

[Home](#)

Consciousness, Science, and the Nature of Explanation

copyright Marion Gothier, April 1998

Abstract: Conscious experience is one of the last truly mysterious phenomena we know of. No satisfactory scientific account of phenomenal experience has been offered to date. This situation has led some to suspect that such an account may be impossible in principle. However, the arguments offered to support this suspicion seem to rest ultimately on intuition and are therefore not wholly convincing. I argue that both the supporters and the detractors of a scientific explanation of conscious experience base their positions on a triple misunderstanding of the nature of consciousness, the nature of science, and the nature of explanation. Once those misunderstandings are corrected, the debate can be resolved.

1. The problem of consciousness

1.1 Pretheoretical description of conscious experience

Consciousness is a word with many meanings, and so every discussion about the topic has to start with a clarification of which of the many phenomena it may refer to is the one under consideration.

This paper is not concerned with consciousness in the sense of knowledge (as in 'I am conscious of the fact that Paul is an interesting man, but that doesn't mean that I want to go out with him'), or in the intentional usage of the word ('He consciously tripped the waiter so the poisoned soup wouldn't get served').

The aspect of consciousness I am concerned with is phenomenal consciousness, or conscious experience; this is the sense in which the word is used in the sentence 'I am conscious of the soft quality of this Stradivarius' tone. Conscious experience in this sense refers to the quality of experience, the feeling of what it is like to be an experiencing subject. It is that which we have such difficulty communicating when we speak about an experience to someone who hasn't encountered it yet - the taste of a gator steak, say, or the particular smell of a San Francisco Muni train.

1.2 The reductive claim

The question I am concerned with is whether conscious experience is scientifically explainable. Can we account for the existence of conscious experience within the scientific framework? Is it possible to construct a causal chain that leads from the building blocks of matter and the forces that act between them all the way up to experience?

Claims to this effect are abundant in the scientific literature on consciousness. Take for example John Searle (1990?):

'Conscious states are caused by lower level neurobiological processes in the brain and are themselves higher level features of the brain. (...) The smell of the flower, the sound of the symphony, the thoughts of theorems in Euclidian geometry - all are caused by lower level biological processes in the brain; and as far as we know, the crucial functional elements are neurons and synapses.'

In a similar spirit, Francis Crick (1994) states

'You, your joys and sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behaviour of a vast assembly of nerve cells and their associated molecules.'

I am interested in the question of whether it is possible to demonstrate that claims like these are true, using the methods of natural science. I am not directly concerned with the ontological issue of whether or not materialism is true, which I consider to be quite separate from the considerations in this paper.

2.2. Arguments against the possibility of reductive explanation

David Chalmers rocked the world of consciousness research in 1995 when he published a paper that claimed to do away with reductive explanation of consciousness (and in the process brought dualism back into the realm of the thinkable). He focused the discussion about consciousness by making a sharp distinction between the 'easy' problems of consciousness, such as working out how visual processing works and developing a model of human cognition, and the 'hard' problem of finding a scientific explanation for conscious experience itself.

Chalmers' claim is that consciousness is not reductively explainable. From this result, he goes on to propose a new theoretical framework (Naturalist Dualism) in which consciousness stands as an unreduced entity next to basic physical entities like energy or matter. Because Chalmers claims to already have solved the question this paper is concerned with, and because I think that his account is quite typical of the way the discussion is usually conducted, I want to spend a little time looking at his arguments.

Chalmers suggests three different strategies to show that the reduction of conscious experience is not possible (Chalmers 1996, pp.94 - 106):

(1) Argument from conceivability

The argument from conceivability gains its force from the idea that in order to count as complete, an explanation must be logically necessary: it can not leave room for a (logically) possible situation where the conditions it gives are instantiated, but the phenomenon it allegedly explains isn't. Chalmers claims that even a maximally complete physical description of the world leaves room for the possibility that conscious experience is absent (zombies) or that there are systematic differences in phenomenal experience (e.g. inverted spectrum scenarios). If zombie scenarios or inverted spectrum cases are logically possible, then it follows that even a maximally complete physical description of our world is incomplete, and a reductive explanation of consciousness is impossible.

(2) Argument from knowledge

Chalmers then goes on to consider two arguments from epistemology: epistemic asymmetry and the knowledge argument. The argument from epistemic asymmetry basically points to two facts, namely: first, consciousness can not be detected straightforwardly on the basis of external (physical) facts; it is fundamentally a surprising feature of the universe which we would not expect to find just on the basis of the physical data. Second, consciousness is special in that it gives a special role to the first person case - it appears to be uniquely accessible from 'the inside', and inaccessible from 'the outside'. No other phenomenon we know of has these features (there is a problem of other minds, but no problem of other heights, or other lives). Chalmers takes these observations to be evidence for the nonreducibility of consciousness.

