THE NATURE OF EXPLANATION IN A THEORY OF CONSCIOUSNESS

David Bengtsson

Kungshuset, Lundagård

222 22 Lund

David.Bengtsson@fil.lu.se

INTRODUCTION

In his novel "Thinks", David Lodge lets the female protagonist, the novelist Helen Reed, close a conference in Cognitive Science with the words "Understanding Consciousness, it occurred to me this weekend, is to modern science what the Philosopher's Stone was to alchemy: the ultimate prize in the quest for knowledge." The point being that the search for the Philosopher's Stone, the substance with which to make gold out of lesser metals, promoted science in general, even though the main quest failed. The same thing, then, or so goes the argument, with the search for an explanation of consciousness. While the central problem, how subjective experience can arise from physical events, may remain as hermetically closed as ever, the attempts of solving this problem has lead, and will probably continue to lead, to amazing discoveries about consciousness and related matters. Of course, the analogy can only take us so far, and the failure of alchemy should not be thought of as an indication of the possible future success or failure of an explanation of consciousness. We should be careful when looking for parallels in the history of science (as we shall see, not all writers on the subject heed this advice). There is, though, I think, something to be learned from the analogy: The existence of explanations surrounding the central problem of conscious experience may tempt us to believe that "more of the same" would finish the job. There need, of course, be independent arguments for this. But there is also the possibility that these explanations do succeed, or are of the kind that would, without our realizing it. Some writers on the subject would probably say that these explanations already exist, that the explanations we already got are actually sufficient, and that "all further resistance is futile": Not to admit the success of these explanations, they might say, is prejudiced, stubbornly conservative or just plain silly.

In this essay I will try to flesh out the position that the main debate in the philosophy of mind should not be thought of so much as a debate about the nature of consciousness (of course it's that too) as it should as a debate about the nature of explanation. I will also claim that the nature of consciousness is such that the nature of explanation with regard to it needs to be especially carefully scrutinized. Therefore I also claim that explaining consciousness is not just a matter *within* science; it's a matter *about* science, about its very foundation. As such, it's a defence for this wonderful discipline known as Armchair Theory of Consciousness (or just "Philosophy of Mind" for short).

Augustine famously said about time that when no one asked him about it, he knew exactly what it was, but when someone did ask him about it, he didn't know. This, of course, translates beautifully into "Come to think of it, I have no idea what I'm talking about". In the theory of consciousness we are in a similar predicament, with the slight modification that in some sense we know exactly what we are talking about, but we have no idea how to explain it within a unified scientific theory. And knowledge without explanation is very disturbing to the civilized mind. The purpose of this essay is also to scrutinize the different instruments available for explanation. Dennett (1993) persuasively claimed that we can not settle for some "and then some miracle occurs" version of explanation. Even though this indeed would amount to an explanation, since miracles are famous for their remarkable causal powers, invoking them would be scarcely different from giving up. Another explanatory instrument is the inference of laws, in this case *psycho-physical* laws, and it seems to some that all we could ever discover, concerning the relation between brain-states and consciousness, is law-like correlations. Does the invocation of laws plus the addition of consciousness as a distinct realm of reality add up to giving up? This is an important test of our intuitions. In some cases, addition may give explanation. After all, if there is such a realm, explanation without it would be incorrect or incomplete at best.

The essay is divided into two parts, one picking out what problem of consciousness we are concerned with, and the other dealing with the role played by the nature of explanation in explicating this problem. I

Lund University Cognitive Studies - LUCS 106 ISSN 1101 - 8453

come to the conclusion that the requirements that we put on an explanation may be motivated by a prejudged conception of what the explanandum is. I argue that the nature of explanation is not, or at least not obviously, topic-neutral. I therefore suggest that a "scientific" explanation of consciousness may leave us with an intact philosophical problem, since the standards for scientific and philosophical explanation may differ.

In the conclusion I draw some rather farfetched conclusions that may earn the epithet "speculative" about the limits of scientific explanation and the necessity for philosophical scrutiny for assessing these limits. It's more or less an act of self-defence on behalf of the essay itself, giving reasons for why the literature to which it belongs is a proper study for the scientific mind. The importance of the arguments that I give in the essay is not, I believe, restricted to the foundations it provides for these speculations. I therefore invite the reader to make such use for them that he or she finds appropriate.

PART I: SPLITTING THE REFERENCE

Narrowing it down

Consciousness is a mongrel notion. That is, there is not just one problem of consciousness, because there is not just one phenomena of consciousness. All the phenomena of consciousness are in need of explanation and many of them are to some extent already explained, or about to be. David Chalmers (1995) listed the "easy problems" of consciousness by which he meant the questions about how the brain does things, how certain functions are performed (how we discriminate between stimuli, how we report our mental states etc.), and contrasted these with the "hard problem" of consciousness which is really the problem with which we are concerned here. Ned Block made a substantive distinction between what he called "access-consciousness" and "phenomenal consciousness" thus singling out the hard problem from the likewise hard, but at least conceivably solvable problems. In his highly influential and frighteningly titled 1982 article "Epiphenomenal Qualia" Frank Jackson, then a dualist of sorts, presented us with some of the most thought-provoking arguments yet to have risen out of the extensive literature of the philosophy of mind. He there asked us what to make of the fact that even if we knew everything there was to know about what happened, when we experience a thing, in physical terms (in his case, colour), wouldn't there still be something left that we *didn't* know, namely what it was like to experience it? And, of course, Thomas Nagel's more compellingly named article "What is it like to be a bat" (1974)[1] put the matter admirably. The problem of consciousness, with which we and the present company are concerned, is the question about the phenomenal aspect of consciousness, the "what's it like"-ness of conscious experience. Admittedly, this question is usually the one that pops into mind when someone mentions the problem of consciousness, and a vast majority of the literature dealing with consciousness as something problematic takes this aspect. But still there is, I believe, reason to scrutinize the matter further, and to single out the core of the problem. This part of the essay is mostly about pointing out exactly what part of consciousness that poses a specific problem for explanation. That is, what the *explanandum* is.

The problematic features of conscious experience are often referred to with the philosophical term *Qualia*. Qualia (sing. *quale*) are our experiences, not what we experience. It is important to understand this early on, since much of the confusion in this matter turns on the point that it is the *content* of conscious experience that poses the big problem. In the literature, this has to do with the emphasis on the *privileged access* we are supposed to have to our own consciousness. Obviously, it is easy to interpret this claim as if we had privileged access to what our conscious experience is *about*. The thing is that it is this access itself, and not what it is about, that is problematic. I will make the perhaps controversial claim that the subject matter of our conscious experiences is quite irrelevant[2]. Qualia are our direct contact with what goes on in the head (and, to some extent, our most direct contact with what goes on in the world). The fact, noticed by Nagel and Jackson, is that we are in some sense directly aware of some of the things going on in the head. This directness is to be contrasted with *mediated* awareness, in the sense that I am aware of the fact that my brain consists of two roughly similar hemispheres etc. But it is important to understand what this claim actually amounts to.

It has often been noted, and some have even based a substantive philosophical argument upon the fact that the content of our conscious experiences are not brain-states[3]. But no interpretations, as far as I have seen, have been given for what conscious experiences are about instead. The reason, I believe, that these experiences are not about brain-states is that content is determined by interpretation, and we simply have not learned to identify brain-states in this way, or, to put it the other way around, to interpret our conscious experiences as being about brain-states. The

[1] For some amusing, though not very seriously meant, answers to this question see David Lodge's theory-laden novel "Thinks".

[2] In this respect I side with Paul Churchland (1984) and Richard Rorty (1980). See below.

[3] Jackson (1982) has often been interpreted this way, for instance. But the strength of his argument, I believe, surfaces even more in the other interpretation.

important thing to note, then, is that the problem of consciousness with which we are concerned here is not the problem of the content of conscious experience, but of the *fact* of conscious experience, that is, the *qualia*.

The perhaps most common example of a quale is seeing red, even though it is highly uncertain whether we ever have a simple experience like this[4]. Tasting wine, listening to music, the feel of being touched, all these are phenomenal states and the mystery to be solved is why there should be anything like this at all. It is important to note that nothing forces us to choose only "simple perceptions" as qualia. The interpretation and higher processing of incoming stimuli does not "distort" our access to qualia, since qualia just are this access. Nothing, this is the big point about them, could distort qualia. What could be distorted is the memory of qualia, the connections between qualia, and the interpretation of qualia. Seeing the brain function with the guidance of modern neuroscience the question raises itself: couldn't all this have been accomplished without this peculiar phenomenal feature? Working our way through the layers of the brain leaves it absolutely incomprehensible where such a thing as phenomenal consciousness could take place. But then again, what did we expect? The problem is that we have some concept of what it would take for the brain to accomplish certain tasks, to implement certain systems and perform certain functions, while being left quite in the dark as to how it would accomplish phenomenal consciousness. Indeed, we lack a conceptualization of how *anything* could accomplish phenomenal consciousness. Or at least, some of us do. Daniel Dennett (1993, for instance) claimed that function is all there is that is in need of explanation and that our propensity to make judgments about our conscious experiences are just one more function to explain[5]. It seems quite possible that all these functions could be performed and consciousness still be lacking. The phenomenon with which we are concerned, then, is this, presumably non-functional property which we ascribe to people when we think they are conscious in the same way as we our selves are. Watching them behave, asking them for clues and scanning their brains will serve as very strong indications of them being so. Indeed, it's the *only* indication that could possible do the job, supposing that phenomenal consciousness itself is non-functional and non-causal (and serving as indication is performing a function – this, as we shall see, leads to problems about first-person reports: should we take them at face value, and can qualia be said to cause our beliefs about ourselves being conscious?). But that does not mean that these functions are what we refer to when we talk about consciousness. What we refer to (in the problematic cases), hence the title of this part of the essay, are the *Qualia* and the question is whether they are numerically identical to the implementations of certain functions, if they are co-extensional with them, or if they are straightforwardly distinct. We will now move on to investigate what these options might mean.

Are the problems distinct?

What is the difference between the hard problem and the easy problems? Why couldn't more of the same techniques that have proven themselves so useful for explaining the more local phenomena succeed in solving the hard problem? Some aspects of consciousness indeed work with a "divide and conquer" strategy (Gazzaniga et al. 2002), so why should not the explanatory project proceed in the same fashion? This seems to be the argument put forward by the functionalists and the eliminative materialists. The first group is championed by Daniel Dennett, the second by Paul and Patricia Churchland. Dennett's model, the "multiple drafts" model[6], is presented as a counterpart to the Cartesian Theatre model. This is a rather peculiar argument, considering that few, if any, hold such a view (the Cartesian) of consciousness. The main thinkers of the "Qualia-freaks," Searle, Jackson, Chalmers, Nagel, are as devoted anti-Cartesian as anyone. Dennett argues (1993, following the tradition from Ryle) that there is no discrete region in the brain where consciousness "takes place", and this much seems plausible, and is, I think, acknowledged by most writers on the subject. If consciousness is to be identified with anything in the brain, it seems plausible to claim, it should be with some *global* state of the brain (see Baars (1996); Crick and Koch (2003)). The neural correlate theories presented today work rather with functional states than with the firing of individual neurons, or discrete regions in the brain. Visual perceptions, for instance, depend on the work of many distinct areas, and

[4] Peter Gärdenfors has pointed out to me (in private conversation) that most philosophers seem to be preoccupied with experiential *atoms* like "the simple perception of red". Naturally, if we are ever to reach a proper correlation between brain states and qualia (as opposed to a general solution to the mind-body problem), the qualia part has to be more naturalistically conceived of.

