Review of Thomas Nagel, Mortal Questions,” Philosophical Review 89, 473-477.
Nemirow, L. (1990). “Physicalism and the Cognitive Role of Acquaintance,” in W. Lycan, ed. Mind and Cognition, Blackwell: 490-499.
Stoljar, D. (2005). “Physicalism and phenomenal concepts”, Mind and Language 20: 469-494.
Strawson, G. (2006). Realistic Monism: Why Physicalism Entails Panpsychism. *Journal of Consciousness Studies* 13
Tegmark, M. (2000). “Importance of Quantum Decoherence in Brain Processes”, *Physical Review* E 61, 4194–4206.
Tye, M. (2000). Consciousness, Color and Content, MIT Press.