The second epistemological argument Chalmers gives is the knowledge argument. He uses Jackson's example of Mary, a neuroscientist specializing in color perception who has been brought up in a black-and white room. Even if Mary knows everything there is to know in scientific terms about color perception, she will still gain

knowledge of a new fact when she steps outside of that black-and-white room and sees a color for the first time. No amount of reasoning from the physical facts will tell her in advance what it is like to see a color. Nagel (1974) makes a similar point: no amount of physical or biological knowledge about a bat will make it possible for us to know what it is like to be a bat; similarly, it is impossible to describe to someone who has been blind from birth what visual experience is like.

The upshot of these examples is that knowledge of a fact about experience is different from knowledge of any other fact - there is a special quality that conscious experience has, and it seems impossible to even conceive of a way this quality could be described in terms of physiology, biology, or physics. Chalmers takes this to mean that scientific theories do not and will not ever capture all the facts about consciousness - ergo, consciousness is not reductively explainable.

(3) Argument from analysis of the concept

Chalmers' final argument is that a reductive theory of consciousness would need an analysis of consciousness such that physical facts could conceivably imply facts about experience, but there is no such analysis: A functional account seems to miss the essence of conscious experience (this is what is wrong with most modern accounts of consciousness); and it is hard to see how a structural analysis might do any better (Consciousness as some kind of biochemical structure??). The fact that we do not even have an idea of how such an implication could possibly go shows that consciousness is different from all other phenomena: it is not reductively explainable.

At first sight, these arguments together make a quite a convincing case for the specialness of conscious experience. Unfortunately, however, they do not really go beyond being intuitively compelling.

In the conceivability argument, Chalmers simply asserts the possibility of zombies and inverted spectrum scenarios and sweetens the deal with appeals to intuition and obviousness ('I confess that the logical possibility of zombies seems equally obvious to me', 'it certainly seems that a coherent situation is described; I can discern no contradiction in the description', Chalmers 1996, p. 96).

As far as the knowledge argument is concerned, I believe that this is the one point in his book where his argument is obviously circular. Stating that consciousness is surprising, or unexpected just on the basis of the physical facts is just another way of saying that consciousness is not reducible to physical facts; hence all Chalmers shows is that if we assume that consciousness is not reducible, it follows that it is not reducible. Also, there is no argument for why consciousness should be in principle accessible only from the inside. Chalmers just claims that it is, and tries to bolster his case by pointing to the uniqueness of the case. But we could object that it is not surprising that consciousness should present us with some unique difficulties, given the rather unique relationship we have with it.

Finally, in the argument from the absence of analysis, Chalmers doesn't do much more than state his position and give some circumstantial evidence for it. Simply claiming that an appropriate analysis of consciousness is impossible is not going to convince anyone who isn't already on Chalmers' side.

Second and maybe more importantly, a reductionist can hold his own against all these arguments by asserting that explaining the impossibility of zombies and inverted spectra, and capturing the subjective quality of conscious experience is exactly what his research program is all about. In other words, once we have a reductive account of consciousness, this account will also tell us why zombies and inverted spectrum cases are logically incoherent, and we will have a way to capture experiential qualities in objective terms. Once we have such a theory, the reductionist might say, consciousness will cease to seem to be a surprising feature of the

universe. The fact that we do not have such an account yet is no proof that we will never be able to develop one. 1 (This point is made quite eloquently in Valerie Hardcastle's (1996) reply to Chalmers in the *Journal of Consciousness Studies*).

I don't mean to knock David Chalmers here - I think that he has done all of us a great service by focusing and clarifying the discussion about consciousness, and I think that many of his points do have merit. I just believe that in the quest for an explanation of conscious experience, he has given up too soon. His arguments against reduction don't carry any logical force in the end, and, predictably, they did not convert any of the defenders of reductionism to his view. In fact, he does concede just this point (Chalmers 1996, p. 110):

'It might be further objected that the arguments I have given consist, at bottom, in a collection of intuitions. There is certainly a sense in which all these arguments are based on intuition, but I have tried to make clear just how natural and plain these intuitions are, and how forced it is to deny them.'

To this, I can only say that it is very dangerous to rely on intuition when thinking about an issue that is so wrought with conceptual difficulties and logical conundrums as the problem of consciousness. I believe that we can do better, and this paper is my attempt to do so.