[5] For some years know, Dennett has been involved in what we could call the Battle of Amazement. What is it that is so amazing about consciousness? Is it how "it" does things, or is it its phenomenal character? Both are quite amazing, actually, but the first is at least conceivable to explain, whereas the latter seems like a harder nut to crack. Dennett's argument here is that we do not realise vividly enough just how amazing the functional features of the brain are, and if we did, we would see how they could make it appear as if there were an irreducible phenomenal feature of consciousness.

[6] Dennett (1993) This "pandemonium model" try to establish that the things that we believe are done by a single, phenomenal consciousness is in fact done by a set of unconscious entities, what one could call "local zombies", and that their joint work could amount to (apparent) phenomenal consciousness. This argument turns on the interesting point that our concept of functions usually are *serial*, and we can't see how *such* a function could accomplish phenomenal consciousness, and that our failure is due to our troubles conceptualizing a set of parallel working functional entities.

nothing is found when we follow the neural activation into the dark pathways of the brain that immediately presents itself as a likely candidate for the correlate of consciousness. There is no endpoint of processing where consciousness "takes place." At least, nothing at present speaks in favour of this view.

Seeming as believing

Dennett (1993, 1988) seems to argue that all that needs to be explained when it comes to phenomenal consciousness is the fact that we report conscious experiences. And this, he claims, is easily done. It *seems* as if we have conscious, irreducible experiences, but this only means that we have a propensity for making judgments of a certain kind. And, Dennett claims, judging something to be the case is not a sufficient argument for this really to be the case. Even if our judgments to this effect are sincere, this only shows that we really believe that we have experiences of this kind, and just sincere belief is not sufficient either. Chalmers (1996, p 191) presents a, to my opinion, rather conclusive argument against this, and thus against the core argument of "Consciousness Explained": There are two distinct senses of "seeming": one "phenomenal" and one "psychological." The psychological sense of "seeming" is our being disposed to judge a thing to be the case, the phenomenal sense is just to *experience* in a certain way, and clearly these are distinct. What seems to be the case in this *phenomenal* sense is not a strong indication that it really is the case, but it is a conclusive indication on that this kind of seeming takes place – the analogy is with me reporting my belief: it is a strong argument, not for my belief but for my having this belief[7]. Obviously, I can judge that something seems a certain way, and indeed, this seems to be the only external indication that something seems to me in a certain way, but that does not mean that there is no independent fact of the matter; the fact that I have a certain conscious experience. Without it, my claim would be false. Seeming as disposition to judge and seeming as experiencing in a certain way, are distinct. My saying that something seems to me to be red is obviously not a conclusive argument that that thing is red, neither is it a conclusive indication that I have the experience of red. But having the experience of red is not a conclusive indication that the thing experienced is red either. The seeming in each case is a *non sequitur* for the content of the seeming, but it is conclusive for the seeming itself. The disposition to judge that something is red is obviously not co-existent with having the experience itself (I may have reason to lie or just to keep my mouth shut) which should be sufficient to accept their distinctness. The main argument of Chalmers' *The Conscious Mind* relies on this dichotomy between the phenomenal and the psychological, where the psychological is understood as the functions usually thought to be performed by consciousness. With every phenomenal phenomena there seems to be some associated psychological mechanism. Usually, when such striking parallelism is at hand, one should be open to the possibility of identity. One of the questions one might ask is if anything is done by the one that could not equally be done by the other and in our case this strategy provides further evidence for identity. But then again, phenomenal consciousness is not characterized by its functions (even though we sometimes refer to it as if it were), but by what it is like to have it. It may be, and it does seem likely, that qualia is what it is like to be in certain psychological states, but that does not amount to saying that they are identical, since it is exactly the fact that there is something it is like to be in a certain state that is in need of explanation. Indeed, it seems as if this very fact is what justifies us in believing that they are distinct.

The distinctness of the problems, then, goes something like this: In the "local" (psychological) cases, we can settle for functions, the "local"-ness is an effect of leaving the hard problem out. For every distinct local problem the question can always be raised again: why should the performance of this function be accompanied by conscious experience? Why should the processing of visually presented stimuli amount to visual experience? When we address the local psychological problems we are explaining what we experience, but we are not explaining the experiences themselves. The question, of course, is: can this latter project be successful? And, more specifically: in what sense can it be successful?

The problem of reference

Are we Augustinians when it comes to consciousness? Do we know what we are talking about? To what do we refer when we speak of our conscious experiences? This part is baptized "splitting the reference" because I believe that the fixing of reference is a matter of some importance when dealing with explanatory issues of consciousness. The classical formulation of this problem (Putnam and Kripke, but see also Chalmers (1996) and Jackson (1998)[8]) deals with water. Water, we have strong reason to believe, is H_2O. "Water" refers to H_2O, because water is known as "the watery stuff around" and the watery stuff around is H_2O. Now, according to Kripke, reference is fixed by rigid designators. In that sense, "water" refers to H_2O everywhere, in all possible worlds, even in worlds where H_2O is not a watery substance. This is a useful notion, because we can now say that if something else happened to show

[7] This, of course, turns on the argument presented above, that it is not the content that matters.

[8] For an extensive treatment of what I say below, see Frank Jackson (1998) and his discussion about C- and A-extensions.

up that looked like and behaved like water the discovery of it's failing to be H_2O would disqualify it as water. The other possibility is that water refers to "the watery stuff" everywhere. In this sense, wherever we come across watery stuff, it is water. We are dealing with two distinct intensions, and, really, two, although overlapping, extensions as well. Now compare this with the problem of consciousness. What makes it true that I am in a certain phenomenal state is the existence of some psychological (physical, functional) state. In our world "this phenomenal state" picks out a certain physical mechanism. But it does not follow that in every world where this physical mechanism is present there is a corresponding phenomenal state. We need to "split the reference" in accordance with these distinct intensions, and as long as the intensions do not pick out exactly the same things in every possible world, we cannot assume psycho-physical identity. It may be strange to claim that phenomenal consciousness could be present without *any* physical substrate underlying it, but it is not necessary to claim that they could: all we need to claim is that there is no necessity in *this* physical/functional state underlying phenomenal consciousness. Thus, the relation is the one of *supervenience*, not *identity*. And, of course, if there *could* be the physical correlate without the phenomenal event, then something needs to be added to *our* world in order to account for the fact that this correlate never occurs without the phenomenal accompaniment. This is fairly in accordance with what David Chalmers claims (1995, 1996). We will return to these arguments when discussing the nature of explanation in the second part of the essay.

Disregarding content: The place for qualia

One of the arguments for the "independent" existence of qualia, as we have seen, is the fact that they do not seem to "be about" their neural correlate[9]. When we see an object, we do not witness neural processing. Jackson's (1982) thought-experiment about Mary the colour-scientist tried to show that Mary could know everything there is to know about colour in physical terms and still learn something new about it if presented with a hitherto un-witnessed colour. In short, the content of qualia does not seem to tell us anything about what's happening in the brain when we experience them. There is a Kantian point in this: What we learn about the world when we watch it through our eyes is what it looks like, and what it looks like is not an intrinsic feature of the world, but a feature of the relations between phenomenal consciousness and physical entities. That is: to be in a certain neural state is nothing like it is to watch, or hear, or smell that neural state. Even the finished interpretation of incoming stimuli is just a bunch of neural activations, the nature of which is inaccessible to us from introspection. But then, what is it that is directly accessible to us in introspection? The not so breathtaking answer is: our conscious experiences. There is, as noted, some extent to which we may be said to have introspective access to neural states, that is: if we come to know our experiences by their neural correlate. Then, exactly in the same way as we have learned to conclude that there is an apple present when a certain experience takes place, we can learn to conclude that "my C-fibres are firing" when we feel pain. But, again, it is not the content of these experiences that are the problematic feature[10].

Even though we are certainly interested in the cognitive mechanisms underlying our abilities to act, speak and discriminate, our main concern here is with our phenomenal experiences of doing these things. Indeed, the existence of Qualia seems to be part and parcel of what motivates us in making these efforts to reach explanations. While some may deny this, it seems reasonable to say that our discriminatory powers with regard to colours, for instance, would not be half as interesting if they were not normally accompanied by a difference in qualia. The case of blind-sight might be a refutation of this, for here we seem to have discriminatory capacities without the phenomenal consciousness. But actually, here it is the discrepancy that is interesting. The interesting thing is why, and, how, if we can manage discrimination without consciousness, there should be consciousness at all. And, more-over: why it seems as if it is our access to consciousness that provides the foundation for these discriminations. This provides some indication that phenomenal consciousness might not play the role that we usually think that it does, but that it is the underlying, contingently co-existent, structures that provides these foundations. Arguably, even the links between different qualia (likeness, association and so on) is accomplished by these structures. That is: If, by some "strike of the demon" some qualia should be cancelled, but the causal connections between processes in the brain would remain the same, the associated qualia should probably result anyway. This throws some light on the suspicion that the memory of qualia has nothing of the qualia character, qualia is not stored as qualia, and need to be produced anew every time it becomes actual.

[9] See Gazzaniga et. al. (2002) "A vast amount of research in cognitive science clearly shows that we are conscious only of the content of our mental life, not what generates the content" p 661.

[10] Dennett (1993) argues that there is no such thing as "a picture in the head" (arguing against the Cartesian, as opposed to the Baarsian, theatre), which is all very fine, if we keep in mind that we then should be prepared to accept that there is no such thing as a picture in the world either. What Dennett probably should be thought of as saying is that there is nothing in the head when there seems to be a picture in the head that is like what there is in the world when there seems to be a picture in the world. But that is scarcely interesting enough to even be worth stating.

First-person methods in science

Scientific research within the cognitive area relies on reports of qualia, and only insofar as the "patient" is thought to be sincere in these reports, when trying to map higher cognitive functions. Jack and Ropestorff (2002) in their highly suggestive article wrote that "Introspective evidence is the only form of evidence that directly bears on consciousness and subjective states" p334 (see also Crick and Koch (2003), and Goldman (1997)). Consider the special cases where patients manage to perform normally, but where they claim to be oblivious of what they are doing (for example Corpus Callosum patients laughing at jokes they are not aware of, see Gazzaniga et al. (2002)). We immediately infer that there is something amiss here, because explanations in cognitive science wouldn't be half as compelling if all they did was to explain behaviour and not, as is the case now, being interestingly correlated to how we experience our dealing with certain cognitive tasks. Leaving phenomenal consciousness, the first-person perspective, out of cognitive science would amount to widening the explanatory gap and cognitive science should be about bridging it, or at least about enhancing our confidence in leaping it. Something should also be learned from our curiosity of what it is like to be such a patient. If it is just a matter of the inability to linguistically report what we experience, how could this be? We know what it is like to be unable to express in words what we feel, but what would it be like to be unable to report something that we know exactly how to report under normal circumstances? Is it like not finding a word? But then, why do the patients not report this inability? A satisfactory explanation here should involve an account of this, and the intuitive point here is that there really is a fact of the matter. Methodological triangulation (behavioural measurement, recordings of brain activity and introspective evidence related to each other) in cognitive science ensures that we don't speak totally past each other. Correlating the results of each method might enhance our confidence in going from the one to the other, but correlation must be earned. And, most importantly, correlation is not reduction. But what, then, are first-person methods, and what are they doing in science? As Jack and Ropestorff (2002) noted above, and as Chalmers (1) argues, first-person *data* should be treated as facts in their own right. And though the development of first-person *methods* is still in its infancy as scientific method, there are clear indications how such research might go. There is a very interesting ongoing debate between Dennett (2001b) and Chalmers in this matter. Dennett claims that even if there is an irreducible sense of "what it is like" in having certain experiences, this does not play *any* role in establishing a complete theory of consciousness since the "what's it like"-ness does not amount to anything important. Again, then, Dennett's argument turns on the misconception that conscious experience poses a problem because of its *content*. First-person related issues can be explained from a third-person perspective, he argues, but fails to see that it is not the relation to the first-person that is at issue here, but the *fact* of first-personhood. Attaining the relevant correlation requires a systematic and well developed first-person method, since nuanced correlation, the important part-goal of every sensible theory of consciousness, could not be attained in any other way.

