Consciousness and Cognition

David J. Chalmers

Center for Research on Concepts and Cognition
Indiana University
Bloomington, Indiana 47405.
E-mail: dave@cogsci.indiana.edu
Draft — Comments welcome.

Introduction

The problem of consciousness, also known as the Mind-Body Problem, is probably the largest outstanding obstacle in our quest to scientifically understand reality. The science of physics is not yet complete, but it is well-understood. The science of biology has explained away most of the mysteries surrounding the nature of life. Where there are gaps in our understanding of these fields, the gaps do not seem intractable; we at least have some idea of the direction in which solutions might lie. In the science of mind, things are not quite so rosy. Much progress is being made in the study of *cognition*, but *consciousness* itself is as much of a problem as it ever was.

The term "consciousness" usually serves as a convenient catch-all for all that is truly mysterious about mentality. When using the term, one must therefore be careful not to collapse important distinctions, allowing confusion. The most important distinction in the study of consciousness bears upon the approach we take to studying it: we may take either the *first-person* or the *third-person* approach. We might say that the third-person approach treats consciousness as a scientific problem, while the first-person approach treats it as a metaphysical problem. These two different viewpoints that we can adopt lead to very different treatments of the phenomenon, each of which can supply crucial insights. But the reconciliation of the viewpoints seems to be a difficult matter. Frequently, proponents of the two approaches seem to be talking past each other. In this paper, I will try to explicate carefully the relation between the two approaches, and to argue that they are not so irreconcilable as they might seem.

The third-person approach

The third-person approach has much to recommend it. On this view, consciousness is treated as a problem of science, on the same level as heat, life, or nuclear physics, and amenable to the same methods of inquiry. There has been much recent success taking this approach to the study of general cognition — so much so that a new discipline, "cognitive science", has emerged.

The raw materials of cognitive science are much the same as those of any science -- data gathered from external observation. These data can take a number of different forms. The most obvious kind of observable data are behavioural, and the study of human behaviour is concentrated in the field of *psychology*. Another form of data, more difficult to gather but extremely useful for the third-person approach, are direct observations of brain structure and function. The field of *neuroscience* is beginning to have significant success in explaining how the brain supports cognition. A third, less direct method in cognitive science is that of cognitive modelling — the construction of models, usually computational, that cohere to some extent with behavioural or neurophysiological data. The field of *artificial intelligence* is concerned with producing such models.

This tripartite investigation of behaviour, brain function and cognitive models has led to a significant increase in our understanding of diverse aspects of cognition, such as vision, memory and language comprehension. It is not surprising that these third-person methods might also be used to investigate consciousness. Such an approach to consciousness is urged by Dennett (1978), who goes so far as to specify a skeletal cognitive model for the phenomenon. Another third-person approach has been taken by Jackendoff (1987) (although Jackendoff, unlike Dennett, explicitly recognizes that a third-person approach will not necessarily tell the whole story). Besides these philosophers, some psychologists (e.g. Johnson-Laird 1983) and at least one computer scientist (Hofstadter 1979) have tried their hand at reining in the mysteries of consciousness with third-person accounts. (These accounts, incidentally, usually concentrate on those aspects of consciousness which relate to *self-awareness*, a phenomenon which does not exhaust the problems of consciousness, but which is certainly an interesting aspect thereof.)

At the heart of the third-person approach is the philosophical position of *functionalism*. Roughly stated, this is the view that the correct way to understand mental processes is to perform causal analyses, revealing the underlying abstract causal structure behind brain function. Such causal structure can, it is held, be understood objectively, and duplicated in quite different materials (such as computers, perhaps). (There are other, more specific brands of functionalism, such as those that identify mental events with semantically transparent computational processes (Fodor 1975), and those that require definition with respect to common-sense psychological theories (Lewis 1972). Unless otherwise specified, it is functionalism in the broadest sense which is meant here.) Going along with this approach is the metaphor of *mind as system*. On this view, consciousness is best understood as a particular aspect of the function of a complex system (a simplistic example might be to regard consciousness as that process whereby a system is able to scan its own processing). The view of mind as system is certainly attractive, and it is a view that I share. The question is whether this view is enough, alone, to explain the deepest mysteries of consciousness.

