Disputatio 10, May 2001

UNDERSTANDING CONSCIOUSNESS*

Isabel Góis
King's College London

> "Nothing is more chastening to human vanity than the realisation that the richness of our mental life — all our thoughts, feelings, emotions, even what we regard as our intimate self — arises exclusively from the activity of little wisps of protoplasm in the brain".
>
> V.S. Ramachandran and William Hirstein, "Three Laws of Qualia…", *Journal of Consciousness Studies,* Vol. 4, n. 5-6, 1997, pp. 429.

INTRODUCTION

Consciousness, one often reads, is a mystery — a phenomenon so utterly baffling as to defy our trust in the explanatory powers of science. But, is consciousness really a mystery? If it is, what is the source of the mystery, and why should it make us doubt that science may one day dissolve it? And if not, why are so many people convinced that consciousness is the most perplexing phenomenon we have come across? These are the questions motivating the present investigation, and as such they will shape and direct the main task I have set myself in this paper: To show that explaining consciousness is no harder, or different, a task from finding the answers to any other question we may care to ask about the mind. To be more precise, the claim I will be defending here is that what challenges our ability to explain consciousness is not some unshakeable mystery at the heart of this phenomenon, but the difficulty of dispelling a certain way of understanding consciousness that may be very tempting, but is also very wrong. As an alternative, I will propose an approach to the problem of consciousness which seems to me closer to a

* This article was first presented at the Portuguese Philosophical Society in October 1999. I am grateful to Desidério Murcho and two anonymous referees for comments and suggestions. I would also like to thank Conrad Scott for valuable discussion and difficult questions. Research for this article was generously funded by two grants from JNICT (BM 10514/97 and BM 21511/99).

scientific mode of inquiry and more capable of directing us towards a satisfactory solution.

I

Some philosophers and quite a few scientists are rather sceptical of our ability to explain how the workings of the brain can produce the unique shape in which our minds are aware of events around us or of thoughts and feelings within us. In particular, one idea that has gained much favour in the literature on consciousness is that science may well succeed in explaining several capacities and abilities of the mind in terms of brain processes, but explaining consciousness is (must be?) another matter. Why? Because something more puzzles us about consciousness than about any other mental ability and that is the very experience of being conscious. But this, so the argument goes, is precisely that distinguishing feature of consciousness that seems impossible to capture in the language of a scientific theory. In fact, one of the claims most often made is that, no matter how detailed and complete a theory of consciousness may be, it still won't be able to tell us why it is that conscious experiences alone have a seeming quality to their occurrence while unconscious processes don't. And, of course, from here to the conclusion that science can never break into the secret chambers of consciousness is only a small jump. Consciousness, so these philosophers like to tell you, is a mystery and, moreover, one not likely to be disclosed in the details of a scientific theory. A conclusion that will appeal to many, no doubt, but one I do not share. My belief is that this inference rests on mistaken assumptions concerning the nature of our conscious experiences, and that as soon as we undermine whatever charm such assumptions may have, the subjective features of our experiences will seem much less puzzling and impenetrable. However, before going into that, let us make sure that we are clear on why it is that consciousness seems to defy any attempt to explain its nature.

One way to put the problem is this. Consciousness, as we experience it from within, is commonly seen as a translucent medium in which the contents of our minds come, as it were, to the surface and appear in a unique, private, and often intangible way to each one of us. Thus, although we all talk about our feelings, ideas, thoughts, desires, imaginings, and other contents of consciousness, none of us can share with another the very same conscious experiences. However, if we are to investigate the nature of consciousness, the data we are to use in support of our hypothesis must be publicly available and objectively described. To put it in other words, if an understanding of consciousness is at all possible, the challenge is in explaining its nature by the same methods and with the same terms used in the natural sciences.

But, precisely, here is the problem. How do we combine the standard scientific practice of couching explanations exclusively in terms of facts (and

laws) we can objectively examine with the circumstance that facts about conscious experiences (e.g., their occurrence and content) are available to the outside observer only through the subjective descriptions of those who are undergoing them? That is, how can we objectively explain what it is to consciously appreciate, say, a warm sunny morning without leaving the very subjectivity of this experience outside the explanation ? Yet, if we are to explain our subjective experiences in scientific terms, that seems to be exactly what will happen. Our conscious experiences, at least as they appear to us, seem so utterly unlike anything scientists may put under their microscopes that the prospect of having only an impersonal language of variables, measurements and experimental manipulation to describe and explain all the variety of our conscious experiences certainly raises the doubt whether such language is indeed adequate to account for their subjective character. And indeed, at first sight, such doubt appears to be justified. If the point of a scientific account of consciousness is to show how subjective experiences can be explicitly equated with specific processes in the brain, then we should be able to extract the phenomenological features of our conscious experiences from evidence concerning those same processes. However, it seems clear that as soon as we change the language of mental predicates for the language of brain processes we lose the ability to include those phenomenological features in our theories and, thus, subjectivity itself. But, again, if we cannot spell out in objective terms what subjective experiences are like to those who have them, then no matter how much we may have explained concerning the physical basis of consciousness, we still have to explain everything that needs explaining about consciousness, namely its phenomenology.

I agree that the task seems nearly hopeless. Yet, it is my contention in this paper that this appearance of ‘hopelessness’ is more the result of certain mistaken beliefs concerning our conscious experiences than any particular difficulty in explaining consciousness with the tools of science, or in understanding how the machinery of the brain can give rise to conscious experiences. But more on this later. For now, let us try to determine precisely what it is about the phenomenological properties of consciousness that allegedly makes them impermeable to a scientific approach.

