The genuine problem of consciousness

Anthony I. Jack[1]*, Philip Robbins[2], Andreas Roepstorff[3]

1. Department of Neurology, Washington University in St. Louis, USA
2. Department of Philosophy, Washington University in St. Louis, USA
3. Centre for Functionally Integrative Neuroscience, Aarhus University, Denmark

*Corresponding author: Anthony Jack, Dept. Neurology, Washington University in St Louis, 4525 Scott Avenue, St Louis, MO 63110, USA; tony.jack@gmail.com

Abstract

Those who are optimistic about the prospects of a science of consciousness, and those who believe that it lies beyond the reach of standard scientific methods, have something in common: both groups view consciousness as posing a special challenge for science. In this paper, we take a close look at the nature of this challenge. We show that popular conceptions of the problem of consciousness, epitomized by David Chalmers' formulation of the 'hard problem', can be best explained as a cognitive illusion, which arises as a by-product of our cognitive architecture. We present evidence from numerous sources to support our claim that we have a specialized system for thinking about phenomenal states, and that an inhibitory relationship exists between this system and the system we use to think about physical mechanisms. Even though the 'hard problem' is an illusion, unfortunately it appears that our cognitive architecture forces a closely related problem upon us. The 'genuine problem' of consciousness shares many features with the hard problem, and it also represents a special challenge for psychology. Nonetheless, researchers should be careful not to mistake the hard problem for the genuine problem, since the strategies appropriate for dealing with these problems differ in important respects.

1. The 'hard problem' and its evasion

In the mid-1990's, the philosopher David Chalmers wrote an influential paper entitled "Facing up to the Problem of Consciousness." In it, he observed:

> The really hard problem of consciousness is the problem of experience. When we think and perceive, there is a whir of information-processing, but there is also a subjective aspect. As Nagel (1974) has put it, there is something it is like to be a conscious organism. This subjective aspect is experience. When we see, for example, we experience visual sensations: the felt quality of redness, the experience of dark and light, the quality of depth in a visual field. (Chalmers, 1995)

Chalmers went on to accuse scientists investigating consciousness of bad faith. For, despite pretensions to the contrary, their work typically misses its intended target:

> It is common to see a paper on consciousness begin with an invocation of the mystery of consciousness, noting the strange intangibility and ineffability of subjectivity, and worrying that so far we have no theory of the phenomenon. Here, the topic is clearly the hard problem — the problem of experience. In the second half of the paper, the tone becomes more optimistic, and the author's own theory of consciousness is outlined. Upon examination, this theory turns out to be a theory of one of the more straightforward phenomena — of reportability, of introspective access, or whatever. At the close, the author declares that consciousness has turned out to be tractable after all, but the reader is left feeling like the victim of a bait-and-switch. The hard problem remains untouched. (ibid.)

As a description of current scientific work on consciousness, Chalmers' words still ring true. There are many interesting and important scientific issues that relate to consciousness: how the brain constructs complex representations of the world, what differentiates conscious from unconscious mental states, how the vegetative state differs from normal wakefulness, and so on. Yet, even as we make progress on these issues, there is a sense in which we seem to be moving not closer, but further away from the fundamental issue which the field set out to address.

The problem of consciousness is merely the latest incarnation of an ancient problem: the mind-body problem. Since Plato, thinkers have struggled to understand the relationship between mind and brain. In the field of consciousness studies, the most widespread view is that neuroscience will yield the answers. The rallying cry of this perspective was nicely captured by Nobel laureate Francis Crick: "You're nothing but a pack of neurons" (Crick, 1994). The idea is that mind and brain are one and the same, so once we have explained the brain, the problem of consciousness will have been solved. The philosopher Leibniz realized the inadequacy of this reductionist approach almost three centuries ago, when he wrote:

> Moreover, we must confess that the *perception*, and that which depends on it, is *inexplicable in terms of mechanical reasons*, that is, through shapes and motions. If we imagine that there is a machine whose structure makes it think, sense, and have perceptions, we could conceive it enlarged, keeping the same proportions, so that we could enter into it, as one enters into a mill. Assuming that, when inspecting its interior, we will only find parts that push one another, and we will never find anything to explain a perception. (Leibniz 1714/1989, §17)

If we are going to make better progress, we need to take a different tack. The reductionist strategy has been defended, in one guise or another, many times over the last few decades, most notably by philosophers such as Dennett (1991) and the Churchlands (P. M. Churchland 1985, P. S. Churchland 1997). Despite the ingenuity of some of this work, and the publicity it has received, it hasn't won many converts. We need to recognize that there is a problem of consciousness which won't go away so easily.