The first-person approach

Despite the attractiveness of the third-person approach, there is a feeling that it is side-stepping the really hard problems. The truly difficult questions only seem to arise when we take the first-person approach — when we consider, as Nagel famously put it, what it is like to *be* who we are, and what it might be like to *be* something quite different: another human, a bat, or a computational system. It is with these subjective questions that the deepest questions arise. Why is being me like anything at all? And why is it like it is? These questions are the real content of the problem of consciousness. There may be other senses of the term "consciousness", but in the sense in which it is most commonly used, to refer to the real mysteries of mentality, it is these first-person questions that are being raised.

It is still to easy to fall into confusion or to equivocate when talking of "consciousness", so here I will divide the first-person problem into three parts: the problems of qualia, mental content, and the existence of subjective experience.

(1) The problem of qualia

Qualia are the qualitative aspects of our mental states, particularly of our sensations. The paradigm qualia are our colour-sensations; other favourites are the taste of chocolate, the sound of middle C, pleasure and pain. All of these are very poorly-understood. When we look at a red patch, this sets off a particular pattern of neural firings in our brain. Why should this physical process be accompanied by a rich, subjective sensation? Given that it is accompanied by a sensation, why is it a red-sensation and not a green-sensation? There are two issues here: why qualia exist at all, and why particular qualia accompany particular processes. Is the correspondence of qualia to processes arbitrary, or is there some system to it that we don't understand? (It is not always clear just which subjective aspects of sensation count as qualia, and which don't. I will be using the term broadly, to encompass all phenomenological aspects of sensation.)

Jackson (1982) provided the most recent reminder of the qualia mystery, with a sharpening of the argument of Nagel (1974) before him. A future scientist, living in a time when neuroscience is completely understood, might learn everything there is to know about physical brain-processes. But if she has lived all her life in a black-and-white room, she will still not know what it is like to see red; when she sees red for the first time, she will *learn* something. It seems that the third-person approach, at least as currently understood, is incapable of telling us about the nature of qualia. (Jackson argues from this for a form of epiphenomenalism; my conclusion, as we will see, is not quite so pessimistic.)

(2) The problem of mental content.

When I think about a lion, something takes place in my phenomenal experience which has something to do with lions. Again, a straight physicalistic approach gives no reason to believe that such an experience should take place. What should a pattern of neural firings have to do with *lions*? But somehow, my thoughts are *about* something; they have subjective mental content. It is easy to make *attributions* of mental content to a system, justified perhaps by causal relations with the external world, but for subjective mental content we need something stronger. We need brain-states to carry *intrinsic* content, independent of our systems of external attribution; there must be a natural (in the strongest sense, meaning defined by Nature) mapping from physical state to content. (Searle (1980) calls this "intrinsic intentionality.")

The problem of mental content is not totally different from that of qualia — they are both questions about the nature of qualia. The main difference is that qualia arise during external perception, whereas mental content arises during reflection. We might call mental content "internal qualia". (It should be noted that there is also a third-person problem of mental content, which has been raging for years; particularly that of how we can assign propositional attitudes, such as beliefs and desires concerning the world, to systems and persons. This is in some ways an easier problem, as it may rely on human-defined systems of attributions, rather than needing a natural mapping (though some, such as Fodor (1975) would like a natural mapping anyway), as propositional attitudes *per se* are not items of consciousness. In other ways, the first-person problem is easier, as the contents of subjective mental states do not need to *refer* to the world — they are not so rich. When I think of a lion, my phenomenology bears *some* relation to a lion, but the relationship seems more like shared pattern than like reference. A completed theory of subjective mental content may end up having very little to do with reference.)

(3) The existence of subjective experience

The above two items are concerned with the *nature* of our subjective states — why they are one way rather than another. This is the issue which has usually been concentrated upon in the literature. But we should note that it is as deep a problem just why subjective states should exist in the first place. Why should it be like *anything* to be me? If I didn't *know* that subjective states existed, it would seem unreasonable to postulate them. This is perhaps the deepest question of all, and no current theory has come close to dealing with it.