The classic way of introducing phenomenological properties into the problem of consciousness was given by Thomas Nagel who made himself famous with the claim that these properties correspond to what it is like to be conscious of this or that. In his article “What it is like to be a bat”(1995), Nagel put forward the thesis that it is not in our power to explain how consciousness can arise of the brain’s activity, since the subjective aspects of conscious experiences cannot be apprehended in the objective language of science. That is, we have no means to render intelligible the correlation between consciousness and the brain, because we have no idea whatsoever how to

portray subjectivity from an objective point of view without failing to describe *what it is like* to have conscious experience. The argument sounds familiar by now, but notice that what bears the weight of Nagel's thesis is, on the one hand, the link between a conscious experience and its subject's point of view and, on the other hand, the idea that there is something called an experience's subjective character, which is to say, that there are common intrinsic properties associated with a type of experience which constitute the "what it is like to have it". Thus, if asked "what it is like objectively to experience subjectively x?", the answer is: we cannot say. Science acknowledges no other than the non-perspectival language of objective investigation, in which points of view have no place. And since there is no way around the interconnection between episodes of awareness and the point of view which frames them, attempting to explain consciousness with the tools of science is trying to do the impossible. The two standpoints — objective and subjective — conflict with each other irredeemably: put yourself in one and you will have abandoned the other. So, unless we figure out a way of revolutionising our whole conceptual apparatus — a very unlikely possibility, according to Nagel — explaining what it is like to be conscious is and will remain beyond our power.

Attractive as this argument may seem, there are some considerations we should attend to before, as it were, committing our heart to it. The bulk of Nagel's thesis is that experiential facts are only accessible from the point of view of the subject they belong to. Therefore, we cannot place ourselves in a point of view different in kind from our own, and reproduce the mental context attached to it. Now, if this means that, as in his own example, we cannot turn our minds into bat-minds and have the same experiences as bats do, we must surely grant him the point. But does this also mean that we cannot possibly conceive what it is like to be a bat? I think it doesn't. For one thing, if nothing we can discover from a third-person point of view can tell us what it is like to be a bat, what reason do we have to think that there is something it is like? How can we even suspect, that is, that there are intrinsic, private properties of bat-experience? On the other hand, if what Nagel is really asking is what can we know about the kinds of minds bats have, then investigating the structure of the bat's perceptual and behavioural system, discovering what it is aware of and under what circumstances, should help us answer the question "what it is like to be a bat?" (Dennett 1991, pp. 446-7)

For sure it is tempting to argue against the latter idea that imagining what it would be like to be a bat is not a genuine experience of what it is like to be one. But, then again, the aim of science is not to create bat-consciousness (Nature already did that), nor to allow us to simulate it. Rather, it is to give us a theory of how it works. And doing just this is what Nagel's argument doesn't show is well beyond the powers of science. To insist that it is, seems to me to be the same as slipping from the difficulty of conceiving a mind alien to ours to the claim that it is epistemologically impossible to know the kinds of experi-

ences available to that mind. The latter idea just doesn't follow from the former. Accordingly, until an independent argument is given to the effect that the subjective nature of consciousness is not a physical property of organisms — which Nagel explicitly avoids giving — it remains (at least) an open question whether experiential features are indeed so mysterious that science cannot do justice to them.

Now, Frank Jackson (1997) and David Chalmers (1996) have made precisely the move of arguing that the phenomenological aspects of consciousness are non-physical properties of our conscious experiences and, therefore, science can tell us nothing about them. In fact, these properties are such that you simply cannot know anything about them unless you have the experiences themselves. Notice that these philosophers are speaking about the same properties that philosophical tradition, taking after Nagel, baptised as the 'qualia' of consciousness, the inner feels of awareness, the unmistakable quality that permeates the way in which a mental state is also a conscious one. What distinguishes their claim is the explicit commitment to the non-physical nature of these properties. Accordingly, the verdict they pass on the possibility of explaining the phenomenology of consciousness by appeal to facts about the brain isn't just that this is an unfeasible project within current science; rather, their conclusion is that pursuing such project is committing ourselves to falsity.

Frank Jackson (1997) has illustrated this view with the life-story of Mary, a thought experiment intended to bring out clearly what is missing in empirical accounts of consciousness. Here is the story. Mary is a brilliant neurophysiologist who lives confined to a world where the only colours available to her are black and white. To make this more plausible, suppose she was born with a rare sort of damage in her optical nerves. What is special about Mary is that she knows all the physical facts there is to know about colour experiences, even though she never had colour experiences herself. Mary, however, is also a lucky girl: a state-of-the-art technique in neuro-surgery reverses the damage in her optical nerves, and opens for her a 'world of rainbows'. Now, after the operation, she not only knows all the physical facts about colour experiences her science books could tell her, she can also experience what it is like to see the redness of roses, the yellow of Van Gogh's sunflowers, the greenness of a praying mantis. According to Jackson, this proves that Mary learned something new, something she could not have learned in science books, namely, what it is like to be visually aware of colours. But, so Jackson concludes, if the thought-experiment proves this, then it also proves that there are non-physical facts about colour experiences (otherwise, how could Mary learn something new?), and scientific explanations of consciousness are false in as much as they say that there aren't.