We also need to recognize the importance of facing up to the problem. For decades scientists have claimed that there is no need for them to clearly define the topic they are working on. They have refused to explain how their work addresses the central issue of subjectivity, or they have provided facile and unconvincing explanations. They have suggested that, as work progresses, either the solution will become apparent or the problem will simply disappear. In short, they have used every strategy to avoid addressing the problem directly. As a result, the field is highly fragmented, and

researchers have no common framework within which they can relate their work to that of others.

Finally, and most important of all, we need to understand the true nature of the problem. Getting good answers depends on asking the right questions. A good formulation of the problem of consciousness should meet two criteria. First, the problem should be specified in a manner that makes it clear what counts as progress. A meaningful enquiry is one that is guided by questions with recognizable answers. Second, the problem should be stated in a form that doesn't involve any unnecessary or questionable assumptions. Any assumptions that are made should be acknowledged and they should be testable, at least in principle.

Chalmers' critique of scientific theories of consciousness is aimed at the first of these criteria. As he observes, it is hard to see how the proposed theory addresses the larger issue that the author refers to at the start. It is clear that scientific work on consciousness suffers because the first criterion has not been met. There are widespread disagreements and confusion about what should count as progress in the science of consciousness. The degree to which scientific research meets the second criterion, the criterion of being based on sound assumptions, is harder to address, since researchers refuse to clearly formulate the question they are working on. Nonetheless, as we will shortly return to, there are reasons to suspect that a great deal of work is motivated by a view of the problem that makes dubious assumptions.

How does the 'hard problem' fair with respect to these two criteria? Chalmers explicitly addresses the first criterion: it is just the fact that we have no idea how to answer it that is supposed to make the hard problem so hard. Experience, it appears, isn't subject to reductive explanation. So Chalmers proposes a dualist theory of consciousness, a theory that treats experience as an irreducible feature of the world. Of course, this proposal doesn't do anything to solve the hard problem, and it isn't meant to. Chalmers doesn't treat the hard problem as something that one should set out to solve; rather he interprets the intractable nature of the problem as evidence for dualism.

If Chalmers is right, then it is a pretty sad result. His take on the problem forces us to posit a sphere of phenomena entirely distinct from those that have been studied by science to date. That isn't parsimonious. Furthermore, it isn't clear how progress can be made by studying this alternative realm. Chalmers suggests we can proceed the old-fashioned philosophical way, via *a priori* conceptual analysis and clever thought-experiments. Given the track record of this methodology, that seems unlikely. But even if we are willing to follow him down this path, the hard problem will persist.

We can start the search for a more productive formulation of the problem by looking at what leads Chalmers to state the problem as he does. Chalmers elegantly spells out the underlying assumption: we need an extra ingredient to explain consciousness. The idea is that standard information-processing accounts can, at least in principle, explain how we come to form complex representations of the world and how those representations are used to guide what we say and do, but they also leave something out: the subjective aspect of experience. To account for that, it seems, we need an extra ingredient — something beyond mere information processing, something that will add color to the line drawings.

The 'extra ingredient' view of consciousness is widespread among both philosophers and scientists. For instance, the philosopher Ned Block takes a similar view when he suggests a distinction between two concepts of consciousness: 'access consciousness' and 'phenomenal consciousness' (Block, 1995). According to Block, most scientific work concerns itself with access consciousness, but the real prize lies in explaining phenomenal consciousness. Again, the claim is that something more is needed to account for phenomenal consciousness, something over and above what is needed to explain access consciousness. Scientists are rarely as explicit as philosophers about what they understand consciousness to be. Nonetheless, it isn't hard to see that many scientific theories of consciousness are aimed at providing the extra ingredient. They conceive of this ingredient in different ways. Some see it as a specific process, such as 40-Hz oscillations or the collapse of the quantum wave equation in microtubules.

Recently the trend has been more towards seeing it as an emergent property that arises under specific conditions. For instance, Giulio Tononi (2005) has developed a mathematical formula designed to describe the degree of informational complexity of systems, a factor which he claims is critical for the emergence of consciousness.

The critical point about the 'extra ingredient' view of consciousness is that it rests on the assumption that what we are looking for is something out in the world. Some process or feature must explain why we are phenomenally conscious, not just access-conscious. What we need to do, then, is to uncover the biological basis of phenomenal consciousness (Miller, 2005). Unfortunately, the scientific quest for this 'extra ingredient' looks like a fool's errand, for two reasons. First, the sorts of features and processes which science is capable of discovering just aren't the sorts of features and processes that could possibly shed light on subjectivity. That is the point that Chalmers makes, and we won't belabor it any further here. Second, there is a better way to account for our intuitions about consciousness, one which is both more parsimonious and which finds better empirical support. We don't need to look for any objective process or feature to distinguish a complex information-processing machine from a conscious agent, because the crucial difference doesn't lie with the object we are thinking about, it lies in how we are thinking about it. The evidence suggests that there is a genuine gap between the physical and the mental, but that gap exists in the apparatus we use to think about these things, not in the things themselves.