Not many people believe in Zombies (that is, humans of normal behaviour without any subjective mental states), but at least some believe in Functional Zombies: beings which duplicate the functional organization of humans, perhaps computationally, without being conscious at all (e.g. Searle 1980, Block 1980). A question "to what class of entities may we ascribe subjective states?" is of great popular interest. In particular, is any intelligently-behaving entity conscious? I will be arguing, combining first-person and third-person considerations, that the idea of a Functional Zombie is at implausible at the very least.

There are some other commonly-raised first-person problems not explicitly listed above. The problem of *self-consciousness* (or self-awareness) I take to be a subset of the problem of awareness, of which the difficult aspects are covered by (1) and (2). The problem of *personal identity* is a separate issue and a very deep one; but Parfit's exhaustive analysis (1986), which combines the first-person and third-person approaches to great effect, gives reason to believe that our first-person intuitions here may be mistaken. It is the three problems listed above that seem to be the residual content of the traditional Mind-Body problem. I will be using the term "consciousness" broadly to cover all of these problems. If you prefer, replace every occurrence of "consciousness" with "the subjective experience of qualia and mental content."

One difficulty with talking about first-person problems is that for every first-person (conscious) mental state, there is a corresponding third-person (objectively understandable) mental state. (Perhaps there are not two different mental states, but simply two different ways of viewing one state; this is unclear. In any event, it is uncontroversial that for every subjective mental event there is a corresponding physical event; the physical event may be viewed via functional abstraction, in the cognitive science tradition, as a third-person mental event). For every subjective *sensation* there corresponds a *perception*, objectively characterizable, and so on. This dichotomy in (ways of looking at) mental states makes things a little confusing, but it will be useful later on.

The relation between the approaches

Of course, the first-person approach and the third-person approach are not entirely independent of each other. Indeed, I will be arguing that much insight can come through considering them not singly, but juxtaposed. Before this, it is necessary to enumerate a few facts about the two approaches.

(1) The third-person approach is sufficient, in principle, to yield a complete explanation of human behaviour.

This follows from the explanatory completeness of physical laws — a hypothesis which is relatively uncontroversial. Physical phenomena have physical explanations, and behaviour is a physical phenomenon. There is no reason to believe that human behaviour is any different in this regard from hurricanes, rivers or stars. With a complete understanding of the laws of physics, and a good understanding of the way physical entities combine into systems, behaviour will be understood. (Some dualists may believe that the laws of physics break down inside the human cranium, but I won't be considering this position. Physical laws may leave some room for *randomness*, but this is not an obstacle to explanation.)

This is an extremely strong conclusion, of course, and it goes a long way towards explaining the dominance of the third-person approach to the study of mentality. Cognitive science is doing its best to break down the removable barrier implied by the "in principle", and there is no reason to believe that within a few centuries it should not succeed. If it does not, it will probably be because of technical difficulties and problems of complexity, rather than any deep metaphysical barrier. We might trivially restate (1) as

(1a) The first-person approach is not needed to explain anything about human behaviour.

Of course, the first-person approach may well prove to be a useful short-cut in explaining behaviour, and a powerful tool, but the point is that it is not *necessary* — it is in principle dispensable, as far as behaviour is concerned. Another restatement might be

(1b) Behaviour would still be exactly the same, even if we had no subjective mental states.

This is not meant to beg the question of whether it is *possible* that we could lack subjective states, but the idea at least seems coherent.

(2) The third-person approach (as currently conceived) is not sufficient to explain all first-person phenomena.

Statements (1), (1a) and (1b) make a powerful argument for the primacy of the third-person approach. Nevertheless, this approach cannot tell the whole story. Consciousness, or the subjective experience of qualia and mental content, are simply not explained by a third-person account. No amount of neuroscience and cognitive modelling can explain the qualitative nature of a sensation-of-red, or even why such a subjective sensation should exist.