The relations between the physical and the phenomenal spheres

There is no question about whether there is a dependency-relation between the phenomenal and the physical; there obviously is one. Pulling the plug on the brain makes the light of consciousness go out without residue. I will assume that this claim is uncontroversial. The question is as to the nature of this connection. That the physical world is causally closed, if it is, does not rule out the existence of other features. That is: it is not obvious that a mere physical duplicate of our world, keeping the causal relations invariant, would contain phenomenal consciousness. But withdrawing phenomenal consciousness from the causally closed world is exactly what makes it so suspicious in scientific (physical) terms. All other "respectable" properties seem to have a place in a causal network (see for example Hacking (1983), and Barnes (1994[11])). But, of course, bringing causal role in as a requirement for ontological respectability is just begging the question. That would amount to assuming functionalism from the start, and a theory of consciousness that from the beginning rules out epiphenomenalism does not even get off the ground. One might want to object here that if causal role is not invoked at this early stage, there is no end to the amount of properties that can be postulated by would-be scientist. This argument fails, though: We have *independent* evidence for the existence of Qualia, namely from introspection, and if *anything* like this would turn out to be the case for any of these other proposed properties, then surely we would be wrong to rule them out at this early stage too. The point I want to make, and I will return to it in the second part, is that we know the phenomenal part of consciousness *directly*, and not by what it does. This special status is also what motivates the speculations on phenomenal consciousness as an *epiphenomenon*.

Some notes on epiphenomenalism

The argument presented so far has dealt with the

[11] Eric Barnes (1994) argues that explanatory power must be given the invocation of "Brute Facts" (that is: not further explained entities). But his argument will regrettably not work in our case, since the brute facts on his account are picked out and gain their status with reference to their place in some kind of causal network.

problematic fact that phenomenal consciousness is not known via any causal connections. Indeed, to some extent it seems to be the other way around: causal connections are known via phenomenal consciousness. Epiphenomenalism claims that this fact motivates the position that phenomenal consciousness does not have any causal powers[12]. There is something profoundly counterintuitive about this, since we so often, especially when dealing with agents like ourselves, appeal to conscious experiences as contributing causes for their behaviour. We realise that having certain experiences are not sufficient for causing behaviour, and we might even acknowledge that they are not necessary for causing this particular behaviour, but that a causal history of the event would not be complete without reference to this experience. I speak the truth when I say that the reason that I keep listening to Bartóks fifth string quartet over and over again is how it sounds, how it makes me feel. So how could it be true that conscious experience causes nothing? Actually, I believe that it is true that if Bartóks fifth string quartet did *not* sound that way to me, or did not give rise to any conscious experience at all, I would not keep "listening" to it. But that does not mean that my conscious experience causes me to keep listening. This is so because epiphenomenalism can admit that there is *some* connection between the causal system of the brain and the conscious experience: If the experience is lacking, this is because some underlying structure is missing, and this structure is what contributes the sufficient causes for my actions. We *know* these causes primarily through our conscious experiences, which accounts for why we tend to ascribe the causal power to the experiences themselves. This is what we should learn from the discussion of the split reference above.

Naturally, the investigation of the things that do what we tend to think that consciousness does is a worthwhile project, especially if we want to find which structures eventually result in conscious experiences, and so some intimation of what the *relata* are in the psychophysical relation. Baars' (1996) global workspace theory[13], for example, gives what he himself calls "the most complete account to date of the interplay of conscious and unconscious processes in perception, imagery, action control, learning, attention, problem-solving and language." The content of consciousness, he claims, becomes globally available to many unconscious systems. Consciousness appears to be a necessary condition for discrimination, at least in creatures we believe to be conscious. But, first of all, this is not totally true, as the case of blind-sight brings to the fore, and secondly: even if it where true, it would not show that global workspace and phenomenal consciousness are not distinct. Baars writes that there is clearly some sort of causal interaction between our personal experience and our information-processing account of limited-capacity interference. Epiphenomenalism is consistent with there being some relation between the two, but not with this relation being reciprocally *causal*.

The close relation is also shown by the observation that we, if given enough feedback, can come to control the firing of particular neurons. Baars notes this and Paul Churchland (1984) considers it as a refutation of epiphenomenalism. But clearly, this refutation could only go through if epiphenomenalism is supposed to be false from the start. When dealing with explanatory issues in the second part, then, I will treat functionalism and epiphenomenalism as the two main contestants and try to see what requirements on explanation can be combined with, and ultimately lead to the confirmation or refutation of, the two theories. The reason why I choose these two theories for closer scrutiny is that the central, and controversial, point of the causal role of qualia is especially clearly emphasized in them.

Emergentism Revisited

There is a third proposal (there are, of course, several), the consideration of which I will restrict to this section of reasons I will shortly provide. This theory is called *Emergentism* (Sperry 1980, Dewan 1976), and it depicts consciousness as something like a *system property*. As the name may lead on, this idea suggests that a system can have properties not attributable to any of its parts. Such properties are rather common in nature and, arguably, most properties we deal with are of this kind. Things are true about molecules that aren't true about the atoms that constitute them, etc. The lesson to be learned, obviously, is that ontological reducibility is not equivalent to nihilism about complex properties. Molecules "just are" constellations of atoms, but that does not mean that they are not real, they are just not ontologically primary. They earn their separate ontological status, so the argument goes, by way of their non-reducible properties. The argument from the emergentists in the case of consciousness, then, boils down to this: Consciousness is a property of this kind, and our inability to see how this property can result from scientific entities of the traditional kind is due to the trickiness of emergent properties in general. But, as noted, we deal with emergent properties (and, indeed, emergent "entities") all the time, so why should the inexplicability of consciousness be taken as a suggestion as to its emergent character? Quite surely, if consciousness is to be explained functionally, the function in question would more or less have to be emergent in this sense. When I speak of functionalism above and below I take emergentism

[12] On a lighter note, epiphenomenalism is the position that enables you to answer the "well, after all, your feelings are just a bunch of electrical charges in the brain" with a "that's not exactly true".

[13] Baars is responsible for bringing the "theatre" metaphor for consciousness back from disrepute.

to be included under that term, and the arguments that apply to functionalism applies likewise to emergentism in the functionalist interpretation. Are emergent properties causal? Sperry (1980) suggests that they are and holds this as an argument in favour of emergentism. It is quite clear from the argument presented (as it is in Dewan (1976)) that the phenomenon in question is what we have called *psychological* consciousness. The causal properties usually attributed to consciousness are most likely fittingly described as emergent, but, as I will continue to argue, it is at least conceivable how these properties come about. But the property we want to have explained is *phenomenal* consciousness. And just saying that *this* property "non-mysteriously" emerges seems like a contradiction in terms: Until it is clear *how* phenomenal consciousness can come about at all, the claim that it emerges explains practically nothing. The argument presented (in favour of epiphenomenalism), then, could be said to be about the *interpretation* of emergentism, or about the explication of the relation between the properties at hand. If we settle for the causal version, then we have functionalism (in Dennetts sense), if we settle for a non-causal version, then we are stuck with epiphenomenalism. Surprising features, emergentism tells us, can arise from complex systems, without our accepting anything mysterious on the fundamental level. Sure, one might answer, just show me how it's done. Emergentism does not solve the hard problem, it merely rephrases it and replaces (or extrapolates) the metaphors and analogues that, sometimes quite helpfully, apply. The examples given by Sperry and Dewan all deal with the performance of functions. Nowhere is the problem of the phenomenal character dealt with, and still it's assumed that the fact that the alleged emergent function-properties are identical somehow with qualia. Accepting this as an explanation is slightly reminiscent of the "and then some miracle occurs" version of explanation rejected in the introduction.

The Ontological challenge

The question of whether we need to add another realm to reality in which to place qualia, or a fundamental law that connects physical properties to phenomenal ones, is really a question that belongs as much to the issue of the nature of explanation as it does to the issue about the nature of consciousness itself. What is clear is that when explanation breaks down, or does not seem to go all the way, there is always the possibility of making additions to the ontological list of fundamental properties. In those cases, explanation comes to an end because ontology comes to an end. The necessity to accept fundamental properties and laws is the world's way of answering the "why" question with a "because." We should be open to the possibility that the lack of a successful *reductive* explanation for phenomenal consciousness might depend on the fundamental character of qualia. Reductive views fail, on this account, because they try to make do without something that is fundamental to consciousness. True, we should, at least to some extent, try to restrict our ontology, and for every phenomenon make extensive effort to try to see how something more fundamental could succeed in bringing it about. But when these efforts fail, adding fundamentals is always a possibility. Occam's razor reasonably tells us not to accept more entities than we can make it without, but the claim given by non-reductionists is exactly that we can not make it without these irreducible properties. Again, this position should not be either accepted or refused to be taken seriously from the start. What should be settled before moving on to solve this harder question is what would count as an explanation of consciousness. In the second part of this essay I try to give an account of the difficulties facing us from this part of the problem.

Brief intermission and summary

Before moving on, let's recapitulate briefly what has been stated so far. In this part of the essay I have tried to outline the problem of consciousness that, I believe, is most interestingly correlated to the problem of the nature of explanation. The established terms for the aspect with which we are concerned here are "phenomenal consciousness" or "qualia", and I have been using them interchangeably to denote the "what it's like"-ness of conscious experience. I have suggested that we should be prepared to "split the reference" when talking about consciousness since what makes it true that there is such consciousness may or may not be the whole story to be told about consciousness. That is, the reference should be split between the experience itself and that which makes it true that that experience takes place. Splitting the reference in this way is not, yet, to suppose that phenomenal consciousness is distinct from some function performed by the brain or, indeed, from that brain itself: it merely assures that we do not start of with assuming some controversial identity statement. It does, or so I believe, help making the problem clearer and it forces us to take it seriously. I have also claimed that it is not the content of qualia, but rather the fact of qualia that is in need of explanation. I have noted, and this will be central in the concluding part, that the possible causal role of consciousness is a controversial affair, and my choice of functionalism and epiphenomenalism as the two main contenders in the philosophy of mind is motivated by their being unambiguous in their treatment of this role.

PART II: BRIDGING OR LEAPING THE EXPLANATORY GAP?

What is and where do we find explanation?