The problem of consciousness is driven by a powerful intuition, which is captured by the (in)famous philosophical thought-experiment concerning a 'philosophical zombie'. We can imagine a system that does all the same information processing that we do yet which lacks phenomenal experience. Many philosophers and scientists believe that this intuition accurately reflects reality. They think that the origin of the problem of consciousness is a feature of the world, the 'extra ingredient' that explains the difference between a zombie and a conscious being. Intuitions are important. They often guide great scientific discoveries. But they are also fallible. Although they often reflect how the world is, they don't always. Sometimes they are more like visual illusions: errors in how

we see the world that arise as by-products of the heuristics we use to see. In the Müller-Lyer illusion, for example, we see two lines of equal length as differing in length because we mistake their relative depth. Illusions like these tell us more about our cognitive architecture than they do about how things actually are in the world. Likewise, there are good reasons to think that the true origins of our intuitions about consciousness reside not (as it were) in the stars, but in ourselves.

2. The genuine problem of consciousness

> *"'Thoughts' and 'things' are names for two sorts of object, which common sense will always find contrasted and will always practically oppose each other."*
> James (1904): Does 'Consciousness' Exist

Consider the following thought-experiment (Jack & Shallice, 2001). Imagine a very sophisticated robot. We will call it 'René', after the great proponent of dualism, the philosopher René Descartes. René lives in a rich environment containing numerous complex objects, including other robots that are very similar to René. René has been designed to learn about the world for himself by forming his own concepts. In order to simplify this learning process, his creators split René's systems into two categories. One system is dedicated to understanding the physical world and to mechanical reasoning. This system processes information about the environment that René lives in and all the objects he encounters. It allows René to figure out how to negotiate the environment and exploit it to his advantage. The second system processes information about other robots that René encounters, and it also processes information about René's own internal states. René is constructed so that he has internal access to a limited amount of information about his own inner workings. The second system is dedicated to two different, but equally important processes. The first is social interaction. It controls how René deals with other robots. The second is self-representation. René has been endowed with the capacity to develop a dynamic model of how external conditions affect his internal states, as well as how internal conditions influence what he can do. This model helps René to

foresee the effects the world will have on him and that he will have on the world, so allowing him to self-regulate and plan. These two processes are dealt with by the same system because this takes advantage of the similarities between René and other robots. René can use his model of himself as a rough model for how other robots will behave, and he can also use his model of other robots as a rough model for himself. René's interactions with other robots are also facilitated by his self-model in other ways. For example, it helps him to understand how he is perceived by others, and allows him to influence (you might say, manipulate) those perceptions by telling narratives about why he has acted as he has (Boyer, Robbins, & Jack, 2005).

René learns from his experiences, and he develops elaborate conceptual systems both for understanding the physical world around him and for understanding himself and other robots. However, because these two systems process different kinds of information and serve different purposes, they have developed quite independently. René has no way of relating these two conceptual systems to each other. He has learned to recognize and categorize many of his own internal states, yet he cannot recognize them or categorize them using the physical or mechanical concepts he has developed. When René thinks about himself and other robots on the one hand, and objects in the physical environment on the other hand, he sees them through such different lenses that they seem to him to exist in different metaphysical worlds. As a result, René is inclined to believe these things really do belong to different worlds. He is inclined to dualism. And even if we convince René to believe in physicalism, the view that his internal states are physical states, his understanding of this identity will be very limited. The philosopher Thomas Nagel (1974) captures the point very elegantly in his landmark essay 'What is it like to be a bat?':

> Usually, when we are told that *X* is *Y* we know *how* it is supposed to be true, but that depends on a conceptual or theoretical background and is not conveyed by the 'is' alone. We know how both '*X*' and '*Y*' refer, and the kinds of things to which they refer, and we have a rough idea how the two referential paths might converge on a single thing, be it an object, a person, a process, an event, or whatever. But when the two terms of the identification are very disparate it may not be so clear how it could be true … At the present time the status of physicalism is similar to that which the hypothesis that matter is energy would have had if uttered by a pre-Socratic philosopher. We do not have the beginnings of a conception of how it might be true. (Nagel, 1974)

There is growing evidence that our own evolved cognitive architecture is much like René's. We will only provide a sketch of some of the evidence here; readers can find it discussed in greater detail elsewhere (Bloom, 2004; Robbins & Jack, 2006). One of the first indications of a separation between the cognitive processes involved in these two types of thinking came from individuals with autism, who show deficits in social reasoning but not in mechanical reasoning (Baron-Cohen, 2000). Further evidence to this effect comes from functional neuroimaging (Gallagher, Jack, Roepstorff, & Frith, 2002). Similar brain regions are involved in thinking about ourselves and in thinking about others (Frith & Frith, 1999), and these regions differ from the regions that are active during a broad range of cognitive tasks requiring problem solving and attention to the external world (Duncan & Owen, 2000). A recent study is particularly telling, showing that regions devoted to external attention are negatively correlated with regions devoted to internal attention, even while the subject is resting (Fox et al., 2005). This and other evidence (again, see Robbins & Jack, 2006) suggests that there may be an inhibitory relationship between these two modes of thinking: when we think about minds we tend to turn off those parts of our brain that we use to think about physical mechanisms, and vice versa.