This statement, though surely a majority view in the philosophy of mind, is certainly not undisputed. Those who have spoken the loudest in dispute of (2) are Dennett (1982, 1987 ?) and P. M. Churchland (1986). I believe their arguments to be flawed, but will not argue the issue here, as it has been eloquently argued by Nagel and Jackson, among others. Those who disbelieve (2) may choose to stop reading here, then, in the belief that consciousness is not really such a problem. They might however read on, to find an account of first-person problems which is motivated by third-person issues at every point, or at worst to find a reductio ad absurdum of the first-person position. The remainder — those who believe that the Mind-Body Problem is still a real problem — may continue undisturbed.

It should be noted that (2) refers to the third-person approach *as currently conceived* — that is, the study of the brain as a physical system with the usual physical ontology, abetted by cognitive models and functional analyses. I would not like to rule out the possibility of an eventual “objective phenomenology”, to use Nagel’s term, but it is difficult to see what it would look like. At the least, new constructs would be needed, and our ontology would need to be expanded. (This is not as bad as it might sound...) To get from the physical and the functional to the subjective, we would need *metaphysical bridging principles*. It does not seem impossible, however, that these could be stated from a third-person viewpoint. The beginnings of such an approach will be outlined later in this paper. In the meantime, we can restate (2) as

(2a) The third-person approach is not sufficient, in the absence of extra metaphysical bridging principles, to explain all first-person phenomena.

The need for such a bridge is put nicely by Jackendoff (1987), who says that computational/functional accounts of mind may have solved the Mind-Body Problem, but what we need to do now is to solve the “Mind-Mind Problem". This would provide a bridge from the “computational mind” to the “phenomenological mind” (in the terms I have used, from third-person mental events to first-person mental events). That computational and functional accounts may have already got us halfway there is a point which will resurface later.

(3) Our claims about consciousness are a fact of human behaviour.

This is trivially true. (If you believe “claim” to be an irreducibly subjective term, then replace it by “utterance” or something more neutral.) When I say “I’m conscious, and I’m totally baffled by it,” that’s a fact of human behaviour. When somebody writes “The ineffable sensation of red is inexplicable by physical premises”, that’s a fact of human behaviour. This is harmless enough, until combined with (1) to yield

(4) The third-person approach is in principle sufficient to explain our claims about consciousness

or worse,

(4a) The first-person approach is not needed to explain our claims about consciousness.

Everything that we say about consciousness is, in principle, amenable to the usual kind of physical/functional analysis. There is no need to appeal to *any* mysterious metaphysical constructs to explain the things we say. On this account, the last 2000 years of debate over the Mind-Body Problem would have gone exactly the same in the absence of subjective experience (if such a thing were possible). In particular, our sense of bafflement about the first-person problems manifests itself in numerous third-person ways — the things we say, the things we write, even the things we *think*, if regarded in terms of neural processes — and is thus amenable to the methods of cognitive science[1]. The correct cognitive-science explanation of why we claim to be conscious might be a little while coming, but there should be no deep metaphysical problems. This leads us to an important principle:

The Mystery Principle: Consciousness is mysterious. Claims about consciousness are not.

This can equivalently be stated as

The Surprise Principle: Although consciousness should surprise us, claims about consciousness should not.

Most people will agree that consciousness is a surprising property. If it weren't for the fact that first-person experience was a brute fact presented to us, there would seem to be no reason to predict its existence. All it does is make things more complicated. By contrast, the things we say about consciousness are common-or-garden cognitive phenomena. Somebody who knew enough about brain structure would be able to immediately predict the likelihood of utterances such as "I *feel* conscious, in a way that it seems no physical object could be", or even Descartes' "Cogito ergo sum". In short, our bafflement about consciousness can be understood purely as a problem for cognitive science. There are very rich pickings awaiting anybody who takes this path. I believe that it's not as difficult a path as it might seem, and towards the end of this paper I will give a very brief account of why we might expect such bafflement.[2]