This, then, is Jackson's story about Mary. But how good a story is it, and how forceful is the conclusion on us? To begin with, let us query the assump-

tion that Mary learns something new. Isn't this question begging? After all, if Jackson says that Mary's problem is that she never had visual experiences because of the damage in her optical nerves, how does this prove that she learns something new after the operation? Why not say, instead, that having gone through surgery, Mary acquired further neural capacity to react to stimulatory cues and, thus, has gained new neural sensitivity but no new knowledge? Jackson's story doesn't eliminate this interpretation and, thus, we are not forced to grant him that new knowledge comes to Mary through experience. But, even if we agree that Mary learns from experience what she did not know before from books, why should this make us conclude that after the operation she gains access to some non-physical facts about herself? If Mary does indeed learn something new in virtue of having her optical nerves operated on, it seems more appropriate to conclude that this operation has given her the ability to recognize typical differences in colours in more than one way, where before she could only recognize those distinctions in one sort of evidence (supposedly, on the basis of evidence concerning the different properties of light when reflected by surfaces). That is, where before the spectrum of colour experiences available to Mary only included black and white, after the operation she has a broader spectrum of colour experiences, and thus more than one way of referring to colours. Still, how this can fail to be knowledge of a new physical fact (in this case, a fact about the scope of her conceptual abilities), is something I don't see that Jackson's argument can prove (Lewis 1997).

Now that we know why some philosophers have taken phenomenological properties to be impermeable to empirical investigation, it is worth exploring in connection with both arguments presented here why such properties should deserve to be accorded special treatment. After all, we've been taking it for granted that it seems to make sense to talk about conscious experiences in terms of these properties, but 'seems to make sense' doesn't necessarily mean that it 'makes sense'. We should, therefore, ask the question: what is the evidence for assigning empirical meaning to the claim that such properties exist and determine the subjective character of our conscious experiences? That it isn't any sort of objective evidence is part of the above philosophers' argument, so let us pursue the matter fairly and emphasize that what the question is asking is if there is a way leading to something that would count as decisive proof that such properties are unmistakably present.

As Nagel's and Jackson's arguments make clear, all the relevant evidence for phenomenological properties comes from what subjects report about their conscious experiences. What is more, that is all the evidence we need, since what these reports tell us is precisely what it is like for the subject to be conscious. (Remember that Nagel (1974) says that no amount of objective evidence will ever give you a hint about what it is like to be conscious, and Jackson goes even further in claiming that subjective properties

— 'qualia', to be more precise in his case — have no physical causes or effects (Jackson 1982).) So, it seems that if we want to know precisely what it is like to be conscious of, say, the contrast of black and white in this page what you have to do is pay attention to your experience and the specific quality of that experience will immediately become obvious. Or not? It is true that we normally take individuals to be privileged authorities on their conscious experiences, but it is also a well known fact that introspective reports of subjective experiences are not immune to error, and that more often than not subjects, as it were, rearrange the facts when reporting on what it is like to be aware of something. Thus, we are all familiar with cases where subjects report being aware of more than they actually were, and even failing to report being aware of things that they in fact were conscious of. Also, our everyday life is full of examples when prior descriptions of what it was like to be conscious of something are either corrected or dismissed (e.g., you report that a certain sauce tastes sweet and later say it's actually sour), and others when we simply cannot tell what it is like to be conscious of some event despite the fact that we are sure to be aware of it (e.g., what it is like to be yourself?). It follows, then, that being owners of our conscious experiences doesn't necessarily makes us privileged authorities on what it is like to have them[1]. But if introspective reports cannot be taken as accurate descriptions of what it is like to be conscious of this or that, then what does it mean to say that conscious experiences have an unmistakable qualitative feel to them? Clearly, it cannot mean that phenomenological properties are what properly constitute our conscious experiences although subjects may be wrong about them, since that would be laying the 'red carpet' for science to come in and claim to be in a better position to explain what it is like to be conscious than subjects themselves. But also, it cannot mean that phenomenological properties are intrinsic to consciousness but sometimes subjects are incapable of reporting their distinguishing mark, for then we undermine the only evidence we had for the existence of such obvious properties. The only option left is to say that these properties are simply mysterious, and beyond that nothing more can be said. However, if this is what someone chooses to say (like Nagel and Jackson do), than I think we have every reason to doubt that there is any determinate empirical content to the claim that there is a specific feel associated with each conscious experience, since even those who believe in its truth can say nothing more than that is what they believe. To put it in fewer words, the point is that introspective reports provide no reliable evidence that these 'spooky' properties to which philosophers attribute so much importance are anything more than a figment created by our ordinary way of characterizing conscious experiences. And since Nagel and Jackson are the first ones to say that no other evidence is available, then I see no reason to believe that there is

[1] See Nisbett and Wilson (1977).

anything so special about conscious experiences that may prevent us from building an empirical theory of consciousness.