2.5 Summary

The arguments against reduction gain their credibility from the intuition that there is something fundamentally wrong with connecting experience with the kinds of entities that science deals with; there is a sense of absurdity that surrounds the reductionist claim that is even felt by the defenders of the scientific explainability of experience (hence eliminativism, or attempts to deny that what Mary learns actually is knowledge). But there doesn't seem to be a clear sense of what it is that makes this claim incoherent. Chalmers comes close when he says that experience is not analyzable; but he doesn't pursue this claim.

We end up with a kind of trench war: on the one side are the defenders of a scientific explainability of experience, who rely on their faith that someone will come up with an analysis of experience that will make the step from physics to consciousness natural and intuitive; on the other side the detractors, whose arguments in the end come down to the intuition that such an analysis is impossible.

I believe that the entire issue rests on a fundamental confusion about the nature of scientific explanation on both sides of the discussion. Once this incorrect idea is replaced with one that more accurately reflects reality, the argument can be resolved.

3. Foundations

3.1 The nature of science

The model of scientific explanation I want to take as the starting point of my discussion is the one laid out in Karl Popper's 'The Logic of Scientific Discovery' (1959). Science proceeds by the two-step process of hypothesis and testing: the scientist concerned with a particular phenomenon will conceive of a possible explanation for it (the hypothesis), and then derive predictions from this account, which are tested against reality. As long as a theory's predictions agree with the results of the tests, the theory stands; any disagreement between the result of a test and the predictions will lead to a modification of the theory, or to its replacement. In this way, or so it is hoped, we are able to inch closer and closer to an accurate model of reality.

In a nutshell, this is the scientific method. In spite of its simplicity, it is extraordinarily powerful. The knowledge we have gathered with its help has helped us cure diseases, put astronauts on the moon, and transform our environment and our way of living to an incredible extent. Its success has been so overwhelming that many ordinary people, and more than a few scientists, regard scientific knowledge as the only valid knowledge, and the scientific model of the world as in some way more real than the world of everyday experience.

I believe that we have to be careful in making such claims. As we will see below, a large part of the confusion surrounding the consciousness debate stems from this mistaken idea of what science is and what it can do. In order to cut through this confusion, we need to be absolutely sure which of the ideas about science we entertain are based on fact and which are based on conjecture or wishful thinking.

So what can we say about scientific method as described above? The first observation I wish to make is that it is a method for gathering accurate working knowledge (not a method for gathering absolute truth, as some TV-science programs may lead us to believe). The second observation is that scientific knowledge is systematic knowledge about how things work, not about what they are. Science does not talk about ontology, and it is a fallacy to take scientific knowledge as proof for any kind of ontological system. It is easy to see why when you consider brain-in-a-vat scenarios: if we were all brains in vats (with the appropriate machinery to give us consistent sense-impressions), we could still have science; we could still build theories and perform experiments; but the scientific knowledge we would acquire, although useful as a systematic way for predicting sense-impressions, would not describe reality.

I'm not suggesting that we are actually brains in vats, or anything of that nature. What I am trying to say is that, strictly speaking, we can not make the step from scientific knowledge to ontological truth. The materialist argument that because we have never found anything that doesn't fit within the scientific framework there is nothing else is a logical fallacy. The scientist-in-a-vat could say the same thing and be pitifully mistaken. Science does not imply materialism in any way (and neither would science be impossible in a non-materialist world, as long as that world has lawful regularities in it).

Too much of the discussion about the scientific explanation of consciousness centers around ontology - what is consciousness? In what form can it be said to exist? without once stopping to realize that science has nothing to say about ontology. Science is a method for gathering accurate knowledge about the functional and structural properties of the world around us, nothing more and nothing less. With respect to the ontological question, science is silent.

We will see that much of the confusion around the consciousness question dissipates when we settle down to realize that science does not and can not answer the question of what things are. Science only describes how things work.

Before we can see how this applies to the hard problem, we need to take a closer look at consciousness itself, and the proper way of describing it.

3.2 The nature of consciousness

The model of consciousness I wish to employ is that proposed by Max Velmans (1995). Velmans has pointed out that most people, when talking about conscious experience and its relation to physical reality, make the mistake of assuming that consciousness is a phenomenon in the head, and physical reality is somehow 'out there', outside of ourselves. But if we think about the matter a bit more carefully, we can see that this is not an accurate description of how things are. Most perceptual experiences are not in the head, but 'out there', in the

'external' world: when I see a flower in front of me, the visual experience of that flower is located in front of me, where the flower is, not somewhere inside my head. A pain in my foot is located in my foot, not in my head. In Velmans' terminology, we live in the world-as-perceived, and the world-as-perceived is constituted by conscious experience.