The situation with which we are now faced seems to border on the absurd. If everything we say, write and even *think* about consciousness is explanable by cognitive science, why is there any need to posit any great mystery about consciousness in the first place? Why not concede that the appearance of great metaphysical mystery is simply an illusion, and regard consciousness itself as a cognitive process, on much the same level as memory or learning? It becomes a third-person high-level term, playing a similar role in cognitive science as "heat" does in thermodynamics. As for this weird "first-person experience", then insofar as it is not a purely cognitive phenomenon, it simply does not exist. This is the approach taken by Dennett and others, and there are powerful arguments in its favour, as evidenced above. But this paper is based on the premise that the Mind-Body Problem is still a real problem. We are not yet prepared to concede that (2) is false, and that the third-person approach as currently conceived is sufficient to explain all of the first-person mysteries. As first-person loyalists, we have to bite this bullet and struggle on, trying to produce a coherent account of the matter. I believe that the pressures on a first-person account produced by (4) and (4a), and the third-person constraints that are thus acquired, are of great value in shaping a solution that might be acceptable to partisans of both the first-person and third-person approaches.

---

1. McGinn (1989) requests an explanation of why it is so difficult to clearly articulate the the problems of consciousness. I believe that the reason is closely tied up with (4) and (4a). Everything we can *say* about consciousness is physically caused, and somehow, at the back of our minds, we are aware of this. So when the words come out of our mouths, we are aware of how easy it might be to explain our claims without invoking any great mysteries. Further, our words can easily be taken as referring to *third-person* mental states: perception rather than sensation, and so on. Our claims seem very inadequate to do justice to the real phenomenon,

2. There are other reasons why claims of consciousness should not surprise us. It seems plausible that any system which (in a purely functional sense) "perceives" aspects of the world and further has sophisticated linguistic capacity might profess perplexity about the nature of its perception; one might expect puzzlement at both the "qualitative" and "indexical" aspects. And if such a system has the capacity to "think about what it's thinking about" (a functional notion, once again), it might very well claim full-blown self-consciousness. No metaphysics needs to be invoked in our analysis of such a situation. (Computer A: "I know I'm just a collection of digital circuits, but I can't explain this strange *feeling* I have of having thoughts of my own." Computer B: "Yes, I'm inclined towards dualism myself.") This is a point which deserves further development.

This is the low point for the first-person partisan. This is as bad as it gets. But we have to confront the problem posed by (4) and (4a) head on. This has rarely been done in the literature. "Qualia freaks" (Jackson's term) have generally been content to argue for the autonomy of first-person phenomena, from independent considerations, or more frequently to simply assume it. But if we don't take (4) and (4a) seriously, then any account that we produce is open to question. The question which faces us, put slightly differently, is

(4b) How can first-person experience have any bearing on what we *say* about first-person experience?

or, using our preferred shorthand,

(4c) How can consciousness have any bearing on what we say about consciousness?

I believe that in the proper understanding of this question lies the seed of a solution to the Mind-Body Problem. All the same, it is surprising how rarely the question has been considered in the literature. (The more general question of how the mental can affect the physical of course a perennial favourite. Question (4c) is much more specific, and raises a different set of problems.) It has appeared in various guises in a few places, however; we can take these as delineating various choices in how we might answer the question. The choices spread themselves out before us as follows.

(A) We might choose to reject premise (1), and deny that the third-person approach can even give a full account of human *behaviour*. This is the approach taken by Elitzur (1989), who argues from the fact that we talk about consciousness to the conclusion that consciousness plays an active role in determining human behaviour, and deduces that the laws of physics must therefore play an incomplete role in determining behaviour. This is certainly an interesting argument, but it is a little implausible and should be regarded as a last resort. The laws of physics are a little too important to us to be given up so easily. (A similar account of the matter would presumably be given by dualists from Descartes to Eccles.)

(B) We might choose to answer (4c) by saying, simply, "It doesn't." This is the answer which might be given by Jackson (1982), who argues that qualia are completely epiphenomenal, playing no causal role (although he does not explicitly consider utterances about consciousness). Now, epiphenomenal qualia are already a little unparsimonious under general considerations, but considered in the context of question (4c) their implausibility increases. Are we to say that our claims about consciousness are completely independent of the fact of our consciousness? That any similarity manifested between our claims and the actual facts are mere coincidence? While this conclusion cannot be completely ruled out, it certainly seems inelegant. It fails what I will call, later, the "Coherence Test" — its explanations of (1) consciousness and (2) claims about consciousness fail completely to cohere with each other.