Now, I have little doubt that many people find the latter claim about conscious experiences simply outrageous. Given the importance that subjectivity has in our life, and the fact that few of us would care to go on living if they could no longer enjoy their experiences, it is easy to think that whenever someone says that there is nothing special about consciousness, what they are really saying is that there is nothing special about our lives. This, however, is not what I'm saying. Nor am I trying to deny that conscious experiences have subjective features that clearly distinguish my experiences from yours. What the objections raised against Nagel and Jackson are meant to remind us of is that we are not privileged authorities on our conscious experiences and, therefore, often come to entertain mistaken beliefs about what we are conscious of and what we aren't. What is more important, it is precisely because we are prone to have mistaken beliefs about our conscious experiences that striving for an empirical account of consciousness makes sense, and gives us a better chance to understand this amazing capacity our minds to be aware of the world and themselves. Thus, in defending that there is nothing mysterious about consciousness I am not saying that we should, so to speak, give up our souls. But I am saying that if we want to value our minds for what they really are, we must learn through empirical investigation what they are truly made of.

Now, I agree that none of what has been said so far provides any assurance that science can indeed explain consciousness. As we saw above, what really motivates the view that consciousness is (must be) a mystery is the certainty that scientific explanations cannot reduce the difference between consciousness as seen 'from the inside' and consciousness as looked at 'from the outside'. It seems plausible to argue, then, that until we can look at what the man in the lab-coat supposedly has to show us about consciousness, and see the very same experiences we have while conscious, no one has any good reason to believe that consciousness can be scientifically explained. However, as I see it, this is not a plausible demand to make on scientific theories, whether they're trying to explain consciousness, or the molecular structure of carbon dioxide. Science is meant to tell us what physical conditions in the world make possible which observable phenomena, not to prove that every observable effect of those conditions can also be seen in their causes (why should it?). To demand such proof is to misunderstand the aims of scientific explanation, and, worse, to ignore the fact that not every observable phenomenon is worth saving in our theories (namely, because those observations are mere 'tricks of the eye' and the alleged phenomenon can be explained away). Still, not to be accused of destroying other people's hopes in vain, let me make clear why the above 'proof' is not a reasonable demand to make on a theory of consciousness.

What people like Ned Block (1997) and Chalmers (1996) require of a suitable theory of consciousness is for it to render transparent the experiential features of consciousness from the postulated properties, processes or functions the theory identifies them with. That is, presented with the latter we should be able to read off the former with the aid of such theory. Accordingly, none other than an identity of the kind 'conscious state x = physical / causal role y' (delete as appropriate to the target theory) would fulfil the explanatory standards of those sharing the intuition mentioned above. Such identities, however, are reputed to fail on counter-factual grounds: you can always imagine some possible situation where one side of the identity is present without the other also being present, or present in the same ways as in the actual world. Hence, suppose that in the actual world anxiety is identified with the release of some neurochemical substance. Can't you easily imagine a situation in which the same compound is released and no anxiety is felt? And can't you as easily imagine a situation in which you feel anxious but some other chemical or none at all is discovered to be associated with what you feel? If you can, then, for all their merits, scientific theories of consciousness go no further than establishing brute correspondences between introspective reports and empirical evidence about the brain that always fall short of grounding sufficient and necessary conditions for conscious experience.

As I see it, there is a lot to be said against the idea that scientific theories must make their identity claims transparent or plainly obvious. To begin with, let us grant — for the sake of argument — that the core criticism embedded in the argument is correct: straightforward identities between conscious mental states and physical / functional states (again, delete as appropriate to the target theory) do not withstand the test of counter-factual reasoning. Must we conclude that we do not have, and probably will never acquire, the proper conceptual tools to develop a scientific account of consciousness? That is, must consciousness remain a mystery? As we've just seen, some will claim this to be so, but I wish to argue that although there is indeed a lesson to be drawn from the embarrassments of crude forms of reductive materialism, such lesson in no way entails the utter mysteriousness of consciousness. Consider the assumption that scientific explanations must make it obvious why the flanking terms of an identity claim are said to be co-referential. Why should this be a reasonable standard to hold? Supposedly, because such transparency is what allows us to see that, no matter with which side of the identity we work, the same conclusions are deducible from the other. But this can't be right; why must it be obvious that the same conclusions follow? And obvious to whom? What is demanded of a scientific theory is that it gives a true explanation of the phenomena falling under its scope, not that it should make it obvious why that explanation is true. Of course, if the explanation is not only true but also 'obvious', then we are fortunate for the bonus; but if it isn't, the theory that proposes it is no less true just because we find it hard to

understand its explanations. Hence, it can hardly be an objection to an empirical account of consciousness if it doesn't make it obvious how you can deduce conscious experiences from the relevant physical evidence, as long as it allows us to do it.

A different line of objection to this same standard can be generated by considering the claim that successful empirical accounts of consciousness must endorse strict identity claims. Both the proponent of psycho-physical / psycho-functional identities and those who object to them presuppose that it is a clear cut matter what we're referring to when we say, for example, 'I'm angry'. Accordingly, they both agree that the aim of an empirical account of consciousness is to build a sort of theoretical *cerebroscope* which would allow us to look at pictures of the brain and see there faithful portraits of conscious experiences. Now, this would sound like a good idea, were it not for the tacit — and mistaken — assumption that our idioms for conscious experiences are theoretically in order to sustain such identities. It is true that our ordinary ways of singling out conscious phenomena trick us with the conceptual illusion that we are successfully referring to well-defined theoretical attributes and well-behaved real entities. But any quick survey of our common-sense taxonomy of mental states soon reveals the error of taking common sense too literally. For, even if we wish to discount the fact that introspective reports are no reliable guide to the underlying ontology of conscious mental states (and I don't see how we could get away with that), it's still true that what we ordinarily refer to as 'being conscious of' is more like a covering term for different sorts of psychological states, many satisfying more than one phenomenological description (often, even contradictory ones), and each too vaguely defined to delimit a set of conditions telling us what to classify as conscious. In other words, our common-sense understanding of conscious mental states, by itself, does not permit us to establish a definite description of any of its phenomenological categories as a way of fixing its reference, and, therefore, cannot serve as an authoritative constraint on the kinds of theories of consciousness we should frame. Thus, although it is true that identity-definitions of conscious experiences are bound to meet with counter-examples of one kind or another, this in no way necessitates the conclusion that consciousness cannot be explained in scientific terms. Quite on the contrary, what this shows is that we should avoid committing ourselves to a dubious taxonomy of conscious mental states, and let empirical investigation teach us how to carve the mind at its joints.