There is nothing mystical or perplexing about this claim. What it amounts to in familiar scientific terms is that the human cognitive system projects experience out into the world, to approximately the location from where it judges a stimulus to have originated from. Velmans has amassed a good amount of psychological evidence for this projection mechanism (see Velmans (1996)).

However, looking at things this way has consequences for where we make the cut between conscious experience and physical reality. Often, this cut is made vertically, somewhere beneath the skin - anything inside is consciousness, anything outside is physical reality. Velmans' model shows that the cut should instead be 'horizontal': The things we find around us - tables, flowers, sunshine, the stars, computers - are essentially conscious experiences. We exist in the world-as-perceived, and it is only through this world that we have access to the underlying physical reality.

It must be stressed that this model implies in no way that the underlying physical reality does not exist. Velmans is quite clear about his position that the scientific evidence pointing to the existence of physical reality is much too extensive to allow any reasonable doubt. Neither should it be misunderstood as implying any kind of dualism. If conscious experience can be shown to arise from physical reality, that account won't falsify Velman's model.

What this model really does is make an epistemological point: we have direct access only to the world-as-perceived. There is no such thing as direct knowledge of the physical world. Our quest for knowledge starts and ends in the world-as-perceived.

3.3 The nature of explanation

Since the main arguments of the anti-reductionists all rests on the intuition that a explanation of conscious experience in terms of physical reality is somehow incoherent or unimaginable, and because the claim that consciousness is reducible is itself a claim about explanation, it is prudent to take a look at the nature explanation itself.

In 1994, Eleanor Rosch published an interesting paper in the Journal of Consciousness Studies dealing with the psychology of explanation (JCS 1, Vol 1,pp 50 - 65). I will follow her account here, but of course my exposition is much too short to do her work justice.

Rosch takes as a framework the logical treatment of causality in the Buddhist Madhyamika (a branch of Buddhism particularly concerned with the nature of mind), which she argues accurately reflects the ways in which we can see events to be connected to each other. The Madhyamika gives four possibilities for the arising (causation) of a thing:

- (1) at thing arises from itself
- (2) a thing arises from something that is different
- (3) a thing arises from both
- (4) it arises from nothing

Examples of type 1 explanations (that involve something arising from itself) are deductions in logical and mathematical formal systems (where in a sense the theorems are already contained in the axioms plus rules of deduction); mechanical causality (as when a ball bumps into another ball, which then moves away: here movement gives rise to movement); and folk explanations that involve the manifestation of some essence (he doesn't lie because he is honest', 'opium induces sleep because of its dormitive power').

The problem with type-1 reasoning, of course, is that it is tautological. Explanations of this kind add nothing to what we already know. They are simply restatements of themselves in other words. For this reason, science does not accept explanations of this kind as valid scientific explanations. Scientific explanations are always type 2 or 4 explanations (I'm leaving 3 out because it is just a combination of 1 and 2; this doesn't change the gist of the argument)

Examples of type 2 explanations (that involve something arising from something that is different) would be scientific explanations such as that heat is motion of molecules, light is a stream of particles, or illness is caused by bacteria, as well as religious explanations where the world is explained in terms of an invisible 'other' reality involving gods, angels, souls, and so on.

Type 4 explanations (where something arises from nothing) are chance events or events that are governed by probability - such as winning the lottery and other cases of good and bad luck. Scientific accounts also employ explanations of this form, e.g. in Roger Penrose's virtual particles that spontaneously pop in and out of existence, or in the probabilistic nature of quantum theory.

Rosch argues that psychologically, the only explanations we regard as coherent and causally bound are type 1 explanations. Type 2 through 4 are psychologically unstable in the sense that they will not be perceived as coherent unless they can somehow be seen as really belonging to type 1. For example, people affected by a chance event (such as winning the lottery, or getting breast cancer) will tend to rationalize this by searching for meaning in the event (where the meaning is then seen as the ground that gave rise to the event). In the case of type-2 explanations in cognitive science, attempts to turn the scientific story into one where the outcome is contained in the ground are common. Rosch mentions, for example, work on form perception by Hinton (1979) and Marr (1982) where 'Retinal stimulation goes through a series of stages at each of which the representation looks more like the percept. In fact, 'looks like' is taken quite literally by the computationalists; we may be shown actual pictures of how the stimulus supposedly looks like at each stage so that one can see its similarity to the percept' (p. 59).