We encounter explanations daily: we ask for them, we give them, we infer them and we make an effort to grasp them. The perhaps most common type of explanations that we come across in our everyday life are reason-giving explanations for intentional actions. One of the most prominent things about these everyday encounters with explanations is that we will almost always settle for less than complete explanation. All that we ask for is enough to fill the gap between what we already know and what we want to understand. Understanding is reached when we believe that we can conclude the rest ourselves. For example: when asking someone for an explanation of why he voted for a particular party in the last election, we do not need to hear a complete causal history about what eventually lead to his putting the vote in the ballot box. All we need is some reason, some preference, or some agreement in point with the party in question and then we complete the explanation ourselves. In most cases we do not even exactly complete the explanation, it's enough that we see how it might be done. In the controversies of fundamental science and philosophy we are looking for something more than this, namely something like a *complete* explanation. This, in effect, is something like making explicit the reasoning we usually infer in less controversial cases. As is well known in philosophy at least since Wittgenstein, explanation has to come to an end somewhere, so what we really would want is this explanation to end on first principles on which we all can agree. Is this, then, what it takes for an explanation to succeed? That is: a smooth movement that terminates in principles and facts that are widely accepted? It seems plausible to claim that what an explanation needs to do in order to be successful is to reach out, to bridge the gap, to what the agent already knows, what he or she takes as given. But it is important to realize that different agents have not just different initial knowledge when presented with an explanation, but also different standards for explanation. What counts as a good explanation for one person may not count as a good explanation for another, since their background assumptions may differ. And, indeed, one and the same agent usually have different standards for explanation within different contexts. What we demand of an explanation in fundamental physics is not the same as what we demand of the explanation from the kid who smashed our window with his football. In short, we will accept an explanation when we feel satisfied that the *explanans* reaches what we already hold to be true of the *explanandum*. As we shall see, this account poses problems for the theory of consciousness.

There is also the opposite question: What does it mean for an explanation to fail? Is it just the sensation that we can *not* "take it from there" and complete the explanation with help from what we already know or take to be true? Take for example the claim that Relativity Theory solves the Twin Paradox. Let's, for the sake of the argument, assume that I fail to see how relativity theory accomplishes this, but that I also realize that my knowledge of theoretical physics is sadly inadequate. Now, there seems to be at least two ways to deal with this; one is modest and the other is…not so modest. The modest way is to acknowledge that there is an explanation in front of me, but that I just fail to realise how it succeeds. The other is to deny that I have been given an explanation at all, since I still do not understand how Relativity Theory solves the Twin Paradox. What could this possibly mean? My guess is that the modest answer is guided by a belief that the lack of understanding is due to my lack of prerequisite knowledge in theoretical physics, or a lack of theoretical imagination. There is still an explanatory gap, but I realize that it's "on my behalf." The not so modest answer is guided by my belief that my lack of understanding indicates that there is still some distance to be covered by the explanation. My lack of insight into theoretical physics adds up to me not having the full explanation present. But, of course, *no* full explanation could practically ever be *in front* of me, in that sense. How I choose to deal with the problem, arising from my lack of understanding, seems to be a question of confidence as much as of anything else. We often seem to say things like "I have had quantum mechanics explained to me over and over again, but I still do not understand it"[14], thus admitting that an explanation is given, and presumably a successful one, it is just not successful in making me understand. In this case perhaps I can see *that* someone else with more thorough knowledge in theoretical physics might "complete the explanation" but not exactly *how* this could be done, and that would amount to recognizing the explanation as an explanation. What we do here is that we use the good judgment of authorities as "first principles" at which explanation may stop. One of the great benefits of having a human mind is that we can use abstract representations as premises (see Tomassello (2001), for instance).

There is nothing strange in our failing to see how an explanation works, and still recognizing it as an explanation. But on the other side, nor is there anything strange in our realizing how an explanation does it's work, and denying that it succeeds, perhaps because the entities it refers to do not exist (as in God causing earthquakes etc.). What we have here is thus

[14] This is, if one should believe Niels Bohr, a good sign that one is starting to understand it.

an ambivalence in meaning of what it is for an explanation to succeed. It can succeed in convincing, and it can succeed in explaining. And, importantly, it might succeed in one respect while failing in the other[15]. Complex and fundamental theories challenging the hard core of paradigmatic science or just having a disturbing influence on common sense are always likely to be challenged, and rightfully so: explanatory power should not be distributed lightly; it is something a theory has to *earn*.

The relevance of this discussion for our main concern in this essay is clear: If we fail to see how a certain theory might explain consciousness, is this sufficient for us to refute the possibility of its doing so? Chalmers (1996), as we shall see, argues that the lack of conceptual implication from the explanatory facts to the fact to be explained indicates that something is lacking in the explanation, and suggests that this might be something that we have to postulate, something *primitive* in the epistemological sense. McGinn (1989) (but see de Leon (1995); see also below) suggested that the explanation in question might be forever beyond us. That, of course, does not mean that there *is* no explanation, only that there will never be a full explanation for us humans[16]. The fact that we are not convinced is not a conclusive reason to refute an explanation, but it is certainly a good reason to challenge the explanatory power of an alleged explanation. There seems to be the possibility that an explanation could be given, but that we lack the cognitive powers to grasp it. But this, I would say, just points to the fact that we can not *complete* the explanation ourselves, perhaps because something is still lacking, and this something should be contributed by the explanation for it to be a successful one. Answering the "lack of conviction" objection by claiming that some kind of explanation is all there is, then, is too say that the last step of the explanation should be taken in something like a leap of faith.

The last possibility is that we might have a full explanation but fail to realize that we have it. This possibility differs from the "leap of faith" variant in that it claims that our recognizing it just consists in an adjustment to the facts, not in any cognitively adventurous leap. Adherents to the reductive views will probably argue that their claim is of this kind; while their opponents will try to put it in the "leap of faith" category, thereby discrediting it.

There is a further problem, namely the problem of multiple realizations. Two explanations might be absolutely equivalent as to their simplicity, their ability to predict experimental outcomes and their coherence with the larger corpus of unified science. And, quite frankly, they could both be correct. Over-determination is not unheard of in everyday situations, and might be present also in more fundamental explanations, even if they are very seldomly acknowledged in science[17]. I will not treat this problem here (if, indeed, it is a problem), there is enough problems finding *one* plausible explanation for us to worry about there being several. The main issue in this part of the essay, then, will be about what could reasonably be asked of an explanation, and specifically if the *explanandum* happens to be phenomenal consciousness.

Explaining consciousness

In the first part I tried to spell out what it is that is so special about phenomenal consciousness. In this part, I will try to argue that this special feature of consciousness, its phenomenal quality, leads us into speculations about the nature of explanation with regard to it.

Let's begin by making away with the suggestion that the problem of explaining consciousness is due to self-reference. We cannot explain consciousness, so this argument goes, because the thing to be explained is identical with the thing making the explanation. But there is no inherent problem of self-reference (see Kripke (1975)). Take for example: "This sentence consists of six words." True, self-reference is a superior tool in creating paradoxes, but that does not mean that there is something inherently wrong with self-reference. Every instance needs to be assessed separately, and not ruled out in advance. And, anyway, the problem of explaining consciousness does not just consist in explaining the consciousness that does the explaining, but in explaining *any* instance of consciousness. Still there is something to be said for the self-reference objection. Marion Gothier (1998) made a rather sophisticated suggestion of how the self-reference objection might go. She claimed that: "To put it bluntly, science doesn't describe physical reality. It is instead a systematic description of regularities in experience." What we usually do in explanation is that we explain what we experience. "What is heat" is the question why some things feel hot or have other causal properties whose effects we somehow experience. "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." Gothier's point is that we cannot explain conscious experience since conscious

[15] A lackmus-test for your intuitions about explanation might be the following: Does Newton's laws explain anything? It has been said that they suffice to explain the behaviour of everyday middle-sized objects, but is it a *true* explanation? Does it succeed in both the senses mentioned?

[16] This echoes beautifully of Kafka's "There is hope, but not for us."

[17] The question that presents itself is whether necessary entailment from *explanans* to *explanandum* is really sufficient for proper explanation: We might still leave something out. Usually, though, over-determination has other causal properties than just determining that specific outcome. In our case this might not be true, if epiphenomenalism turns out to be true. Our case is different in many respects, and that is, really, what is at issue here.

experience is already part of the *explanans*, and cannot enter into the *explanandum* without (vicious) circularity. Scientific explanation, she claimed, tries to give an account of the underlying reality, assuming some kind of isomorphism between the world and our experience of the world. The explanation of the outside world thus just takes the experience out of the picture. But taking the experience out of experience leaves us with nothing, which is why scientific explanation of experience is not possible, on this account. Gothier says that "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." The objection towards Gothier's argument is quite obvious: the conscious experience in the explanans need not be identical to the one present in the explanandum: thus we dodge the vicious circularity argument. What Gothier needs is the further, much stronger claim that the same *type* of event cannot figure on both sides without vicious circularity. Assuming isomorphism, in the way proposed by Gothier to be generally employed by science, might also lead us to the underlying reality of consciousness itself. Indeed, this seems to be the subject matter of much cognitive science. Gothier does not consider the route by which this isomorphism is reached. The underlying reality explored in cognitive science accounts for how some features of reality presents itself, and thus *earns* its explanatory powers[18]. But, and this is important, this does not *explain* conscious experience itself. Correlation or explication of the "Supervenience-basis" is not identity, and the further connection, the true identity condition, is really what we need in order to reach full explanation. Gothier's admirable account rests heavily on a conception of science that might be questioned, but it at least gives an intuitively compelling explanation of why the hard problem of consciousness seems to be so hard. Gothier concludes that physical science gives the explanations that are possible. This means that she does not consider the postulation of psychophysical laws, or irreducibly phenomenal properties, as amounting to full-fledged explanations. This apparently controversial conception of the limitations of science throws some light on the problem, I think, but is not necessary for reaching negative conclusions about reductive explanations of phenomenal consciousness.

Arguments for reductive explanation

Why should there be some feature of the universe that is fundamentally inaccessible to our explanations? Phenomenal consciousness seems to be restricted to living creatures[19] and all their other features are subject to explanations of various sorts, evolutionary and others, so why should not this feature eventually yield to explanation? It seems to be just obvious that there must be some evolutionary history to be told about how phenomenal consciousness came about. The evolutionary gain of consciousness in the *psychological* sense seems to be quite straightforward, so the story to be told there shouldn't pose too much of a problem. But what about the phenomenal character of experience? Granted that the psychological *could* be identical, and the phenomenal be absent or just different, we would have likewise adaptive behaviour and the same evolutionary success whether the phenomenal would be present or not. The arguments for explanation have two edges: First, the normative edge: there obviously *should* be an explanation. The other is factual: We already *have* explanations of the required kind, so by inference they are possible. Physicalism claims that the only basic properties there are are physical properties; so if phenomenal consciousness could be explained on this account it better be accomplished with reference to the properties available. Bridging the gap, the physicalist/functionalist would say, is about realizing that some "extra" phenomenal property isn't needed in order to reach full explanation. This is, I believe, what Dennett's whole argument is really about. Most straightforwardly his argument for accepting this kind of explanation surfaces in his choice of analogies found in the history of science. The most prominent of these is the reference to the "vital spirit" in his reply to Chalmers (1995). Here Dennett tried to show the similarities between the "hard problem" of consciousness and the belief that life could not be reduced to physical processes and that a "vital sprit" was needed. Chalmers in his reply to this objection (1997) noted that the "vital spirit" was invoked because people could not see how the functions performed by living creatures could be performed by mere mechanisms. But the problem was still acknowledged to be about the performance of functions. In the case of phenomenal consciousness, it is exactly the non-functionality of this feature that grounds the suspicion that functionalism or physicalism cannot account for it.