So, to sum up the argument so far, my claim is that philosophers and scientists have for too long taken for granted that, since each individual must know better than anyone else what being conscious of x or z feels like to him (since he is the first to know what is going through his mind), science should either provide physical evidence for the truth of those introspective impressions, or bow to the conclusion that it can never know our minds as well as

they know themselves. Faced with the difficulty of pursuing the former task, most prefer to subscribe the latter conclusion. Consciousness, they say, is the most astonishing mystery we have yet come across because we cannot imagine how to explain it in naturalistic terms. On the contrary, I have argued that it is time we call our assumptions and deep-seated intuitions about conscious experiences into question and establish their true value, before jumping into the conclusion that consciousness cannot be explained by empirical methods. More, I have argued that once we realize that the so-called mystery of consciousness is entirely due to a persistent mythology around our notion of introspection, there is no reason to think that consciousness cannot be explained scientifically. Rather, there is every reason to believe that empirical investigation will put us in a better position to appreciate the wonders of our conscious minds. In what follows, I will draw some methodological proposals which I believe set us on the right track towards a scientific account of consciousness and clear the way for a proper understanding of our subjective experiences.

II

Supposing you agree that until now no one has presented a conclusive argument against the possibility of a scientific study of consciousness, for sure there is still some doubt in your mind regarding how such a project can be accomplished with any possibility of success. After all, it is much easier to say that — in principle — science can explain the nature of consciousness, and quite another to show how that can be done. In fact, as we saw before, one of the main reasons why so many philosophers and scientists doubt that we will one day arrive at a satisfactory explanation of consciousness is precisely because they are convinced that we cannot even imagine how the brain, locked away in the dark of our skulls and having only electrochemical impulses to guide its work, can produce all the variety and liveliness of our subjective experiences. And how many of us are not tempted to agree with them? For, think about it, can you really conceive the idea that a red-brownish (not gray!) mass of convoluted brain tissue is home to all the variety of your conscious experiences? In other words, can you really make sense of the fact that all our conscious thoughts and feelings just are neuronal processes going on in our brain when our subjective experiences, at least as we describe them to ourselves and others, look nothing like a brainwave or feel nothing like a bunch of neurochemicals rushing across our brain? Where, then, is the persuasiveness of a thought, the urge of a wish, the uneasiness of anxiety, the expectancy of our hopes or the shivering of our fears, the ‘weirdness’ of our dreams and the painfulness of hurt? For sure one doesn’t have to believe in some ghostly properties not captured in brain-scans to see that there is a difference between observing consciousness ‘from the outside’

and experiencing it 'from the inside'. So, even if you agree that consciousness has everything to do with the relevant cerebral activity occurring at the appropriate places in the brain, for sure you will also agree that to explain how one thing leads to the other is going to be extremely difficult, if not downright impossible. How do I propose to 'bridge the gap' then?

It seems clear that one of the most important concerns in starting an empirical investigation of consciousness is to get the right initial foothold. An immediate problem is to determine what needs explaining in a manner which is both neutral and non-question-begging. At first sight, this might seem like an easy task: what we need is a theory of consciousness which explains its causal ancestry in the brain. This, however, tells us nothing about the specific phenomena that fall within the scope of such a theory and, thus, leaves us without criteria with which to measure its success. In other words, we may know that in order to understand the mind we need to know how the brain functions, but without an idea of what sort of phenomena those functions are supposed to be underlying, little can be learned about the brain that might help us understand consciousness (or the mind). Unfortunately there seems to be no easy way to settle on a list of conscious phenomena that is both comprehensive and neutral. Most examples of conscious experiences collected from our everyday talk tie them too closely to paradigms of self-consciousness and this, besides begging the question against what a theory of consciousness should be a theory of, isn't of much help either. We are all happy enough to say that sometimes we are aware of, say, doing something that we are not aware that we are aware of doing. Accordingly, although self-consciousness is without question one of the phenomena to be included in an explanation of consciousness, there is no reason to think that it is the only one, nor to take it as the defining criteria of what it is to be consciously experiencing x, y or z. To make things only more difficult, there is no clear definition of consciousness that may assist us in deciding what falls in and out of a prospective theory. Giving examples of conscious experiences is easy, but as soon as we try to pin down as neatly as possible what consciousness is very soon we find ourselves at a loss for words. And if we do find the words, quickly enough we also find that one person's obvious description of consciousness is to another an obscure characterisation of it. The problem here is that our pre-theoretical ways of referring to consciousness are too messy and vague to constitute an availing ground for scientific explorations on the nature of consciousness. How, then, can we determine what needs explaining?