At this point, we are left with a somewhat perplexing situation: psychologically, we only accept type 1 explanations as coherent. But science explicitly rejects such explanations as tautological, and therefore deals only in type 2 and 4 explanations. In fact, we might say that scientific explanations derive much of their explanatory power from the fact that they are not obviously true in the sense that type 1 explanations are, but can be demonstrated to make valid predictions nevertheless. Science is a bit like a magic trick in this respect. In a sense, science buys explanatory power at the price of losing intuitive coherence.

In the following discussion, we will see how, taken together with Velman's model of consciousness, this result gains enough force to cut through the mystery surrounding conscious experience.

4. Scientific explanation and conscious experience

4.1 The nature of scientific explanation

As we have seen above, the traditional picture of how science proceeds is as follows:

- 1) We start with a phenomenon as our explanandum (e.g. heat, or exposed photographic paper)
- 2) We construct a scientific hypothesis to explain that phenomenon (motion of molecules, or x-rays)
- 3) We test the hypothesis for logical consistency, and deduct predictions from it (transference of motion, production of electrical charges)
- 4) the predictions are then tested in experiments (Brownian motion, Geiger counter readings).

However, if we take Velman's model into account, the story becomes a bit more complicated. Basically, Velman's model implies that the account above fails to make the distinction between the world-as-perceived and the postulated underlying reality. Scientists talk as if we had direct access to regularities in the physical world; but if we look at the situation correctly, all we can observe are regularities in the world-as-perceived. We can get away with conflating the two when we are investigating the external world, but once we start to investigate consciousness, the distinction becomes all-important.

So how can we integrate this distinction into the scientific method? Velman's model tells us that the goal of science is not to explain phenomena in the underlying reality, but to explain phenomena in the world-as-perceived i.e. in consciousness. To put it bluntly, science doesn't describe physical reality. It is instead a systematic description of regularities in experience.

This is a strong claim, and it goes against our usual picture of what science is and does; for this reason it deserves some careful justification.

First, I would like to point out that with respect to explanation (which is what I am concerned with here), science does not explain regularities in the physical world at all. It measures them. Questions such as why Newton's laws are the way they are, or why the speed of light is ~ 360000 km/sec, are considered to be meaningless, or outside the realm of science. However, the question of why a cup of hot tea is hot is considered to be answerable, and is in fact answered by a story about wiggling molecules and their interactions with their environment. Now notice that the latter question is a question about experience. The question about the nature of heat is the question of why hot things feel hot.

It might be objected that not all problems in science have this simple form. In what way, for example, can a theory about radioactivity said to be a theory about experience? We can not see, hear, feel, or taste nuclear radiation, but science nevertheless tells us that it exists. Doesn't this mean that science encompasses more than our experience?

This objection rests on a simple misunderstanding about the status of entities like x-rays, atoms, force fields, and the like in science. What is assumed is that those entities have some kind of independent, absolute existence that we can uncover using the scientific method. But we have seen above that the step from what we know to what exists in an absolute sense is not a valid step. Why do we believe in x-rays? We believe in them because they allow us to account for a phenomenon in experience, namely our visual impression of funny exposure patterns on photographic paper. X-rays are an explanatory construct that allows us to explain those visual impressions, and connect them to other otherwise unexplainable phenomena in the world-as-perceived, such as clicks in geiger counters, and clusters of cancer cases around nuclear plants.

This point is bound to be misunderstood as some kind of crazy Berkeleyan idealist position, and therefore I want to emphasize again: I am not claiming that x-rays or atoms don't exist. For all I know, they might - but as far as scientific knowledge is concerned, the discussion about existence is beside the point. The scientific method is not a tool to uncover ontological facts. All the scientific method is is a method to uncover lawful regularities in the world, and in our case the world in question is the world of experience. So the minimal, safe assertion about what scientific knowledge is is to say that it is a systematic description of regularities within experience, and that the physical 'underlying reality' is a model designed to explain these regularities.

What can we say about the nature of this model? We have seen that science excludes type-1 explanations: explaining a phenomenon in terms of itself will not yield a valid hypothesis, because such arguments are circular. Taken together with a Velmanian view of science, this noncircularity requirement leads to an interesting restriction on the class of permissible scientific hypotheses: because scientific theories explain conscious experience, they must be formulated in such a way as to exclude experiential qualities - that is, the underlying reality can not contain any phenomenal qualities.