The general lesson is that no historical parallel is fitting in advance. Exposing historical parallelism is a luxury to be enjoyed when the problem is conveniently solved. And the same goes for metaphors in general. Though, admittedly, metaphors and parallels can provide useful clues as to how problems might be solved, we should not be blinded by their apparent fittingness. Seeing how an explanation might do its work could be something like grasping which analogies or metaphors that apply. But understanding a metaphor is not sufficient for understanding the phenomena in question. To refute an argument from analogy it is sufficient to point to some respect in which the cases differ and explain why that respect is relevant for the argument

[18] This is, in effect, a Heideggerian solution to the scepticist argument founded upon Cartesian doubt

[19] But AI might provide challenges to the essentialness of this connection.

not to go through. And that is exactly what Chalmers does in his reply to Dennett. What is disputed in the debate we are dealing with here is whether certain metaphors and historical parallels are good ones, and this cannot be settled in advance. In advance it is more likely to lead to prejudice and empty rhetoric.

Baars (1996 and 1997), as we have seen, argued that the only explanations we need, or are likely to find, are found in the amazing correlation between conscious phenomena and neural processing in what he called a "global workspace." What this argument tries to establish is that whenever phenomenal consciousness is present, some psychological process takes place: consolidating memory, moving attention, controlling action etc. This suggested some sort of operationalism for consciousness; and then, of course, the identification of consciousness with these operations is a free lunch. But this correlation is not even close to solving the problem posed by phenomenal consciousness. To identify what usually appears in conjunction with phenomenal consciousness is not to explain consciousness. Most people seems to accept that there is some connection between these processes and consciousness and some people even claim that the connection is *identity*, but just claiming it does not amount to a very strong argument. How does processing in a global workspace bring about phenomenal consciousness? As noted in the first part of this essay, it is easy to identify phenomenal consciousness, which, by argument, has no causal powers, with that with which it is associated, and has the causal powers we usually attribute to consciousness. But, as we have seen, this temptation should be resisted, or else we are just begging the question. The debate, after all, is about whether consciousness can be picked out functionally or not. As we shall see below, Baars' account is subject to the "bridging principle"-objection (posed by Chalmers (1998)).

Koch and Crick (2003) have a similar approach, but they straightforwardly abolish the hard problem and instead outline a "framework for consciousness." On the first page they write "It appears fruitless to approach this problem head-on. Instead we are attempting to find the neural correlate(s) of consciousness." This is a very sensible and admirably modest strategy. Given that the phenomenal aspect of consciousness is not all that is important, one can bypass it and continue to flesh out an explanation of the psychological aspect, thus using first-person reports of phenomenal consciousness merely as indications of the psychological trait in question. This, indeed, is giving up on the hard problem, and focusing on the easy ones[20]. And note that this bypass is possible exactly because we have taken the epiphenomenalist view on phenomenal consciousness. Perhaps there resides in this strategy a hope that somewhere later on explanation of the phenomenal aspect will be attained or realised to be unnecessary. Epiphenomenalism, after all, leaves room for a full explanation of the causally closed universe that does not bring in the phenomenal aspects. All it means is that a complete physical description of a world does not fully determine the ontology of that world.

[20] It is important to note here that the 'easiness' of these problems is not to be understood as an underestimation of the technical complications of this kind of explanation, only as a indication that it is conceivable how an explanation of this kind might go.

Arguments against reductive explanation

In this section I will briefly try to do justice to the arguments against reductive explanation of phenomenal consciousness presented by McGinn (1989), Nagel (1974), and Jackson (1982). But the emphasis will be on the argument developed by Chalmers (1996) and Jackson (1998). It should be noted that none of these arguments are directed against *explanation* as such, not even against *scientific* explanation, but against *reductive* explanation. In McGinn's case, the argument is against the possibility of our *grasping* the reductive explanation. In his 1989 article he writes that we know that brains *de facto* provide the causal basis of consciousness, but that we have no idea how this can be so. What's lacking, in our terms, is the feeling that we can "take it from there." "The problem arises" McGinn writes "because we are cut off by our very cognitive constitution from achieving a conception of that natural property of the brain (or of consciousness) that accounts for the psychophysical link" (p 350). McGinn introduces the concept "cognitive closure" where a type of mind M is cognitively closed with respect to a property P (or theory T) if and only if the concept-forming procedures at M's disposal cannot extend to a grasp of P (or an understanding of T). In short: some properties are accessible to our minds, and some are not. Now he suggests that we could be cognitively closed with respect to the psychophysical link: the physical property that entails phenomenal consciousness, and the theory that explains how this is possible. But that we are cognitively debarred from reaching such an explanation does not mean that there *is* no such explanation, the grasping of which could be within reach for some more evolved creature. McGinn notes that our knowledge of consciousness-related issues is mediated in two fundamentally different ways: via perception and via introspection. "Crudely" he writes "you cannot form concepts of conscious properties unless you yourself instantiate those properties" (p355). As Jackson (1982) and Nagel (1974) have shown (see part 1), there are properties of the brain that are closed to perception of the brain. McGinn merely asks whether the property P above is one of those properties. He suggests that it is, and that it is also closed to *introspection*. The property P should bridge the gap between properties perceived and properties introspected, but this property can not

be accessed via introspection or perception.

The most straightforward objection to McGinn's argument is that it seems presumptuous to merely *assume* that this extraordinary property should be *physical*. The argument he gives is successful merely in showing that we need not conclude that this property is mysterious in any way. But, of course, it has to be mysterious enough to evade our cognitive abilities, which is rather rich in itself. What McGinn tries to establish is that there is no *philosophical* distinct problem of phenomenal consciousness: The relevant property is just as natural as any; it is just that we lack the capacity to grasp it. In terms of necessity, then, this alleged physical property P must conceptually necessitate that the thing that has it is conscious in the phenomenal sense[21]. But this amounts to a postulation of an irreducible phenomenal entity and a natural law connecting it with physical organization. Physical properties as we know them can not logically necessitate phenomenal consciousness, and so this "further property" P that McGinn postulates is not mentioned in our physical theories, and, given that the physical properties *as described by our physical science* are as they are, phenomenal consciousness does not follow *from these properties*. This is more or less exactly what the sensible property dualist claims.

Chalmers (1996) (but see also Jackson (1998)) develops what is probably the strongest argument against reductive explanations. In the chapter named "Naturalistic Dualism" he sums up the argument as follows:

> In our world, there are conscious experiences.
> There is a logically possible world physically identical to ours, in which the positive facts about consciousness in our world do not hold.
> Therefore, facts about consciousness are further facts about our world, over and above the physical facts.
> So materialism is false.

This argument has become famous in the literature as the "Zombie" argument. The point is that there is a possible world in which there is a physical duplicate of you, engaged in the same kind of reasoning, formulating the same sentences, processing the same stimuli, but who lacks phenomenal consciousness. A *mere* physical duplicate of our world does not entail consciousness[22]. In our world, so this argument goes, physics does not exhaust ontology. As noted above, it should be realised that this argument rests on the fact that *our* concepts of physical properties does not enable us to see how they could entail consciousness. But this, of course, does not mean that these properties themselves are not absolutely responsible for there being consciousness. The point is that we do not know them as such, and so there is some extra property (and this could be a second-order property), the nature of which we are not familiar with, that accounts for consciousness. This is exactly what the sensible property dualists (see Searle (1992), for instance) have claimed all along. The argument against reductive explanation inherent in this does not say that explanation is impossible; it just says that to bring about a conceptual entailment from physical properties to phenomenal ones, something more is needed than is available in physics (or neuroscience) at this time. Our present concepts of these properties do not reach conceptual entailment, since it is conceivable that they could fail to bring about consciousness. Note that this argument is almost exactly parallel to McGinn's, with the slight modification that Chalmers, instead of appealing to cognitive closure, suggests that we should just accept qualia as irreducible, and our first priority now is to find the principles that tie them to physical/functional properties. This brings us to the concept of bridging principles.

As we have seen, it is not an uncommon strategy to connect consciousness to some functional organization, then to explain that functional organization and then claim that consciousness is thus explained. Chalmers (1998) noted that "These principles of interpretation are not themselves experimentally determined or experimentally tested. In a sense they are pre-experimental assumptions." Bridging principles (as Baars' (1996) global workspace for example) merely beg the question, or at least attain their plausibility in less than conceptually binding ways. Full explanation of the bridging principle does not amount to full explanation of consciousness. Localizing the neural or functional correlate of consciousness is a worthwhile project, but it is important to note that as yet, identity is not attained. Bridging principles are assumptions, and as such they are in need of justification. The question is what is required of these justifications.

[21] The relation between this property and the theory seems strange to de Léon (1995), and he builds much of his argument against McGinn upon it. And upon what it is to grasp a property. He claims that it should be something like knowing all, or some, or the typical causal potentialities it possesses. If this is so, then surely we could know the property in question, without knowing how it enables consciousness. I do not believe that de Léon's argument works, though, since McGinn suggest that this property accounts for the *link*, and that is precisely what the theory is supposed to do. Given the property, the theory is a free lunch, and vice versa. We must give McGinn that T probably would escape our conceptual abilities if a sense of "what it is like" is to follow *logically.*

[22] What the "mere" in this sentence amounts to is that the physical description must contain what Jackson (1998) called a "stop-clause." That is: to get the desired possible world, the physical description must be completed with an "and that's all there is" clause, thus debarring extra properties that may entail phenomenal consciousness.

Arguments against the arguments against reductive explanation

In this section I will briefly treat an argument presented by Patricia Smith Churchland (1997) against the arguments treated above, but the main issue is the arguments against the possibility of Zombies given by Dennett (2001a). The argument Churchland turns against is the argument from ignorance. This argument says that we do not know anything about property P, and thus P can never be explained, or nothing science could ever discover would deepen our understanding of P, or P can never be explained in terms of properties of kind S. Churchland here obviously joins the traditional philosophical parlour-game of flogging a dead horse: no one reasons in this way. Boiling a type of argument down to its essential structure is a nice enough philosophical strategy, but in this case, what Churchland is boiling away is the additions that make the argument plausible. What the *real* argument is about is that something is missing in the reductive explanations given, and that we should be open, due to the apparent hardness of the problem, to think about it in different terms. That we do not understand how an explanation does the job it is supposed to do is, admittedly, not a sufficient argument for showing that it fails. But claiming that one sees how an explanation may succeed is not sufficient for showing that it succeeds either. And when it comes to explanation, my intuition is that it is really the people who claim that they understand the phenomenon in question that carry the burden of evidence. Churchland argues that saying that one fails to see how neurons can enable consciousness is just begging the question. But surely, saying that one sees how it can be done just begs the question the other way. This argument leads us nowhere.

To deny the possibility of zombies one needs to show that the physical nature of the brain *necessitates* there being phenomenal consciousness. Dennett (2001a) claims that the conceivability of zombies is *only* apparent and that we, if we would grasp what we are actually doing when ascribing phenomenal consciousness to ourselves and others would realise that Zombies are not possible, since there is no independent phenomena to be inferred. Chalmers' requirement for fundamental laws for getting from functions to phenomenal consciousness, Dennett claims, arises from the "Zombic Hunch[23]" – and that is just a fling that will eventually pass. Of course, Dennett's argument begs the question: to establish it, he needs to search for independent arguments, or let the success of his line of reasoning talk for itself. Dennett's conjecture that future generations will laugh at the hunch in question is actually quite unalarming: future generations have no birth-right to truth.