Clearly, not by conceptual analysis alone. Philosophers have been trying for centuries to make sense of the cluster of conflicting categories that make up our ordinary understanding of consciousness, and it is safe to say the armchair inspiration has until now shed but little light on the subject. More importantly, the difficulties mentioned in the above paragraph make it clear

that there aren't any apparent conceptual facts of the matter about consciousness so indubitably true as to secure a pre-theoretical answer to this question. But, is there an alternative? There is one: empirical research. That is, my conviction is that it is a mistake to try to solve this question by conceptual analysis. Instead, we should see it as an empirical matter to be settled by empirical investigation. Does this also answer our concern of setting scientific research on consciousness on the right foothold? As far as I can see, this is the only good answer. Not only does it spare us the interminable battles of intuitions that have fuelled philosophical debates for so long, it also gives us a better chance of avoiding misconceptions of the phenomena in question. To put it clearly then, the claim I am making is that we stand a better chance of understanding consciousness if we start with a rather minimal conception of it, and let empirical research help us refine a better one.

Supposing the reader grants the above point, two questions seem nevertheless to raise difficulties for an empirical study of consciousness. First of all, if I am right in claiming that subjective reports on conscious experiences can't be taken at face value, what data is there to support testable hypotheses? And second, how can we even begin to apply the scientific method to something so vague and ill-defined as the above assumption? Clearly, the answer to the first question is to find a way of gaining access to conscious experiences in a manner that is scientifically respectable (Dennett 1991; Baars 1988). Since, admittedly, these experiences are only directly accessible from a first-person perspective, the problem is how to go from introspective reports to an objective characterization of subjective experiences without prejudging the accuracy or information contained in those reports. As to the second question, the best way to capitalise on it is to unpack it in a set of specific questions which are both basic enough to allow a rudimentary classification of conscious phenomena and, at the same time, empirically addressable (Churchland 1995; Hirst 1995; Johnson-Laird 1987). Thus, instead of taking the above assumption and ask: "What is consciousness?", my proposal is that we leave this question on the side, and ask instead more specific questions which have the advantage of being both pivotal to issues about consciousness and empirically tractable. In other words, what I am suggesting here is a sort of basic guide to the study of consciousness, which permits us to establish facts about subjective experiences in terms of actual experiments and observations. The way I see it, if we put these two strategies together, we have a solid empirical entry point into the 'secret chambers' of consciousness. But before arguing for that, let me present each of them in more detail.

How, then, can we collect reliable evidence on conscious experiences without failing to meet the appropriate standards of scientific investigation? In his book *Consciousness Explained* (1991, pp. 72-98), Daniel Dennett suggested a method which he dubbed heterophenomenology, and which seems to me quite appropriate to the task at hand. Broadly stated, the heterophe-

nomenological method is applied to an experimental situation in three stages. First, we make multiple recordings of the entire experiment (these will range from videotapes to electroencephalograms, passing through all the standard ways of monitoring what goes on 'in the lab'). Second, several written transcripts of those recordings are elaborated and compared, which then serve to fix an objective description of the data collected during the experiment. The third, and most crucial stage, comes when those transcripts are interpreted as speech acts (pp.76-77). That is, we take the uttered 'noises' and 'bodily movements' and interpret them "…as things the subjects wanted to say, of propositions they meant to assert, for instance, for various reasons."(p.76) For this, we must adopt the Intentional Stance towards the subject's — verbal and non-verbal — behaviour, and interpret it as exhibiting intentionality. In other words, we attribute to him beliefs, desires, and so forth, which explain why he behaves the way he does (e.g., the subject says 'The light moved from left to right' because he believes that it moved in that direction.) What warrants the inference from 'noises' and 'bodily movements' to conclusions about what is in the subject's mind? First, the assumption that the subject is a rational agent who acts in accordance with the beliefs and desires that can safely be attributed to him. In fact, if we couldn't trust that subjects (in general) act in a rational manner (and, consequently, that their actions are intentional), there was no point in conducting experiments since their behaviour would be totally inscrutable, and therefore unintelligible. Second, the fact that we take care not to attribute to him more than those beliefs and desires which would suffice to make the best sense of what he does and says. That is, the inference from 'noises' and 'bodily movements' to intentional content is guaranteed by our concern to keep our interpretations as deflationary as possible (see Dennett 1971, 1978, 1987).

This, then, is the heterophenomenological method in action, and how we can collect objective evidence on the alleged content of subjective experiences. Why alleged? First, because the method is supposed to be nothing more than an objective description of what subjects intend to say and do. Hence, it makes no claim regarding the accuracy or truth of what subjects affirm to be the content of their experiences. Second, because until we can confirm that something close enough to what subjects describe is going on in their brains, those introspective reports remain without independent validity. That is, without evidence that neural states have properties sufficiently similar to the properties of the experiences described by subjects, those descriptions are better understood as fictions (Dennett 1991, pp.78-81). And, precisely, the advantage of treating the intentional contents of conscious experiences as fictitious entities is that it allows us to establish (provisional) facts about subjective experiences, while avoiding inflating those facts with claims that go beyond the evidence (namely, the claim that 'heterophenomenological items' correspond to real happenings in people's brains.)