The picture we are constructing for the scientific process now looks as follows:

- 1) We start with a phenomenon in the world-as-perceived as our explanandum (e.g. heat, or exposed photographic paper)
- 2) We construct a scientific hypothesis in terms of an underlying reality. This step crucially involves stripping the phenomenon of all its experiential qualities (motion of molecules, or x-rays)
- 3) We test the hypothesis for logical consistency, and deduct predictions from it concerning regularities in the underlying reality (transference of motion, production of electrical charges)
- 4) the predictions are translated back into predictions about regularities in the world-as-perceived and tested (Brownian motion, Geiger counter readings).

This new model introduces two new steps, namely the one in 2, where we go from the world-as-experienced to the underlying reality, and the one in 4, where we move back up. These two steps are, logically speaking, non-sequiturs, and intuitively speaking, non-obvious. The assertion that heat is the motion of molecules has nothing to do with logic (it is not possible to derive this explanation from the nature of the experience of heat), and it is, psychologically speaking, incoherent - it takes quite a bit of arguing and experimentation to convince people that it is actually true. Since neither logic nor intuition will give us a reason to believe in the validity of the steps in 2 and 4, how do we justify those steps? We do it in two ways:

- (1) We assure ourselves of the correctness of our models (contra intuition) by conducting experiments (this is the scientific method)
- (2) and we assume an isomorphism between the world-as-experienced and the underlying reality (this is the materialist hypothesis).

Of these two justifications, only the first carries any real power - and it is the one that is enshrined as the core of science in the scientific method. It is what distinguishes science from other explanatory systems that deal in underlying realities, like religious systems. The materialist hypothesis is merely a hypothesis that can not even be proven or disproven with scientific methods itself (as we can see from brain-in-a-vat scenarios). It is a crutch designed to give scientific explanations psychological power.

Again, I want to emphasize that I am not concerned here with proving materialism wrong or anything of the sort - I am simply saying that if we ask ourselves what the scientific method is designed to do, it turns out that the ontological question is outside its realm. I believe that it is the insistence of science that what it has uncovered is ontological truth, that it explains what things are (namely matter and energy and nothing else) that is ultimately responsible for the intractability of the mind-body problem and the hard problem of consciousness.

To summarize, the correct model of scientific explanation that arises from this discussion is this:

- (1) Hypothesis: a scientific explanation explains our immediate, experiential reality in terms of an underlying reality which does not contain experiential qualities. It is the fundamental difference between the two that gives scientific explanations their explanatory power (at the expense of psychological coherence).

(2) Testing: A scientific explanation then constructs causal chains within the underlying reality which causally connect phenomena that have expressions in both the world-as-perceived and the underlying reality. These causal chains are themselves entirely given in terms of the underlying reality (they are type-1 arguments), and serve as tests of the validity of the hypothesis.

4.2 The scientific explanation of consciousness

We are now ready for the question of what a scientific explanation of consciousness would look like. Recall that scientific explanations explain phenomena in terms of an underlying reality that is devoid of experiential qualities. So what do we get when we try to construct a scientific model of experience itself? The standard answer to this (given by Kripke (1980), Chalmers (1996), and numerous others), is that we are left with nothing - when you take the experience out of experience, there is nothing left. But if we take Velmans' model into account, the answer turns out to be different. Velmans' answer to the question of what a scientific explanation of experience is is that it is simply science itself. When you take the experience out of consciousness, you get the physical universe.

What, then, is the status of cognitive science? No explanation in science can be truly complete until we have told a story about how the physical event in question interacts with the human nervous system and cognitive apparatus: An explanation of heat will have to include the story about how molecules of high kinetic energy initiate a nerve signal, and how this signal is processed in the brain. But there is no distinction between an explanation of heat and an explanation of the experience of heat. They are both parts of the same story. Cognitive science is the effort to complete the scientific enterprise. It is part of the quest for an understanding of experience. Once we have constructed a complete scientific model of reality that includes the human cognitive system, the scientific enterprise as defined will be complete.

In this sense, those who claim that a model of the human cognitive system is what we need to understand conscious experience are correct. The scientific story about my experience of red is the story of light of a certain wavelength and its interaction with my cognitive system. In terms of science, there is nothing else to say. (I believe that this is the sense in which Dan Dennett's position is right; or at least this is the only perspective from which his claims make sense to me)

4.3 The nature of the mystery

If this is true, then where does the sense of mystery come from, the feeling that even a maximally complete scientific explanation of the universe would not satisfactorily explain consciousness?

Remember that the sense of satisfaction we derive from a scientific explanation rests on two things: One, the fact that we give a model of a phenomenon in terms of an underlying reality which is crucially different from the phenomenon to be explained, and two, because we connect it via a causal chain to other phenomena that seem unrelated at first sight.