It is, of course, a complicated matter to relate brain activation to how we think. The quick analysis we offer here raises as many questions as it answers. Nonetheless, the parallel is striking. One of the most intractable intellectual puzzles ever known concerns the relation between the mind and the body. Now we discover that one of the clearest patterns that can be discerned in the brain is a spontaneous tendency for activity in regions involved in attention to the external world to be negatively correlated with activity in regions involved in attention to self. What this suggests is that the 'extra ingredient' approach to consciousness does in fact rest on an illusion. The problem of consciousness, the mind-body problem, only looks like a problem about two types of thing. It is really a problem about two types of thinking. Searching for the biological basis of consciousness is like searching for the biological basis of beauty. Certainly the physical structure of an object is what makes it beautiful, but beauty cannot be captured or understood by examining the physical structure of objects. To understand beauty you

need to look in the eye of the beholder. The same thing is true of understanding consciousness.

What does this mean for the science of consciousness? Is everything lost? Not at all. Certainly it means that the problem of consciousness isn't what many people have supposed. Our reformulation of the problem lacks the romantic allure of a quest for nature's special secret, that deeply hidden property or process that is supposed to magically explain consciousness. But it has other features to compensate.

First of all, it should be clear that scientific research can shed light on the origins of the problem of consciousness. This is a momentous fact. For literally *millennia* we have relied almost exclusively on philosophers to clarify the mind-body problem. Now we can see how research by cognitive neuroscientists on the structure of cognition — specifically, on how we construct representations of minds, on the one hand, and physical mechanisms, on the other — will prove essential to fully understanding the problem. This research raises a number of interesting and important questions. For instance, to what degree is the separation between these conceptual systems innate? How do these systems break down into sub-components? Do some sub-components interact and overlap more readily than others? These and related questions can be asked and addressed at both at the level of brain architecture, by functional imaging, and at the cognitive level, by behavioral methods.

We have made a small start on some of these issues, reviewing evidence that thinking about minds splits into two types (Robbins & Jack, 2006). The first type involves thinking of complex systems as agents whose behaviors depend on what they believe and what they want, which we call the 'intentional stance' after Dennett (1981). The second type involves thinking of complex systems as conscious, feeling beings, which we have dubbed the 'phenomenal stance'. The intentional stance helps us to compete and co-operate with others, whereas the phenomenal stance plays a special role in social bonding and moral thinking. When we think about an object as a physical mechanism, by contrast, we are taking the 'physical stance.' One interesting observation

is that we don't seem to have too much difficulty taking the intentional stance and the physical stance towards the same object. For instance, we don't hesitate from talking about our computers in intentional terms, using phrases like "it is trying to print" and "it decided to erase my document". There seems to be a greater tension involved in taking the physical stance and the phenomenal stance towards the same object. This tension contributes to our sense of the problem of consciousness as a particularly stubborn aspect of the mind-body problem. It is also plays a role in damaging social interactions. Thus, the psychologist Paul Bloom has observed that social groups often emphasize the base physical features of persecuted minorities (describing them as dirty, for instance) as a precursor to treating them in a manner that denies their moral status as conscious feeling beings (Bloom, 2004).

Scientific research does not represent an exclusive means by which we can arrive at a better understanding of the problem of consciousness, although arguably it provides the most solid grounding. A number of philosophers have arrived at a closely related conclusion through a priori reflection on the nature of phenomenal concepts (Loar, 1997; Tye, 1999). Chalmers (1995) suggests that we face up to the hard problem of consciousness by accepting experience as a basic and irreducible feature of the world. There is already a philosophical precedent for arriving at a more modest and parsimonious conclusion, which still acknowledges the problem: experiential concepts are basic and can't be reduced to physical concepts. By looking at how the problem of consciousness may originate from our cognitive architecture, we provide empirical support for this conclusion. More importantly, our approach also helps to draw attention to the implications of this conclusion for the science of the mind.