I nevertheless believe that there is a grain of truth in epiphenomenalism, but that it needs to be spelt out more carefully. If we can explicate just *how* consciousness is an epiphenomenon of physical processing, then question (4c) may admit a more parsimonious answer. I will be defending a kind of "double aspect" theory which bears some similarity to epiphenomenalism, while at the same time being much more palatable, I hope, to physicalists.

(C) We may choose to deny premise (2), and argue that in fact the third-person approach tells us everything there is to know. By denying that consciousness is an essentially first-person, metaphysical phenomenon, we deny that question (4c) poses any problem. If there is anything at all called "consciousness" that plays a causal role, then it does so in exactly the same way that centers of gravity play a causal role in physics, or that temperature plays a causal role in thermodynamics: as a convenient third-person abstraction. This, without doubt, is the approach that the likes of Dennett and Churchland would take, although I have not seen them argue the matter explicitly.

The consideration of questions like (4c) has been used as an explicit argument for the third-person approach in a few places. Foss (1989), arguing against the Nagel/Jackson arguments, argues that a "Super Neuroscientist" could in principle know everything that a beingwould say about its qualia-experience, and everything that it *might* say. From this, he draws the conclusion that the Super Neuroscientist would know everything that there is to know about qualia. Watkins (1989) argues that if what we say about qualia is to reflect the existence of qualia at all, then qualia must be physical.

Clearly questions like (4c) are a goldmine for third-person arguments, but I believe that these analyses are a little simplistic. They "solve" the Mind-Body Problem by the method of denying that there is a problem. This runs counter to our premised belief that there is a substantial problem to be solved. Nevertheless, I believe there is a grain of truth in these arguments. There is perhaps *some sense* in which qualia are physical. There is perhaps *some sense* in which consciousness is understandable from the third-person viewpoint. What must be explicated is the nature of this sense. The arguments above do not solve the Mind-Body Problem, but they do indicate that it might be soluble. However, there is more work do be done. A flat statement of physicalism is true in no sense a complete solution. We need to know *how* it could be true. The nature of the work that remains to be done will be outlined shortly.

Before we leave this section, we should note that a well-known argument of Shoemaker (1975) bears some resemblance to the arguments in (C). Shoemaker was arguing against the possibility of "Absent Qualia", or Functional Zombies. These are supposed to be arguments against functionalism, demonstrating that we might duplicate any given functional process perfectly without any accompanying subjective experience. Shoemaker argues that we *know* about qualia; our knowledge is caused by the existence of these qualia; therefore qualia play some causal role. The conclusion is that Absent Qualia are impossible, as removing qualia would mean changing the causal structure. This is generally construed as an argument for the third-person approach (for functionalism, in particular). Nevertheless, I am sympathetic with this argument and with functionalism, and believe that Absent Qualia are impossible. However,what needs to be explained is *why* Absent Qualia are impossible, and *how* qualia could play a causal role. This question is nowhere considered by functionalism. Again, more work remains to be done. Both functionalism and epiphenomenalism have plausible aspects, but neither gives a complete account of the problems of consciousness. I will suggest that a correct theory of consciousness shares aspects of both functionalism and epiphenomenalism, and that the correct answer to question (4c) lies somewhere between (B) and (C).

The Coherence Test

The Coherence Test is a test which any completed theory of mind must pass. It is motivated directly by questions like (4c).

The Coherence Test:
A completed theory of mind must provide
(C1) An account of why we are conscious.
(C2) An account of why we *claim* to be conscious; why we *think* we are conscious.
Further (C3), accounts (C1) and (C2) must *cohere* with each other.

It should be noted that (C1) is a completely first-person matter; it involves answering the metaphysical question of how subjective experience is possible. (C2), on the other hand, is understandable on strictly third-person terms — it is, in principle, a matter for cognitive science. (If, again, in your vocabulary "think" and "claim" are irreducibly subjective terms, then replace them by appropriate third-person terms that deal with behaviour and brain function.) Another way in which to phrase (C2) might be "An account of why consciousness seems so strange to us." Viewed appropriately, this is also in principle understandable by a cognitive science approach. Anyway, it is the coherence condition (C3) that is the most important here. This ensures that completely independent accounts under (C1) and (C2) will be ruled out; that the accounts must bear a substantial relationship to each other. This provides a vital link between the first-person and third-person approaches.