My contention is that the fishiness of the physical identical twin who happens to be a zombie may be dependent on the "this-ness" of reference. We say "how could anything be physically identical to *this* and lack phenomenal experience?" Our consciousness is so intimately connected to our physical make up that we can for no practical purposes be required to "split the reference." The splitting of "psychological" and "phenomenal" features is not easily done, due to our mixed up intuitions about, for example, the causal powers and structure of consciousness. Qualia is, after all, a rather technical concept.

In a mere physical duplicate of this world, all the third-person evidence would be in place for there being phenomenal consciousness, Chalmers' twin would still be writing his articles on the hard problem of consciousness, using the same arguments[24]. Given this, we need to understand that the *only* valid evidence we have in *our* world for the phenomenal aspects are first-person evidence. That is, if we could somehow witness the physical duplicate world, we would still be at a loss as to whether there existed phenomenal consciousness in it or not. Admittedly, we are likewise at a loss to whether *our* world, with the exception of ourselves, isn't thus lacking in phenomenal quality. But that exception is an important one, and the one that makes us confident in judging that something similar is also present in creatures that behave roughly as we do. Of course, watching the other world we would have likewise good reason to judge them to be conscious[25]. The point is that their being conscious does not follow necessarily from their behaviour and their physical make up. The intuition is not ready for extinction just yet.

Dennett has in article after article tried to convince us all to get on the bus, to confidently make the explanatory leap (or to cross what he persuasively claims is a bridge), but arguments of the kind presented by the "qualia freaks" are highly successful in making us cautious about making this leap. Dennett rhetorically asks us what could possibly go wrong if we accept it, since nothing is thought to follow from the fact of conscious experience, but this kind of "no-harm done pragmatism" hardly seems a proper way to deal with the hard question of consciousness. There is obviously some kind of prima facie plausibility to arguments of the "there is no such thing" sort, since keeping our entities to a minimum yields simplicity, which is something of a master virtue of fundamental science. But of course, an absolutely functional

[23] In Dennett (2001b) he writes that the answer is to include the Zombic Hunch among the facts to be explained by a good theory of consciousness. This, then, is the Heideggerian strategy.

[24] And on twin-earth, philosophers would still be divided into the same camps that they are in our world. The only difference, so the argument goes, would be that on twin-earth Dennett would be right and Chalmers wrong.

[25] But then again, we should never be very confident in using similarity arguments when dealing with other worlds.

counter-argument to this kind of argument is "oh yes, there is."

In the end, the argument turns of what could reasonably be required of a theory of consciousness.

Requirements on explanation

What is it that we want an explanation to accomplish? We want it to bridge the gap between things with which we are familiar. In the case of consciousness, we want to see how facts of a non-phenomenal type can make facts of the phenomenal type true. In scientific explanations, understanding is reached by conjoining fundamental laws with initial conditions. So what fundamental laws in conjunction with the initial condition of the physical state of a normally functioning human brain bring about phenomenal consciousness? Identifying phenomenal consciousness with some function would make explanation easy: the physical make-up of the brain in conjunction with fundamental physical laws necessitates the function. But, as we have seen, the bridge we have been asking for is exactly the one inherent in this identification. How is it justified? What about the function, or nexus of functions, in question necessitate phenomenal consciousness? As we have seen, the arguments in favour of reductive explanation have turned on the fact that everything that is usually believed to be caused by consciousness is in fact caused by something else. But epiphenomenalism claims that consciousness in itself causes nothing. Its possible non-functionality is exactly what makes it so hard to pin-point within science.

The requirements on explanations vary. Some may settle for a close enough correlation between facts of the one sort and facts of the other, others may surrender when the explanation enables predictability, and, finally, some may require strict logical implication from one type of phenomena to the other. My claim is that reductive explanation of phenomenal consciousness may live up to the weak requirements, but not to the strong one. The question is what this means for the success of reductive explanations.

Each and every one of us has reasons to believe that we ourselves are conscious in the phenomenal sense. Our reasons to believe that other persons are conscious in this sense are not so straightforward[26]; the reason we have here is rather that we have no reason to doubt it. Similar functions in similar material probably give the same result in our world[27]. Now, the strength of the reasons we have for our belief in our own phenomenal consciousness seems to be in a very peculiar way disconnected from the actual causal chain of justification. Consider our Zombie Twins (if you can). By argument they will make the same judgments caused by the same mechanisms, and in this sense, their *reasons* are as strong as ours. And still, the essential part, the part that *really* provides the reason-giving force in this case, is lacking. For the Zombies there really is a discrepancy between seeming as believing and phenomenal seeming: they have the former but not the latter, and *our* reasons for believing that we are conscious stems from both sources[28]. This may be controversial. This concept of reason is not a straightforwardly *cognitive* concept. It is more of an *epistemological* one[29]. The weak requirement rests on the intuition that there is no difference between these two reasons, and thus on the impossibility of zombies. But the thing is that the inference of this impossibility seems to require that one accept the weak set of requirements. Of course, the opposite argument holds that the same thing seems to go for the strong requirement.

Are zombies possible? And, if not, are they *necessarily* impossible? Here, the controversy turns on the status of conceptual analysis, and this status, as we shall see, is deeply entangled with the notion of the requirements on explanation. Nothing in the hitherto given attempts of explanation of consciousness has shown that zombies are impossible. It has certainly been *claimed* that they are impossible, but none of these explanations has shown that phenomenal consciousness follows with strict necessity from the functional organization of the brain.

Weak requirements

Bernard Baars (1996) suggested that the hard problem becomes hard when we adopt an implausible, perfectionist standard for explanation. He writes: "Far more practical criteria are used everyday in medicine and scientific studies of consciousness" (p211). Baars, then, in his search for explanations of consciousness,

[26] Some, Dennett for one (1993), argue that we have exactly the same reasons to believe other people to be conscious as we have for believing that we ourselves are conscious. He finds some support for this in the suggestion that we probably discern other people as subjects before we form any concept of ourselves as such. This is all very good, but there is something amiss here. The reasons are *not* the same. It is our sense of what it is like to be a person that grounds our belief that we are conscious (the Dennettonian response is fitting for the *psychological* sense of consciousness –and the more plausibly culturally constructed sense of consciousness).

[27] It would be rather strange if, say, only fifty percent of us where conscious (even though it might account for much of the recent debates in the philosophy of mind).

[28] We conclude things about ourselves in the roundabout way too: and that is where Dennett, for instance, becomes lost: he believes that our reliance on "external" evidence means that this is the *only* evidence we have.

[29] There is a problem here: there is no causal chain from the qualia to the judgment and probably not from the qualia to the belief either, which is, according to some theories, required for them being knowledge. This is a problem for epistemology in general, and not just for the theory of consciousness.

settles for "close connections" between neural processes and conscious phenomena, and justifies this by the entailment of workable scientific standards. But leading to workable standards in science is scarcely strong enough evidence for something as controversial as an identity claim in the theory of consciousness. It invokes phenomenal consciousness just insofar as it is essentially connected to more traditional scientific entities. But is this enough for explanation? Baars' mistake comes in his explication of what he takes to be the implausible requirement inherent in the hard problem. He takes it to mean that "To know whether you, the reader, are conscious, I must know what it is like to be you." But this is not true; all I need to know is that *there is something it is like to be you*. The "empathy criterion" invoked by Baars is admittedly too strong, but it is not identical to the strong requirements of explanation treated below, and it is not needed to refute the kind of suggestion that Baars gives.

Paul M. Churchland's (1984) argument against the stronger requirements is even more straightforward: Surely we can not require a true theory to entail a false one. His point is that the "theory" of phenomenal consciousness is false. The only thing that needs to follow from the theory to which phenomenal consciousness is to be reduced is something "roughly equivalent" to phenomenal consciousness: "what does license a full-fledged identity statement is the comparative *smoothness* of the relevant reduction." The same point is made in the following passage: "All we can properly ask of a reducing theory is that it has the resources to conjure up a set of properties whose nomological powers/roles/features are systematic *analogues* of the powers/roles/features of the set of properties postulated by the old theory." First of all, there seems to be infinitely many sets of postulate-able properties of this kind, so in that respect, this requirement is too weak. Second, if one of these roles/powers/features is the phenomenal one in the old theory, and nothing entailing this is present in the new theory, then the theory has not reduced phenomenal consciousness, and is thus not a reductive explanation at all. The big mistake here obviously resides in the attempt to explain a nexus of properties by a theory that explains *most* of these properties. This is, obviously, not sufficient. What Churchland needs to show, then, is that some feature of the new theory entails the phenomenal feature. It does not suffice with it entailing something "roughly similar" since, as we have seen, the phenomenal feature is what's in need of explanation. Admittedly, there is no *explicit* reference to this "smoothness-criteria" in Churchland's refutation of the "old theory of phenomenal consciousness," but it is easy to see that this criterion is at the bottom of the refutation.

Later on, Churchland's argument turns directly on the deducibility requirement: "The deducibility requirement would (…) trivialize the notion of reduction by making it impossible for *any* conceptual framework to reduce any other, distinct conceptual framework." This is so because the terms in the theory to be reduced are not present in the lexicon of the new theory. But we do not need to deduce any *theory* from the old one; we need to show how the entities and laws of the new theory necessitate the phenomenal facts. This is what has to be made intelligible. And I do not see why this should be impossible.

Churchland tries to prove his point by showing what we would gain by using neurophysiological terms to refer to our conscious experiences (see also Rorty (1980)). And sure, we could come to know our experiences by name of their neural correlate; but changing words does not explain away the phenomenal character of the experiences thus referred to (as decades of attempts to replace the phenomenal concept of consciousness with the psychological have made painfully clear). That we may in this way eventually come to forget that we ever thought the psycho-physical relation to be mysterious does not show that there is nothing mysterious about it[30].

There are a lot of arguments of the type "scientific entities gains their respectability by way of their place in a causal network, or by way of their role in explaining other phenomena" about (see, for example, Hacking (1983)). But, as Stubenberg (1996) writes: "Notice that this is not a scientific claim; it is an enormously strong philosophical claim about science. It states, in dramatic fashion, what has been called the principle of scientific realism (PSR): Only properties whose attributions helps explain something scientifically are admissible in our world view as truly descriptive of reality." But these arguments exclude non-functional entities from the realm of the scientifically respectable at the start, thus restricting both the area of science and the respectability of phenomenal consciousness as an object of scientific study, in a rather unjustified way. This requirement is both too strong and to weak, and should by no means be accepted in advance. Even the weak requirement seems to involve a very strong requirement.

Strong requirements

In the strong requirement camp we find, above all, Chalmers (1996) and Jackson (1998), but also Searle (1991) and Nagel (1974) can be said to belong to this fraction. And, I might add, I myself tend to sympathize with the reasoning lying behind these requirements. An explanation of phenomenal consciousness must entail it logically, so this argument goes, since everything short of it would leave it open for (logically) possible worlds where the

[30] Actually, the Churchland/Rorty strategy strikes me as kind of Orwellian.

explanans holds and the explanandum doesn't. The phenomenal is not absolutely determined by the physical, where "physical" is understood as "property known to present-day physics." That is, there is a possibility of a Zombie-earth. The objection that this is preposterous is easily met: The point is not that there is something strange about the connection; it is just that it should be acknowledged as a primitive fact about the world. Given this, we have our logical entailment. The point is that something has to be present in *our* world to account for the fact that we are conscious in the phenomenal sense. This requirement is necessary for an explanation of consciousness to come through because nothing else could, given that phenomenal consciousness is non-functional.