Consider Nagel's take on the problem, captured by the quotation above. Nagel suggests that our grasp of the idea that the mind is the body is similar to the grasp that someone ignorant of modern physics would have of the idea that matter is energy. Indeed, on the surface, the problems look similar. In both cases the identification seems mysterious because we don't know how the disparate concepts relate to each other. But deep down, the problems are very different. In the case of problems in physics, even

when the concepts are disparate, they nonetheless belong to the same larger network of concepts that are grounded in the same way. Both matter and energy are concepts that help us to make sense of the external world. When physicists tell us that matter and energy are the same, they are telling us that we can describe the world most accurately if we revise our network of concepts so that these two concepts co-refer. Merely telling us that matter and energy are identical isn't enough to achieve this, hence the air of mystery that initially surrounds the identification. If we were to take the trouble to understand the theoretical framework that gives this result, however, that would suffice.

In the case of the mind-body problem, integrating the concepts is not merely a matter of revising a single network. The concepts we use to understand minds, as opposed to that the concepts we use to understand physical mechanisms, appear to be parts of disparate networks, grounded in disparate ways. The most obvious and significant difference is that the concepts we use to understand minds don't just help us to comprehend a class of external phenomena that are publicly observable (namely, overt behaviors); they also help us to comprehend a class of internal phenomena that are not observable in this way (namely, experiences). Our understanding of our own and other minds is partly grounded in a type of information that plays no role in modern science, namely, internally available information about our current perceptual, cognitive and affective states.

It might seem that the problem of consciousness should be solvable by constructing a theoretical framework in which phenomenal concepts and their physical counterparts merge (P. S. Churchland, 1996). But this proposal overlooks the possibility that the two frameworks are incommensurable, at least to some extent. If that were true, it wouldn't be too surprising. After all, these frameworks evolved for different purposes, and they deal with very different types of thing. Though there seem to be areas in which phenomenal and physical concepts can be brought into close proximity, there also seem to be areas in which they will forever remain miles apart. For instance, it is likely that we will be able to find a smooth translation between the phenomenal properties that differentiate different types of pain, and mechanistic accounts of those different pain

states. However it is hard to see how mechanistic accounts could ever express another aspect of pain: that is its essential badness (in the evaluative sense). The badness of pain is something we know immediately from the first-person perspective, but it is essentially absent from the distant third-person perspective of mechanism. Furthermore, an appreciation of the badness of pain of is not something we can do without. Understanding it is essential to effective self-regulation. You won't make good plans if you don't make a point of avoiding pain. It is also essential to successful social interaction. You won't make lasting friends if you fail to see their pain as bad. As such, the realization that pain is bad forms part of our bedrock understanding of the world from a personal (including an interpersonal) perspective. Mechanistic accounts can, in principle, explain what is going on in our brains when we are thinking pain is bad. But they can't explain the most important fact that we need to know in order to function in the world, namely the badness of pain itself.

The genuine problem of consciousness is a problem about explanation, but it isn't the sort of problem that can be solved by a theory of consciousness. We have two different ways of understanding the mind: we can understand it as a physical mechanism, and we can understand it from a personal perspective. The problem is that contemporary scientific psychology aims almost exclusively at mechanistic explanations of the mind. This is, ironically, no less true of most supposed scientific theories of consciousness than it is of the regular business of experimental psychology and cognitive neuroscience. Yet, for reasons both intellectual and practical, mechanistic explanation is not enough on its own. We can't understand the mind unless we can understand it for ourselves, from our own personal-level perspective. If we are right that physical and phenomenal concepts belong to fundamentally distinct networks, then it is a problem that may never be definitively resolved. Nonetheless, it is a problem we can make progress on, for even if these networks always remain distinct, they can still be integrated into a more coherent whole. The genuine problem of consciousness is the challenge of achieving this large-scale integration of our conceptual scheme.

Why should we care about this problem? More specifically, why should the *science* of psychology concern itself with the personal level of understanding? There are several reasons. First, it is the personal level that we primarily care about. With chronically ill patients, it is the level of pain they experience that concerns us. Only by extension do we concern ourselves with their galvanic skin response, their cortisol levels, or the activation of their anterior cingulate. Second, consider how the science of the mind might guide interventions that improve our lives. One of the most interesting features of accounts that work at the personal level is that they can serve to directly alter mental function by changing how we self-regulate. For instance, cognitive therapy is a method that works by encouraging subjects to observe their own experiences and to think about them in more productive ways. It has long been a treatment of choice for anxiety-related disorders, and it has recently been shown to be highly effective in treating depression (Teasdale et al., 2002). Personal-level explanations and training protocols represent both the most humane and the most publicly acceptable means of intervention that the science of the mind can hope to offer. Finally, the personal level is important because it can inform interventions that proceed at the genetic, neural and pharmacological levels. The availability of interventions at these levels is likely to increase exponentially as neuroscience progresses. In order to provide adequate informed consent for these procedures, we shall need to be able to explain the impact they will have on how people experience their everyday life. It seems unlikely that explaining the neural and behavioral consequences of interventions in brain function will prove adequate. Thus the genuine problem of consciousness, the problem of integrating personal-level and mechanistic ways of understanding the mind, is not merely an abstract intellectual issue. The degree to which we make progress on it, or continue to ignore it, is likely to have a very real social impact.