Another way of putting the Coherence Test is that we need accounts of
(C1a) The mind that we experience (a first-person question); and
(C2a) The mind that the brain perceives (a third-person question).

One would hope that these "minds", on any completed account, would bear a very close relationship to each other. If the relationship is not strict identity, it should at least be a very close correspondence.

It is not possible to "prove" the validity of the Coherence Test, but it certainly seems extremely plausible. While it is logically possible that an account of why we are conscious could be completely independent of a cognitive account of why we *believe* we are conscious, it seems inelegant and unlikely. If the Coherence Test in fact fails for a correct completed theory of mind, then we are living in a world where the things we say about consciousness and the Mind-Body Problem bear at best a coincidental relation to the way these things actually are. Faced with the possibility of this complete unreliability of our verbal reports, it would probably be better to join the third-person camp and regard the mysterious nature of "subjective experience" as a mere illusion.

Given that we accept the Coherence Test, a correct third-person account of cognitive processes puts severe constraints on possible first-person accounts, and possibly vice versa. Of course, we have not explicitly defined the notion of coherence — instead, I think it is better to go ahead and apply it on a case-by-case basis, and let some understanding of the notion emerge. (If the reader, on going through the following, worries that perhaps *no* notion of mind can pass the Coherence Test, then rest assured that an example of what a Test-passing theory might look like will be given later.)

Applying the Coherence Test

Functionalism is a controversial doctrine when applied to first-person aspects of mentality. But for the third-person study of mental processes, it reigns supreme. Few people these days doubt that the correct way to analyze human behaviour and to explain the workings of the brain is in functional terms. All remotely satisfactory accounts of learning or memory, say, explain these via causal analyses of systems with many interacting parts. Indeed, it is difficult to see what other kind of account there could be. So when it comes to explaining the things we *say* (and even *believe*) about consciousness, there is little doubt that a functional analysis will be the way to go. We may enshrine this doctrine as

(5) The fact that we claim to be conscious holds in virtue of certain functionally specifiable properties of the brain.

Of course, we're not sure exactly what those functional properties *are*, yet — this is part of the reason why functionalism alone is not a completed theory of consciousness. But we are almost certain that the properties will be *functional* ones, considered at the most parsimonious level. And there is no doubt that if we make the correct functional abstraction from the brain, then duplicating this functional specifications will produce the same kind of *claims* about consciousness.

We may now use this uncontroversial claim, together with the Coherence Test, as an argument against certain theories of mind.

Theory 1: We are conscious in virtue of low-level biochemical processes.

Such a view of mind is held by at least a few writers on the subject (e.g. Searle 1980, Block 1980, Horgan 1984). These writers have generally been arguing against functionalism as a theory of mind, at least as a theory of the first-person aspects. Block, for instance, claims that qualia might be a consequence of physiological facts which need not appear in a functional account. Searle thinks it very plausible that the "causal powers" of the brain (meaning those powers sufficient to cause consciousness) lie at the biochemical level.

This view, however, runs a great risk of failing the Coherence Test. If, as according to (5), our claims of consciousness are in virtue of certain functional properties properties of the brain, then the statement that consciousness is inherently neurophysiological suggests that consciousness has nothing to do with the production of our claims. This would seem strange. A possible reply by neurophysiological partisans might be to claim that the neurophysiological locus of consciousness is specifiable functionally, albeit at a very low level. The connectionist movement in cognitive modelling has recently made popular the idea that a correct functional account may inhere at a low level of abstraction. (Recent proponents of such an explicitly low-level functionalism include Lycan (1987) and Clark (1989). But if we accept that neurophysiological accounts can also be functional (as Lycan urges), then this is no longer an argument against functionalism per se — a consequence that Block and Searle would be loath to accept. Insofar as our theory holds that the sources of consciousness are not specifiable functionally, then it is a theory which fails the Coherence Test. If consciousness does not arise from the functional, then our claims are independent of consciousness.