The most common objection to this line of reasoning says that physics necessitates phenomenal consciousness, not logically, but *naturally,* or even *metaphysically* (what ever *that* means). It seems that what these "less than logical"-necessity theorists wants to say is that logical necessity is too strong a requirement, and so they apply to other standards that imply restrictions on what may change from world to world in the relevant domain. The problem is the content of these restrictions – if their argument is to succeed they need to invoke just those psycho-physical laws they are putting restrictions in the domain to avoid; which amounts to making the implication in question a logical one. The point is that the arguments that apply to "natural necessity" by invoking these restrictions tacitly imply the laws that we, with our strong requirements, wants to make explicit.

The second objection goes as follows: Could not some conceptual necessity relation be beyond our reach? Not if the concepts we are talking about are *our* concepts. This argument only goes trough if the thing referred to is "the whole thing," that is: more than we know it as. The things to which we refer surely explain consciousness, but not by implementing the properties we know it by. The desired necessity can only be reached by adding properties or concepts.

The cart before the horse: can explanations put requirements on us?

An explanation, as noted above, is never fully in front of us. Some of the work needs to be done by ourselves. More than an entity, an explanation is an event: what we could call an "explanation-situation", consisting in the internally and externally "given," so to speak. The question is whether failure to see how the functional explanations explain phenomenal consciousness is due to the explanation given or how the explanation is received. The strong requirements lying behind the assertion of the hard problem typically ascribes this failure to the explanation given, while weak requirement-theorists generally put it down to our lack of scientific imagination[31]. What could this later claim mean? It means that we should trust the methods used in present-day science, and recognize their validity for the case in point. Consider the following situation: A proper explanation of heat is given to a child, but it is not very likely that the child could contribute its part to the explanation situation. The explanation-situation is incomplete due to the infants not being a competent "completer" of scientific explanations. Among the things that need to be contributed is the understanding that one type of event can constitute another. Arguably, the child does not "get" the explanation of heat partly because it doesn't have the necessary concepts. But partly, it might be that it just fails to see the connection. It could have access to all the available facts, but fail to see the connection. At some point the teacher will say "and this just is what heat is," and then it's up to the child to trust this claim. At some point, our reasons for believing something exceed our reasons for doubting it. Putting faith in connections is something we *learn*. We have learned to believe in causation, since nothing could conceivably replace it. In all cases of causal explanation one might challenge them with the observation that one does not *see* the causation; the only thing one sees are law-like correlations. This, of course, was Hume's point. Here one sees clearly what the objection to the strong requirement is: if one should take the appeal to it seriously, nothing ever gets explained. We do not see causation, but that does not mean that causation does not take place. "All there is to causation" one might say "is these law-like correlations." I admit that this is reasonable. In most cases, pragmatism of this sort is reasonable. Actually, it's hard to see what else could count as an explanation of traditional natural facts. The tendency to refute this kind of law-like correlation as clear-cut causal and/or identity cases, so the argument goes, is due to our being biased against it, as happens when the phenomena in question is our consciousness. But we should not treat a phenomenon differently just because we happen to be peculiarly connected to it; that would just be chauvinistic. Indeed, we need to realise that we ourselves are as open to science as any natural object. This argument says that the explanations are there, we just need to get off our high horses and acknowledge their explanatory force, even when the item in question is something we want to claim privileged access to.

Actually, this rather shameless *ad hominem* argument turns on the wrong feature of phenomenality. There is *no* such biased resistance against explanations of how our legs work, or how our lungs got their capacity, or even of how we came to prefer red wine over white wine or how we fall in

[31] It is important to understand that this is not necessary: the strong requirement can very well ascribe the lack of conceptual entailment to our not having the proper concept.

love. And these things are arguably as close to us as are the phenomenal character of our experiences. Obviously, then, if there is a resistance when it comes to *this* feature, it has to be attributed to something else. And that something just might justify the stronger requirement on explanation.

The stronger requirement is not necessarily connected to the irreducibility claim. The other possibility for the reductivist is to simply state that our concept of phenomenal consciousness is mistaken. That is: the property we claim to know consciousness by is not instantiated, and nothing should therefore entail it in order for there to be a proper explanation of consciousness. This is an extrapolation of Paul Churchland's (1984) argument, treated above. The *proper* concept could be entailed, so this argument goes, because the proper concept is functional, and *this* we must realize. *This* is the requirement that the explanation puts on us. But the question, again, is what could possibly justify this identification. And *here* an explanation must be given, and it is exactly there the lack of logical entailment is worrying, if we want to live up to the strong requirement.

Dennett (1993) asks us to draw what could be called a Reluctant Conclusion: that all there is to phenomenal consciousness are functions and that all our evidence for the opposite is misleading. But this is not reasonable given the access we have to consciousness. In this case we have reason to ask for more, because we have not yet any account of how we come from *this* to *that*. We have learned to see the connections given in regular science, but nothing has prepared us to make a similar leap when it comes to connecting subjective experience to physical functions. Scientific explanations are of a kind that finds its standards in a public world (see Goldman (1997)); we only have to see how one publicly available event leads to another. But when the explanation concerns subjective experience, these standards are not sufficient because the phenomena to be explained is just how it can be that some feature is *not* publicly available. Cognitive science is in a fix here, since it is supposed to allow for methodological pluralism, and here the different methods seem to render incompatible results. This suggests that we should just make the leap and have it over with, but I still maintain that this leap is best done by way of postulating a fundamental law. Should phenomenal consciousness be granted fundamental status? To turn Dennett's argument against him: what could we possible lose on granting it this status? What we could gain is a systematic study of this postulated law, a serious attempt to try and find what kind of organization of matter that will give phenomenal consciousness. Denying it could in the worst case scenario lead to a replacement of consciousness studies with a neuroscience oblivious of this character. That is (with a borrowed term): what we have to gain is a Science of Consciousness as If Experience Mattered (Varela (1997)). If we think we have reason not to accept simple identity, and I say we have, then we are entitled not to be satisfied with anything short of a fundamental law.

Could there be a scientific explanation short of a philosophical explanation?

A not insubstantial part of philosophy is about scrutinizing the leaps taken and the bridges crossed in science. Questioning the nature of science, the nature of consciousness, the nature of causation, the nature of explanation etc. is part of what makes philosophy the interesting and, to some, irritating discipline that it is. And, of course, for every new domain claimed to be explained by science, or by whatever method, careful philosophical scrutiny should always be undertaken. Science and the "scientific method," if there is such a thing, have proven its worth and should, some might argue, be accepted on account of this. But, surely, one of the pillars of that respectability has been exactly that scientific explanations have never been accepted in advance. The whole point of science is that it proves it worth every time. Its openness to objections is part and parcel of the success-story to be told about science. The scrutiny of scientific explanations can be said to be divided into reasons: reasons to accept an explanation, and reasons to doubt it. In the case of phenomenal consciousness, I have tried to point to the reasons we have to doubt that a certain kind of explanation succeeds. I therefore find my self supporting the line taken by Chalmers in acknowledging these properties fundamental status. Now, I have argued that science in order to be unified and to account for all that "happens" need not deal with phenomenal consciousness, and thus not grant it this status. And that is exactly what I mean by claiming that there could be scientific explanations short of full philosophical explanation.

There is the possibility that there could be a philosophical standard that requires more than scientific methodology could possibly supply. Since phenomenal consciousness by argument lacks causal powers (if it doesn't help collapsing the wave-function, that is…) we do not need it in order to have a causally closed, coherent, neat unified science. Thus we might say that proper science only concerns itself with this causally closed picture of the universe. Baars (1997) writes that "in science, after all, we can only study something if we can treat it as a variable, comparing its presence to its absence." The problem, of course, is that we cannot separately study features that are connected to each other by natural law, if they can not be ascribed distinct causal powers. It is probably the case that the underlying structures will co-vary with phenomenal states, so that no treatment of qualia as a variable *ceteris paribus* is possible. So

if a unified science is all we strive for, no more fundamental entities need to be added. But still, something, namely the phenomenal feature of consciousness, will be missing from this picture. Science might very well turn the phenomenological strategy of "putting the world within brackets" on its head and instead bracket phenomenal consciousness. Explanation is then reached, but at a philosophically high price.

The point I want to make is that we have reasons to demand more of an explanation of consciousness because we have access to it by a different route from the usual order of explanation. Nothing short of reaching *that* could be very satisfactory. The sense of an explanatory gap is a product of the lack of ontological distance between our access and the thing that is accessed. In this sense, we already have all we need with regard to consciousness, what we lack is the connection to the rest of the world.

There is a problem of primacy here. Usually when we settle for a reductive explanation, we give the thing reduced a secondary status; its existence, we say, is derived from that to which it is reduced. But phenomenal consciousness is *epistemologically* primary, and to give it an *ontologically* secondary status needs to be thoroughly justified. Giving it a secondary status with regard to something that we are only familiar with by way of theoretical construction is hard to accept. It is plausible to treat atoms as more primary than chairs, despite the fact that we are more familiar with chairs, because it is conceivable how atoms could *amount* to chairs, and how chairs could be *divided* without residue to atoms. But the thing with phenomenal consciousness is that it is *not* conceivable how it could be the result of mere physical processes.

In our world, the physical facts plus the natural laws necessitate phenomenal consciousness, it is just that among those natural laws are the ones that connects function to phenomenality, and it is the treatment of this connection that usually differs between the reductivists and the non-reductivists[32]. The explanations we are given everyday often obstruct the assumptions needed to finish the explanations, leading to our being more and more oblivious to them. They are just there, preliminary unjustified. In the case of consciousness this is just what is disputed, and therefore should be taken up to the surface for scrutiny.

Given the same requirements we have on explanations in science, one might say, consciousness is likely to be explained. But we have independent reason to reject these explanations, or to consider them inconclusive. This is because the requirements given in science are adapted to handle properties of certain kinds. And, quite frankly, phenomenal consciousness is not one of these properties. In this case, and this case only, we have a kind of access that justifies the stronger requirement, or, rather: that justifies another scope for the stronger requirement. It is not the unbridgeable gap between physics and consciousness that accounts for this, but the *lack* of ontological distance between the observer and the observed in the case of phenomenal consciousness. All explanations leave a gap, and in all these other cases, we are right to jump them. But why should we jump the explanatory gap to consciousness when we are already immediately familiar with it? It is exactly this kind of familiarity that is so interesting about consciousness, and leaving it unexplained is not satisfying. Explanations, as the functionalistic ones, live up to the strong requirement only by denying the phenomena that is really in need of explanation. And only by invoking an irreducible law connecting these properties to functions, or whatever turns out to be the most likely candidate for the correlate of consciousness, can we be satisfied with the explanation given.

CONCLUSION

Evaluations of proposals given in the philosophy of mind have of late been focused on the hypothetical research program they may suggest. Failure to provide guidelines for such a research-program has come to function as a disqualification for a theory to be a theory of consciousness at all. The growing influence of this criterion seems to go hand in hand with the decreasing confidence given to arguments appealing to intuitions. Ideally, of course, a successful argument should make strong appeal to both. Failing that, a good argument should still be able to argue convincingly for the superior relevance of the favoured criteria. My main argument in this essay has been that the settlement of criteria and, in the long run, the preferred version of explanation, are not obviously subject-neutral.. The requirements settled for in science are usually correlated with what we believe suffices for attaining the *truth* about the subject at hand. That means that if we have reason to believe that the truth about the matter is not captured by merely living up to the requirements inherent in that particular kind of explanation, we have reason to doubt its success as explanation. I have claimed that the special character of phenomenal consciousness gives us reasons of this kind. I claim that this means that the problem of phenomenal consciousness is not so much a question *within* science, as it is *about* science, about its limits and its foundations.