3. Reintroducing the subject

> "*The plans which I most favor for psychology lead practically to the ignoring of consciousness*."
> Watson (1913): Psychology as the Behaviorist Views it

How should we go about integrating personal-level and mechanistic ways of understanding the mind? At least part of the answer is obvious: we need to incorporate the subject's point of view back into psychology. That is, we need to find ways of relating our mechanistic understanding of mental function to the personal-level understanding that we use to make sense of our experience. The details of how best to do this are not so obvious. One good starting point is to recognize that, while nowadays psychology and neuroscience are primarily focused on the mechanistic explanation of mental phenomena, this hasn't always been the case. To see this, it helps to consider two major schools of thought, the introspectionists and the psychophysicists.

The introspectionists, such as Edward Titchener, Wilhelm Wundt, and William James, were primarily focused on understanding experience. Here is a quote from James:

> Most books adopt the so-called synthetic method. Starting with 'simple ideas of sensation', and regarding these as so many atoms, they proceed to build up the higher states of mind out of their 'association', 'integration', or 'fusion', as houses are built by the agglutination of bricks. This has the didactic advantages which the synthetic method usually has. But it commits one beforehand to the very questionable theory that our higher states of consciousness are compounds of units; a student who loves the fullness of human nature will prefer to follow the 'analytic' method, and to begin with the most concrete facts, those with which he has a daily acquaintance in his own inner life. (James, 1892)

Titchener and Wundt practiced the synthetic method. Their project famously failed because they could not agree on what the fundamental components of experience were, and there appeared to be no way to decide between their accounts. Like James, we are skeptical of approaches that involve complex analyses of experience in its own terms. James's writings were primarily descriptive; he was concerned with getting the phenomenology right. For instance, in his *Principles of Psychology* (James, 1890), he engages in a long and complex debate concerning the phenomenology of intentional action. On page after page he carefully considers alternative accounts, thoroughly dissects the exact nature of the experience, and presents arguments for his own account. James appreciated not just the importance of understanding experience, but also the work

that was needed to do it. Such attention to phenomenology has long been out of fashion in psychology. Only a handful of the vast number of papers now published in experimental psychology and cognitive neuroscience make any attempt whatsoever to describe the experience of the subject. Where some reference to phenomenology is made, it is almost never put forward as a matter for discussion. We have only rarely read or heard a cognitive scientist suggest that further work is needed to understand what it is like to carry out an experimental task. Authors are expected to devote most of a published paper to reasoning through, justifying, and hedging their interpretations of objective data. Only rarey do they occupy more than a few lines attempting to reason through, justify, or hedge their descriptions of phenomenology.

Next, let us consider psychophysics, which is widely credited with establishing psychology as a science. As the name implies, the idea was to relate the mental to the physical. Around 1834, Ernst Weber and his student Gustav Fechner published work aiming to relate the intensity of sensation to the physical magnitude of the stimulus (initially, felt weight to physical mass). However, they did this in a somewhat roundabout way. They looked at the accuracy with which two nearly identical stimuli can be discriminated, and how that varied with overall magnitude. The idea was that felt intensity could be measured in units of 'just noticeable differences'. It is a remarkably elegant and plausible account, which first gave credence to the idea of a quantitative psychology, and which has captivated generations of psychology students.

Unfortunately, it isn't right. There is an incorrect assumption: that each 'just noticeable difference' corresponds to an identical unit of felt intensity. The failure of the Weber-Fechner law for describing felt intensity wasn't conclusively demonstrated until much later, when Stanley Stevens used a rating-scale measure to ask subjects more directly about the character of their experience. Stevens' power law relates stimulus intensity to felt intensity. Although the Weber-Fechner law failed at its intended purpose, it nonetheless represents an important result. It tells us how much information is transmitted at different intensities — a goal much closer to that of current research.

The story of psychophysics holds important lessons. It is tempting to think that a particular relationship exists between experience and some physical measure. This idea is widespread in contemporary psychology (for examples, see Jack & Roepstorff, 2002; Jack & Shallice, 2001). These claims often sound plausible, but on closer examination they turn out to be false. Arriving at an accurate description of experience, even for the relatively simple task of relating one specific aspect of experience to the physical properties of the stimulus, requires a great deal more work. It requires careful measurements, using well-designed measures that have been assessed for accuracy and validity. More generally, it requires something that is sorely lacking in the current Anglo-American tradition of psychological research, and that is a scientific culture that is sensitive to issues surrounding the measurement of experience. The importance of an active, informed and questioning scientific culture should not be underestimated. Consider that, while we had to wait for 1957 for Stevens to disconfirm the Weber-Fechner law empirically, James already saw in 1890 that they had missed the mark. James saw the Weber-Fechner law as an accurate generalization as to the friction in the neural machinery.