Both parts are absolutely necessary for the satisfaction factor. Any explanation has to be given in terms of an underlying reality that is fundamentally different from the explanandum, because we will run into circularity otherwise. But a simple translation of some phenomenon into an underlying reality does not make for a satisfactory explanation at all - if you knew nothing about science and I just walked up to you and told you that heat is motion of molecules, without giving some further story about how this relates to other phenomena, you will have no reason to accept my theory. In fact, it will probably seem ad-hoc and quite absurd to you. We believe in scientific models because we test them - because of their ability to causally connect phenomena in the world-as-perceived which we did not expect to be connected (such as falling apples and tides, for example).

It should be clear now why the scientific explanation of experience seems so unsatisfactory. As we have seen, a scientific hypothesis - although it will strictly speaking always be a non-sequitur - can be made plausible by showing how it fits into a larger picture, and deriving (via causal chains) surprising consequences from it. However, because experience encompasses the entire world, and therefore an explanation of experience is tantamount to constructing a model of the entire world, that model can not be made plausible in the same way other scientific hypotheses are made plausible. There can not be a causal chain that connects a model of the entire world to anything else. We do not have access to a larger picture to fit it into. With respect to consciousness, we are like a tribal chief who is told by a scientist that sunlight is a stream of particles, without any further elucidation. The explanation is utterly incomprehensible without a context to put it in - and in the case of consciousness, there is no context. We do not understand why consciousness is explained by the underlying reality not because consciousness is mysterious, or because it is not physical, but because of our situation with respect to the gathering of knowledge. With respect to consciousness, we do quite literally not see the forest for the trees.

4.4 Can we understand a reductive explanation?

Maybe this explanation does not go far enough for those who insist that there must be a further truth about consciousness. They might say that psychological convincingness is not the issue - the issue is the claim (by reductionists) that the underlying reality somehow produces the world-as-perceived; that there is a causal chain that connects physical reality with the world-as-perceived.

The bone of contention is the idea that physical reality comes first in some sense, and that experience is somehow derived from it. It is the idea that nothing exists beyond the world that science describes, and therefore conscious experience must be derivable from physical entities, or understandable in physical terms.

It should be clear from the discussion above that this claim is false, not because consciousness is not physical, but because we can not assume that a physical description of consciousness will be understandable in any intuitive way. Here is a thought experiment that hopefully will make this point absolutely clear:

Consider an alien race that possesses a special sense ('a third eye') which allows them to tell which physical structures they encounter have conscious experiences and which don't. Let's say the aliens call this sense-impression gnorkness (in parallel to 'redness'). So presumably they would say that humans are gnork, but wallpaper isn't. Now we have a sense-experience, gnorkness (of course we poor humans cannot imagine what it is like to perceive gnorkness), and gnorkness singles out a set of physical objects that are gnork. A scientific theory of consciousness in such a scenario would be an account - in terms of the underlying reality - of the properties something must have to cause gnorkness (the hypothesis). This hypothesis would be tested by showing how those properties fit in with the rest of the underlying reality, in particular how they interact with the aliens' third eye to give rise to a cognitive representation of gnorkness (this is where the causal chains come into the picture). Given that gnorkness really indicates the presence of conscious experience, our aliens would have a scientific account of consciousness, causal chains and all. Now ask yourself the question:

Do the aliens *understand* their account of gnorkness/consciousness?

The answer is of course no. The aliens are in exactly the same epistemic situation that we are in: Starting from experience, they construct a model of reality to account for the regularities in their experience. In doing so, they, too, must abstract away from experience itself - logic requires that they do so. All the aliens may have is a better account of which physical structures are associated with gnorkness/consciousness than we ever have a chance to construct, because they have independent evidence as to the distribution of the phenomenon.

The misunderstanding lies in the idea that causal chains have the power to explain things in science. They don't. Hypotheses explain things, and they derive their explanatory power from their prima facie incoherence. Causal chains are the means by which we make hypotheses palatable, by showing how much explanatory power a hypothesis has. No causal chain ever constructed in science has ever done what we expect of a bridge across the explanatory gap: make a scientific hypothesis intuitively plausible.

5. Conclusion

I hope to have shown that in the discussion about consciousness, both sides are right: the scientist when he claims that what science offers as an explanation of consciousness is the only explanation there is, and the philosopher when he says that this explanation is utterly mystifying. The reasons for their misunderstanding are mostly conceptual ones: the idea that scientific explanations should be intuitively understandable; the misconception that consciousness is a small part of reality located in our heads; the idea that we can deduct facts about existence from facts about knowledge.