Now, before the objection is made that this is not a proper way to treat mental states, let me advance the following question. What are the grounds for assuming that the putative categories of folk-psychological understanding are the correct way to individuate contentful states in the brain? Put in other words, what reason do we have to think that folksy, mental-like states exist in the brain? I can think of no other but one arising from an a priori commitment to the existence of an ontology underlying common sense psychology. In other words, a desire to say that there really are mental states, the proof being that we are conscious of them. How 'being conscious' could provide such a proof, is something I cannot even guess. After all, let us not forget that from the single fact that subject reports to be aware of this or that, little can be deduced about their actual experiences. As I have often insisted, we may be the only source of data on subjective experiences, but we are no privileged authorities on the contents of our conscious minds. Accordingly, it would be precipitate of us to put much confidence in the validity of introspective reports concerning the working of one's own mind. More importantly, we should bear in mind that some of the distinctions between mental states that can be drawn within our mentalistic folk vocabulary, may prove to be mere 'linguistic artifacts' when looked at from a neurophysiological point of view. In other words, there may be no discernible distinction between the brain processes that underlie your thinking that, say, '867542 is a big number' and your thinking that '867543 is a big number' (Dennett 1991, p.319; Greenfield 1995, pp.154-162.) Of course, it is possible that talk in terms of 'beliefs', 'desires', and so forth, provides a good approximation to whatever it is that is going on in our brains that adds up to consciousness. But, to me, it seems preferable to err on the side of caution and avoid premature conclusions, than make the opposite error of presuming the existence of such entities in the brain and, then, having no means of finding evidence to back up that assumption.

So, now that we have a method to collect the data, what are the right questions to ask? That is, what sort of facts should we try to find out about conscious experiences? From what has been said, it is clear that an empirical study of consciousness must involve a set of questions concerning the nature of consciousness, which are both relevant for establishing the scope of a theory of consciousness, and amenable to experimental investigation. How do we know which questions are relevant? By taking into consideration what we would need to know about consciousness in order to determine which properties go with it. That is, by trying to hit on facts that can help us specify which aspects of consciousness must be accounted for by any given theory. On the other hand, since these questions must also be empirically decidable, it is obvious that we must elaborate them in such a way as to allow actual testing and observation. Now, an immediate difficulty of this project is that questions aimed directly at consciousness have no obvious domain of response (remember that our ordinary concept of consciousness is too ill-

defined to sustain scientific inquiry). Fortunately, this difficulty can be circumvented by appealing to the contrast-class of awareness. In other words, we can design questions that allow us to circumscribe the domain of conscious phenomena by contrast with non-conscious mental events (Baars 1988.) In what follows, I will propose a short list of such questions with a brief commentary on what they are intended to illuminate. To be sure, I make no claim that this list is exhaustive, nor that the specific formulations I propose are the best ones to adopt. My aim is purely to give an idea of what I consider to be a solid empirical approach to the subjective facts of conscious experiences.

Now, the first question can be put the following way: 'What are subjects conscious of?' This is meant to account for the distinction between what individuals can and cannot be aware of. The range of issues covered under this topic include: (a) the distinction between 'internal' and 'external' experiences (that is, awareness of states of the world, and awareness of inner states); (b) the distinction between awareness of processes and awareness of the products of those processes; (c) the distinction between what one can become aware of and what not. One important point to emphasise about this question is that the distinctions it is meant to illuminate should not be presumed to translate sharp contrasts between what falls in and out of a subject's awareness. It is possible that awareness 'comes' in degrees (sometimes we are fully-aware of x, sometimes we are less-aware), as well as for transitions between consciousness and non-consciousness to be graded (remember, for example, the moments before falling asleep). Plus, we should not presume that absence of introspective awareness amounts to absence of awareness tout-court. Reportability is only one among other indicators of consciousness (action-readiness and behavioural integration are others).

A similar question to the one above can be raised to account for what one can and cannot consciously control. Issues of attention (including selective attention), and voluntary action fall clearly under the scope of the distinctions we would be here trying to draw. For example, there are moments when we cannot avoid but paying attention to some event that distracts us from our current occupation, and others when we cannot stop thinking about something. On the other hand, there are items in the environment around us that 'immediately' capture our attention (e.g., the screech of car tyres), while others pass by without us even noticing them (e.g., the cacophony of sounds during the rush hour). Also, some behaviours (presumably) are the product of conscious decisions, while others — as much as we would like to control them — evade not only our powers of decision, but also escape our awareness (e.g., when you swing your leg while sitting in the dentist's waiting room). Again, remember that the distinctions brought forward under this question may not correspond to sharp divisions in what falls under conscious control. In particular, we need to be careful with shadowy boundaries between

so-called automated skills and complex behaviours allegedly requiring conscious monitoring.

Third in place, comes the question of self-awareness. The contrast intended to be illuminated here is that between what we can and cannot be aware of concerning ourselves. Specific questions include awareness and non-awareness of bodily states, 'known' and 'unknown' motivations, and the limits of introspection (both in the sense of accuracy, and in the sense of a reflective capacity on one's own awareness). Of particular interest is the problem of determining whether self-awareness goes beyond the information given. In other words, if in being aware that they are aware, people 'fill in the gaps' of their conscious experiences. As mentioned before, individuals sometimes (unwillingly) report being aware of more than they actually are, while on other occasions reporting less. Plus, they often entertain mistaken or simply false beliefs about their own subjective experiences (as when they believe to be giving an 'honest answer' to the experimenter's questions, when in fact they give what they believe to be the desired answer). Accordingly, one of the aims of the question is to determine the extent to which self-consciousness reflects these unwitting 'constructions'. Another point of interest is to inquire after what kinds of information people attend to when aware of themselves being aware. The aim here would be to establish as far as possible the extent to which self-awareness facilitates or lessens our ability to carry out certain activities (including our ability to engage in social functioning).