How is it that James recognized the disparity between mechanistic accounts and accounts of experience more than a century ago, while psychologists today find it hard to do the same? Why are we (usually) so cavalier in our attempts to account for experience? We do not have space here to go into the complex history, but two points about current practice are of particular note. The first is that there is a widespread belief that introspective reports are unreliable or unscientific. Psychologists prefer ‘objective’ measures, such as discrimination performance, because they seem more trustworthy than so-called ‘subjective measures’, such as direct reports of experience. They don’t realize that, when their target is subjective experience, objective measures are less direct than subjective measures, and hence less likely to be valid. The second is that psychology is still guided by the reductionist view that understanding the brain suffices for understanding the mind. As a result, psychologists commonly fail to realize that they are making an assumption when they substitute objective measures for subjective ones, and so they do not even seek to justify the claim that their objective measure serves as an

accurate proxy for experience. (For an alternative view, see Dennett, 2003; in rebuttal, see Jack & Roepstorff, 2003.)).

At the moment there is a tendency to think of consciousness as an isolated issue, a specific problem that cognitive neuroscience will solve before moving on to other things. Yet when we consider the nature of the problem, and its connection to the history of psychology, it seems clear that the problem of consciousness is foundational for psychology as a discipline. If cognitive neuroscientists wish to provide full and satisfactory explanations of the phenomena they study, then they cannot avoid working on the genuine problem of consciousness. If good progress is to be made, there needs to be a shift in the scientific culture towards one that recognizes the importance of taking account of the subject's point of view.

A number of factors will facilitate this shift. Here we have focused on providing motivation for this change, by outlining why this is an important issue. Another important factor is that of overcoming current prejudices against the use of introspective reports, by explaining the misconceptions that have led to the widespread belief that introspective reports are not to be trusted. Perhaps the most significant factor will be providing solid examples of empirical research that illuminates the links between mechanistic and personal-level modes of understanding. When one looks at the relatively sparse examples of work that fits in this category, two things are clear.

First, there are many ways of integrating concepts of the mental with concepts of the physical. Charting how our perceptual experiences relate to external stimuli (psychophysics) and how they relate to our internal neural states (the neural correlates of consciousness) is just the beginning. We should also seek to answer questions like the following. What generalizations about experience most accurately capture the results of detailed observations? How does our phenomenology reflect the information we have about our own internal states, and how can that information be used to guide our thinking (metacognition)? What neural and cognitive processes are involved in attending to and forming concepts about our internal states? What mechanisms underlie the formation of

introspective judgments concerning phenomenal qualities? These forms of integration move well beyond mere correlation. They enable us to make better sense of, and better use of, our experiences. That should be the main goal of scientific work on consciousness.

Second, these issues lie very clearly in the province of psychology. Progress will depend on making use of the experimental methods and theoretical frameworks that modern psychology has developed. Yet, there is, at present, no coherent theoretical framework or research program that relates these issues to a unified science of consciousness. It is remarkable, for instance, that a great deal of work on metacognition proceeds without making any use of introspective reports — and thus without illuminating the nature and source of internal information. Even when tackling an issue that directly concerns the subject's point of view, experimental psychologists are reluctant to take account of it. This leads us to the final factor needed for the scientific culture to change in a way that will allow us to make progress on the genuine problem. That is a change in attitude.

Taking account of the subject's point of view involves reducing the authority of the experimenter. In the standard approach, the experimenter assumes the role of all-knowing interpreter of the subject's responses. In our approach (Jack & Roepstorff, 2002), the experimenter becomes more of a facilitator for the subject. A measure of authority is ceded to the subject, for the simple reason that subjects are ordinarily in the best position to make judgments about their own experience. To obtain reliable reports, the experimenter guides the subject to make reports that are clear, concrete, detailed, and focused as tightly as possible on the issue under investigation. The process is a dialogue, the success of which depends on the cooperation of both parties. The subject must accept the methodological prescriptions of the experimenter and the ground rules of the experiment. The experimenter must accept that the subject has the most direct access to her experience. Little can be gained if the experimenter is determined to dismiss the subject's claims from the get go; that is, if the experimenter refuses to accept the face validity of introspective reports for measuring experience (Jack & Roepstorff, 2003). In

short, the experimenter must learn to trust the subject. But trust does not imply blind faith. The wise experimenter will use strong experimental designs, for instance, including catch conditions that allow her to assess the reliability of reports and disregard data from subjects that do not pass muster (Leopold, Maier, & Logothetis, 2003).