The problem with consciousness is not that it is mysterious and heat, say, isn't. Heat is mysterious in exactly the same way that consciousness is. *Science* is mysterious. Science deals in hypotheses that are completely out there, utterly unrelated to the phenomena they are supposed to explain - and this is necessarily so. It follows from the very nature of explanation. We reject as circular any coherent explanation - and when we are given an incoherent one, we complain that we can't understand it. We start our quest for knowledge by taking consciousness out of the picture, and then we act surprised when we can't find it in our explanations.²

I have claimed that what I have to say provides a different perspective on the question of consciousness - one from which we may be able to cut through the confusion that usually surrounds the debate. It is now time to make good on that promise. Here are my answers to some of the vexing questions in the consciousness debate.

Are Zombies logically possible? - No. There is no physical universe without consciousness, and no consciousness without the physical universe. The mistake lies in confusing knowledge with reality. When we claim that we are imagining a physical universe without consciousness, what we are really imagining is a system of knowledge without a knower - and this is not possible.

Is consciousness surprising? - Yes, but it is so not because of some mysterious property. It is surprising because in order to construct meaningful explanations, we have excluded it from science at the very beginning of the scientific enterprise.

What about the subjective quality of conscious experience? - We do not find the subjective within science for the same reason that we do not find experience itself: It was excluded from the beginning. Science by definition talks only about those aspects of consciousness that we can share. Those aspects of our inner lives that are essentially private are outside its domain.

Does Mary learn something new when she sees red for the first time? Yes, but what she learns is outside the scope of science by definition.

Is science a means to gather knowledge about reality? - No. Science is systematic knowledge about experience. The step from what we know to what exists in an absolute sense is always a fallacy.

Is there a scientific analysis of consciousness? - Yes, and we've been busy constructing and refining it for the last 400 years.

Will we intuitively understand a scientific theory of consciousness once we have it? - No.

What do you get when you take experience out of consciousness? - The physical universe.

Footnotes

1. Chalmers never really addresses this argument directly. Instead, he argues separately for the logical possibility of dissociation, and then addresses a weaker version of the materialist position (pp.131 - 134), where logical possibility is conceded:

a. The materialist position ('when we know what the scientific account for consciousness is, we will know why dissociation is logically impossible') can be restated in modal terms as saying that we may have demonstrated the logical possibility of dissociation using the primary intension of the notions involved, when we should have used the more appropriate secondary intensions.

b. But this argument is irrelevant with respect to the ontology question. Materialism will be proven wrong if we can find any property - primary or secondary - that is not instantiated when the maximal set of physical facts E is instantiated in a possible world.

c. This is just what we have established using the Zombie and inverted spectrum arguments.

d. Ergo, materialism is false.

As we can see, the materialist argument loses all its force once logical possibility is conceded. But logical possibility is of course the real bone of contention; and Chalmers' arguments for the possibility of zombies and inverted spectra are so weak (relying on intuition and stipulation), that relying on them to refute materialism doesn't really make for a good case. The core of the matter is the question of whether those scenarios are logically possible; and Chalmers doesn't answer that question.

2. I am of course not the first person to make this observation. For example, Erwin Schroedinger came to similar conclusions when he wrote that the reason that we can not find our sentient self in the world picture of science is that it is itself that world picture (quoted in Scott (1995), p.115).

Bibliography

- Chalmers, David (1996), *The Conscious Mind* (Oxford University Press)
- Hardcastle, Valerie (1996). 'The why of consciousness: A non-issue for materialists', *Journal of Consciousness Studies* 3, No.1, pp. 7-13
- Kripke, Saul (1972), *Naming and Necessity* (Harvard University Press)
- Nagel, Thomas (1974), 'What is it like to be a bat?', reprinted in Hofstadter, D. and Dennett, DC (1981) *The Mind's Eye* (Bantam), pp. 391 - 403
- Popper, Karl (1952), *The Logic of Scientific Discovery* (Routledge)
- Rosch, Eleanor (1994), 'Is Causality Circular?' ,*Journal of Consciousness Studies* 1, No. 1, pp.50 - 65
- Scott, Alwyn (1995), *Stairway to the mind*, (Springer:Copernicus)
- Searle, John (1990?), *The problem of consciousness*, *Mind* 60:1
- Velmans, Max (1995), 'The Relation of Consciousness to the Material World', *Journal of Consciousness Studies* 2, No. 3, pp. 255 - 265

Author information:

Marion Gothier, 29 A Sharon Street, San Francisco, CA 94114 Tel. 415-621-2219, e-mail: tram@sirius.com