The fourth, and last, question is the following: 'Is there a distinction between information that is consciously entertained from information that is not?' Put perhaps more clearly, is there a difference between (say) a memory consciously entertained, and that same memory held 'in the back' of the mind? One of the issues intended to be illuminated here is the extent to which the 'personal context' in which particular conscious experiences are inevitably embedded affects the very content of that experience. That is, the point that interests us here is to know if and how the experiential links that bind subjective experiences to prior experiences affect the tone and content of occurring experiences. The other issue is to know how much past experience can influence present behaviour without a person being aware of it. These two issues are closely related, but it is important not to ignore the difference of focus. The first asks for the influence of prior information on occurring experiences when that information is consciously called to mind. The second asks for the same influence when that information is not included in the subject's awareness.

These, then, are some of the question that we can use to design experimental situations and, consequently, determine the scope of an empirical theory of consciousness. If we now combine the careful use of the heterophenomenological method with the neutrality of these questions, you will see

that we have gained an empirical entry point into the subjective aspects of consciousness without giving up the third-person point of view, nor availing ourselves of any assumption that cannot be supported by empirical evidence. What is more, with these two tools at hand, we can set up a research program that guarantees that none of the important features of consciousness will be left out of an empirical account of its properties, while at the same time measuring up to the standards of scientific investigation. To be clear, there is here no promise that scientific research will confirm our common sense understanding of consciousness. Nor is there any promise that in a future science of the mind/brain, consciousness will stand out as the 'mind-pearl' to be found hidden in the workings of an 'oyster-brain'. However, taking into consideration the arguments presented in this paper, it is easy to see that such promises have no place in a scientific attempt to explain consciousness. Put in other words, if we want to take seriously the task of breaking down the riddle of consciousness, we must put aside old habits of thinking about consciousness, and accept that some of the questions that we would like to see answered are born of misguided curiosity. Still, at the end of this chapter, I hope that you'll agree that giving up the hope of seeing those promises carried out is a price worth paying, since to understand how our brains can engender a mind of their own cannot but deepen our appreciation for this fabric of *nice tricks and beautiful stories* that we call consciousness. To be sure, it is always possible that neither we nor future generations will ever be able to fully understand consciousness. True; but like all things in life, there is no failure as long as we keep trying.

Isabel Góis
Dep. of Philosophy
King's College London
Strand, London WC2R 2LS
United Kingdom
isabel.gois@kcl.ac.uk

References

Baars, B. (1988) *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Block, N. (1997) "On a Confusion About a Function of Consciousness" in Block, Flannagan, Guzeldere (eds.) *The Nature of Consciousness: Philosophical Debates*. Cambridge: Cambridge University Press, pp. 375-416.

Chalmers, D. (1996) *The Conscious Mind*. Oxford: Oxford University Press.

Churchland, P. S. (1988) “Reductionism and The Neurobiological Basis of Consciousness” in Marcel and Bisiach (eds.) *Consciousness in Contemporary Science*. Oxford: Oxford University Press, pp. 273-304.

—— (1997) “Can Neurobiology Teach Anything About Consciousness” in Block, Flannagan, Guzeldere (eds.) *The Nature of Consciousness: Philosophical Debates*. Cambridge: Cambridge University Press, pp. 127-140.

Dennett, D. (1971) “Intentional Systems”, *Journal of Philosophy*, 68, pp. 87-106.

—— (1978) *Brainstorms*. Montgomery, VT: Bradford Books.

—— (1987) *The Intentional Stance*. Cambridge, Mass.: MIT Press.

—— (1988) “Quining Qualia” in Marcel and Bisiach (eds.) *Consciousness in Contemporary Science*. Oxford: Oxford University Press, pp. 42-77.

—— (1991) *Consciousness Explained.* Middlesex: Penguin Books.

Greenfield, S. (1995) *Journey to the Centers of the Mind*. New York: Freeman and Company.

Hirst, W. (1996) “Cognitive Aspects of Consciousness” in Gazzaniga (ed.) *The Cognitive Neurosciences*. Cambridge-Mass.: The MIT Press, pp. 1307-1320.

Jackson, F. (1982) “The Epiphenomenal Qualia”, *Philosophical Quarterly*, 32, pp. 127.136.

—— (1997) “What Mary Didn’t Know” in Block, Flannagan, Guzeldere (eds.) *The Nature of Consciousness: Philosophical Debates*. Cambridge: Cambridge University Press, pp.567-570.

Johnson-Laird, P. (1987) “How Could Consciousness Arise From the Computations of the Brain?” in Greenfield and Blakemore (eds.) *Mindwaves*. New York: Blackwell, pp.247-258.

Lewis, D. (1997) “What Experience Teaches” in Block, Flannagan, Guzeldere (eds.) *The Nature of Consciousness: Philosophical Debates*. Cambridge: Cambridge University Press, pp. 579-596.

Nagel, T. (1974) “What is it Like To Be A Bat?” in *Philosophical Review,* 83, pp. 435-450.

Nisbett and Wilson (1977) “Telling More Than We Can Know: verbal reports on mental processes” in *Psychological Review*, 84, pp. 231-259.