This shift in attitude is something that, at present, many scientists and some philosophers (e.g., Dennett, 2003) steadfastly resist. If this resistance can be overcome, the intellectual and social benefits could be momentous.

REFERENCES

Baron-Cohen, S. (2000). 'Autism: Deficits in folk psychology exist alongside superiority in folk physics'. In S. Baron-Cohen, H. Tager-Flusberg, and D. J. Cohen, eds., *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience* (pp. 73–82), 2nd edition. Oxford: Oxford University Press.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18*, 227-287.

Bloom, P. (2004). *Descartes' Baby: How the Science of Child Development Explains What Makes Us Human.* New York: Basic Books.

Boyer, P., Robbins, P., & Jack, A. I. (2005). Varieties of self-systems worth having. *Conscious Cogn, 14*(4), 647-660.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*(3), 200-219.

Churchland, P. M.(1985). 'Reduction, qualia, and the direct introspection of brain states'. *Journal of Philosophy* 82: 8–28.

Churchland, P. S. (1996). 'The hornswoggle problem'. *Journal of Consciousness Studies* 3: 402–408.

Churchland, P. S. (1997). 'Can neurobiology teach us anything about consciousness?' In N. Block, O. Flanagan, and G. Güzeldere, eds., *The Nature of Consciousness* (pp. 127–140). Cambridge, MA: MIT Press.

Crick, F. H. C. (1994). *The Astonishing Hypothesis: The scientific search for the soul*.New York: Charles Scribner's Sons.

Dennett, D. (1981). ‘Intentional systems’. In *Brainstorms: Philosophical Essays on Mind and Psychology* (pp. 3–22). Cambridge, MA: MIT Press.

Dennett, D. C. (1991). *Consciousness Explained*: Penguin.

Dennett, D. C. (2003). Who's on first? Heterophenomenology explained. *Journal of Consciousness Studies, 10*(9), 19-30.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neuroscience, 23*(10), 475-483.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci U S A, 102*(27), 9673-9678.

Frith, C. D., & Frith, U. (1999). Interacting minds--a biological basis. *Science, 286*(5445), 1692-1695.

Gallagher, H., Jack, A. I., Roepstorff, A., & Frith, C. (2002). Imaging the Intentional Stance. *Neuroimage*.

Jack, A. I., & Roepstorff, A. (2002). Introspection and cognitive brain mapping: from stimulus-response to script-report. *Trends Cogn Sci, 6*(8), 333-339.

Jack, A. I., & Roepstorff, A. (2003). Why trust the subject? *Journal of Consciousness Studies, 10*(9), v-xx.

Jack, A. I., & Shallice, T. (2001). Introspective physicalism as an approach to the science of consciousness. *Cognition, 79*(1-2), 161-196.

James, W. (1890). *The Principles of Psychology*. New York: Holt.

James, W. (1892). *Psychology*. New York: Holt.

James, W. (1904). 'Does 'Consciousness' Exist?' *Journal of Philosophy, Psychology, and Scientific Methods, 1*, 477-491

Leibniz, G. W. (1714/1989). ‘The principles of philosophy, or, the monadology’. In A. Ariew and D. Garber, eds. and trans., *G. W. Leibniz: Philosophical Essays* (pp. 213–225). Indianapolis and Cambridge, MA: Hackett Publishing Company.

Leopold, D., Maier, A., & Logothetis, N. K. (2003). Measuring subjective visual perception in the nonhuman primate. *Journal of Consciousness Studies, 10*(9-10), 115-130.

Loar, B. (1997). 'Phenomenal states'. In N. Block, O. Flanagan, and G. Güzeldere, eds., *The Nature of Consciousness* (pp. 597–616). Cambridge, MA: MIT Press.

Miller, G. (2005). What is the biological basis of consciousness? *Science, 309*(5731), 79.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83*, 435-450.

Robbins, P. and Jack, A. I. (2006). The phenomenal stance. *Philosophical Studies* 127: 59–85.

Teasdale, J. D., Moore, R. G., Hayhurst, H., Pope, M., Williams, S., & Segal, Z. V. (2002). Metacognitive awareness and prevention of relapse in depression: empirical evidence. *J Consult Clin Psychol, 70*(2), 275-287.

Tononi, G. (2005). Consciousness, information integration, and the brain. *Prog Brain Res, 150*, 109-126.

Tye, M. (1999). 'Phenomenal consciousness: The explanatory gap as a cognitive illusion'. *Mind* 108: 705–725

Watson, J. B. (1913) , 'Psychology as the Behaviorist Views it'. Psychological Review, 20, 158-177