Can phenomenal consciousness be scientifically explained? Obviously the answer to this question

[32] Van Gulick (1995) doubts this strategy and appeals to the fact that we are dealing with rather complicated scientific entities, and such rarely figure in fundamental laws. But, of course, the "simplicity requirement" can be though to apply to *the other side of the explanation*, in the qualia. *They* would more or less *have to* be connected in a fundamental way to functions.

boils down to what counts and what doesn't count as a scientific explanation. If "scientific explanation" means "explanation invoking only the properties available in present day science" I believe the answer is "no," but the same goes for many phenomena in the world. When dealing with the theory of consciousness, it is easy to get the impression that consciousness is the only thing left to explain in the world, and that all that remains is to connect the complete picture of the (rest of the) physical world with consciousness. This, of course, is not so. "Scientific explanation," then, should be every explanation that makes it conceivable how phenomena of one sort bring about phenomena of the other. The non-reductivist claims that this can only be done for consciousness by invoking fundamental psychophysical laws. As we have seen, this claim is based on the strong requirement on explanation in conjunction with a certain conception of what the explanandum is.

I have also claimed that we can expect a perfectly sensible scientific explanation to leave us with an intact philosophical quandary. A scientific theory of consciousness that serves us in all practical respects can still leave the philosophical problem unresolved. This would amount to accepting that the limit of scientific entities is not the limit of the world. Physics does not exhaust ontology[33]. Of course, *this* conception is based on accepting the weak requirements on *scientific* explanation.

Admittedly, this argument can be claimed to show nothing about the limits of science, and rather to be an argument establishing the irrelevance of philosophy. The non-reductionist position is ultimately founded on the treatment of hypothetical cases guided by conceptual analysis[34] and one might rightfully ask what this technique has ever done to merit explanatory/epistemic status. But this argument, I believe, is founded upon a profound misunderstanding of the nature, and origin, of explanation. The "scientific" conception of explanation is not once and for all given: it is fitting to its subject matter, but it should not be assumed in advance that phenomenal consciousness is part of that subject matter, since that would just be begging the question. But, as we have seen, the non-reductionist view treated here is also based on a predetermined notion of explanation, thus begging the question the other way.

Unfortunately, no knock-down argument is likely to settle this dilemma.

Is Explanatory Pluralism a viable option? What would be the implications for cognitive science if it where? Appealing to different standards for explanation in philosophy and in science is problematic, especially for a scientific discipline that is supposed to bring the two together under a friendly pretext. Cognitive Science, it has been said, is a discipline for philosophers with scientific ambitions and scientists with philosophical ambitions, and the methodological pluralism inherent in this discipline has been held to be it's perhaps most charming feature. The image of human consciousness is getting clearer every day when advances are made in all of these areas. But what would count as a complete picture? The proposal I have been trying to give in this essay is that the answer to this question restricts the possibilities of how a theory of consciousness might go. But I have also pointed to the fact that the more narrowly construed scientific explanations are an important part of the philosophical explanations strived for. The implication of the non-reductionist view, supported by the argument presented in this essay, is that there is a worthwhile project in trying to establish what functions are correlated to phenomenal consciousness. This project should be recognised to be worthwhile by most theories of consciousness. What differs is the interpretation of the results of this project. And, again, no knock-down arguments are likely to settle this matter either.

[33] As Stubenberg (1996) writes: Not physics but physicalism has made qualia into homeless properties.

[34] The Zombie argument is the paradigmatic exemplar of this strategy. This kind of philosophy of mind is part of a bigger project that one could call the "philosophy-of-science-fiction". Nothing short of conceptual analysis could help us much in dealing with controversial counterfactuals of this kind.

REFERENCES

Baars, Bernard J. 1996. 'Understanding Subjectivity: Global Workspace Theory and the Resurrection of the Observing Self' Journal of Consciousness Studies vol. 3

Baars, Bernard J. 1997. 'In the Theatre of Consciousness: Global Workspace Theory, A rigourous Scientific Theory of Consciousness'. Journal of Consciousness Studies vol. 4

Barnes, Eric 1994. 'Explaining Brute Facts'. Proceedings of the Biennial Meeting of the Philosophy of Science Association. Issue Volume One: Contributed Papers

Block, Ned 1995. 'What is Dennett's theory a Theory of?'. Philosophical Topics 22, 1 and 2

Block, Ned 2001. 'How Not to find the Neural Correlate of Consciousness'. The Foundations of Cognitive Science, Branquinho, Jao (ed)

Chalmers, David 1995 (3) 'Facing Up to the Problem of Consciousness'. Journal of Consciousness Studies vol. 2

Chalmers, David 1997. 'Moving Forward on the Problem of Consciousness'.Journal of Consciousness Studies vol. 4

Chalmers, David 1998. 'On the Search for the Neural Correlate of Consciousness'. Toward a Science of Consciousness II: The second Tucson Discussions and Debates. MIT Press

Chalmers, David 1996. The Conscious Mind. In search of a fundamental theory. Oxford University Press: New York

Chalmers, David (1) 'First-Person Methods in the Science of Consciousness' at www.u.arizona.edu/~chalmers/papers/firstperson.html

Chalmers, David 1993. 'Self-Ascription without Qualia: A Case-Study'. Behavioral and Brain Sciences 16:35-36

Churchland, Patricia Smith (1). 'What should we expect from a theory of consciousness?' www.phil.canterbury.ac.nz/tom_bestor/e-texts/

Churchland, Paul M. 1984 'Subjective Qualia from a Materialist Point of View'. Proceedings of the Biennial Meeting of the Philosophy of Science Association. Issue Volume Two: Symposia and Invited Papers

Dennett, Daniel 1993. Consciousness Explained. Penguin Books: London

Dennett, Daniel 1987. The Intentional Stance. MIT-press: Cambridge

Dennett, Daniel 2001a. 'The Zombic Hunch: Extinction of an Intuition?'. Philosophy. vol. 48

Dennett, Daniel 2001b 'The Fantasy of First-Person Science, a written version of a debate with David Chalmers, held at Northwestern University, Evanston, IL,. February 15, 2001, supplemented by an email debate with Alvin Goldman

Dennett, Daniel 2001 c. 'Are we Explaining Consciousness Yet?'. Cognition vol. 79

Dennett, Daniel 1996. 'Facing Backwards on the Problem of Consciousness'.Journal of Consciousness Studies vol. 3

Dennett, Daniel 1988. 'Quining Qualia'. Consciousness in Contemporary Science. Marcel and Bisiach (eds.) Clarendon Press: Oxford

Dewan, E.M. 1976. 'Consciousness as an Emergent Causal Agent in the Context of Control System Theory'. Consciousness and the Brain: a Scientific and Philosophical Inquiry. Globus, Maxwell and Savodnik (eds.). Plenum Press: New York

Gazzaniga, Michael, Ivry, Richard and Mangun, George 2002. Cognitive Neuroscience – the Biology of the Mind. Second edition. W.W. Norton & Company: London

Goldman, Alvin I. 1997 'Science, Publicity, and Consciousness'. Philosophy of Science. vol. 64 issue 4, December

Gothier, Marion 'Consciousness, Science, and the Nature of Explanation'. www.san-fran.com/consciousness/science.html

Van Gulick, Robert 1995. 'What Would Count as Explaining Consciousness?'. Conscious Experience. ed Thomas Metzinger. Schöning / Imprint Academic: Paderborn

Hacking, Ian 1983. Representing and Intervening – Introductory Topics in the Philosophy of Natural Science. Cambridge University Press

Harman, Gilbert H. 1965. 'The Inference to the Best Explanation'.The Philosophical Review vol. 74, Issue I, January

Jack, Anthony Ian and Ropestorff, Andreas 2002. 'Introspection and cognitive brain mapping: from stimulus-response to script-report'. Trends in Cognitive Sciences vol. 6 No. 8 August

Jackson, Frank 1982. 'Epiphenomenal Qualia'. Philosophical Quarterly no 32, April

Jackson, Frank 1998. From Metaphysics to Ethics, a Defence of Conceptual Analysis. Clarendon Press: Oxford

Koch, Cristoph and Crick, Francis 2001 'the Zombie within'. Nature Vol. 411, 21 June

Koch, Christoph and Crick, Francis 2003 'A framework for consciousness'. Nature neuroscience volume 6 no 2, February

Kripke, Saul 1975. 'Outline of a theory of truth'. Journal of Philosophy vol. 72

de Léon, David 1995. 'The limits of thought and the mind-body problem'. Lund University Cognitive Studies LUCS 42

de Léon, David 2001 'The Qualities of Qualia'.Communication & Cognition Vol. 34, Nr. 1&2

McGeer, Victoria 1996 'Is "Self-Knowledge" An Empircal Problem? Renegotiating the Spcae of Philosophical Explanation'. The Journal of Philosophy vol. 93 issue 10 October

McGinn, Colin 1995. 'Consciousness Evaded: Comments on Dennett'. Philosophical Perspectives Vol. 9 Issue AI

McGinn, Colin 1989. 'Can We Solve the Mind-Body Problem?'. Mind volume 98 issue 391, July

Mayes, Randolph G. The Internet Encyclopaedia of Philosophy: "Theories of Explanation", www.utm.edu/research/iep

Metzinger, Thomas 1995. 'The problem of consciousness'. Conscious Experience. ed. Thomas Metzinger, Schöning / Imprint Academic: Paderborn

Nagel, Thomas 1974. 'What is it like to be a bat?'. The Philosophical Review Vol. 84, issue 4, October

Nagel, Thomas. 1998. ‘Conceiving the Impossible and the mind-body problem’. Philosophy vol. 73 July

Nelkin, Norton 1993. ‘What is Consciousness?’. Philosophy of Science Vol. 60 issue 3, September

Rees, Geraint, Kreiman, Gabriel and Koch, Christoph 2002. ‘Neural Correlates of Consciousness in humans’. Nature Neuroscience volume 3, april

Rorty, Richard 1980. Philosophy and the Mirror of Nature. Princeton University Press: Princeton

Ryle, Gilbert 2000. The Concept of Mind. Penguin Classics: London

Searle, John 1992. The Rediscovery of Mind. MIT Press: Cambridge MA.

Shoemaker, Sydney 1991. ‘Qualia and Consciousness’. Mind Vol. 100 issue 4, October

Sperry, R.W 1980 ‘Mind-Brain Interaction: Mentalism, Yes; Dualism, No’. Neuroscience vol 5

Stubenberg, Leopold 1996. ‘The Place of Qualia in the World of Science’. Towards a Science of Consciousness: The First Tucson Discussions and Debates. Hameroff, Kaszniak and Scott (eds.). MIT Press/ Bradford Books: Cambridge MA.

Tomasello, Michael 2001. The Cultural Origins of Human Cognition. Harvard University Press: Cambridge MA.

Varela, Francisco 1997. ‘A Science of Consciousness as If Experience Mattered’. Towards A Science of Consciousness: The Second Tucson Discussions and Debates. Hameroff, Kazniak and Scott (eds.), MIT Press: Cambridge MA.

Varela, Francisco 1996. ‘Neurophenomenology: A methodological Remedy for the Hard Problem’. Journal of Consciousness Studies, June