(Block (1980) argues against a similar argument by Shoemaker, stating that qualia can be independent of any given functional account but still make a causal difference. But such a difference will by definition not bear on any aspect of our functional account. Given that our account is specifically constructed to include our claims of consciousness as a relevant aspect, this reply fails. If qualia are independent of this functional account, then qualia cannot make a difference to our claims. Horgan (1984) makes a similar confusion in defending the claim that qualia are neurophysiological. Considering the case of Martians functionally identical to us but physiologically different, he claims that the issue of whether or not they have qualia might be resolved by observing whether they manifest the same kind of puzzlement about qualia that we do. But *of course* they will manifest similar puzzlement, as this is determined by their functional architecture. Therefore such observation will tell us nothing more than we already knew.)

Even if we we chose to hold that underpinnings of consciousness are specifiable functionally, it may be dangerous to go to too low a level. It seems very likely that when an account of the processes which lead to our consciousness-claims is found, it will be specifiable at a level which lies above the biochemical. If we insist that consciousness itself inheres at a lower level than this, then the source of consciousness is independent of the source of our consciousness-claims, and our theory fails the Coherence Test. (It may be possible that certain *aspects* of qualia are determined at a lower level than the level of our qualia-claims. Such aspects could not make any causal difference relevant to our claims. For instance, the difference between red-sensations and green-sensations is very difficult to articulate, and so the precise nature of these sensations might be dependent on certain low-level facts. But the *existence* of such qualia must be dependent on the higher-level functional account, if the correct theory of mind is Coherent.)

We may conclude that theories which place the source of consciousness at too low a level are very implausible. A similar argument will apply to any theory which does not satisfy Functional Supervenience: the principle that replicating certain relevant functional processes, even in a different substrate, is sufficient to guarantee first-person mental states. (This, incidentally, is a weaker principle than the Functional State Identity Theory, which holds that mental states *are* functional states.) If consciousness exists in virtue of some fact which is quite independent of any functional account, then our claims of consciousness do not reflect the fact of our consciousness. This argument parallels Shoemaker's argument against Absent Qualia in some ways, although unlike Shoemaker I do not wish to argue for the truth of the Functional State Identity Theory. But the argument establishes that if there could exist beings which duplicate our relevant causal structure (perhaps in a quite different substrate), but which do not possess subjective states — in other words, if there exist Functional Zombies — then the correct theory of mind is fundamentally Incoherent.[3]

Theory 2: Epiphenomenalism.

This view does not come close to satisfying the Coherence Test. It holds that we are conscious in virtue of certain mysterious dualistic principles, which yield mental states as a consequence of physical states. Such principles would on the face of it be quite independent of the relatively straightforward functional sources of our claims that we have "qualia" and "subjectivity". All parsimony is thrown out the window. The facts that we (a) have qualia, and (b) claim to have qualia, are related only by coincidence.

This, incidentally, does not rule out all forms of epiphenomenalism. It might be possible to have a *principled epiphenomenalism*, where the "mysterious dualistic principles" above are replaced by something more concrete, showing that mental states arise in virtue of specific, functionally specifiable physical states. If this were done, then it is possible that the Coherence Test might after all be satisfied: we would *be conscious* and *claim consciousness* in virtue of similar functional accounts. But all that this possibility shows, for now, is that epiphenomenalism would need a great deal of development to be a plausible theory of mind.

Theory 3: Identity Theories, Functional and Physical

We have recently been rather sympathetic to functionalism, but it is not close to a completed theory of mind. It performs rather well on part (C2) of the Coherence Test. The source of our consciousness-claims is undoubtedly functionally specifiable. It must be conceded that the flat statement of functionalism gives no insight into just *which* functional properties of the brain are responsible for our consciousness-claims, so there is more work to be done on that point — but we might set that aside as a mere technical difficulty, a matter for cognitive scientists. A more profound difficulty with functionalism is that it does not come close to dealing with (C1). It has often been noted (recently by Searle 1989) that functionalism is not a theory which is motivated by the existence of first-person mentality; rather, conscious mental states are viewed as an *obstacle* for functionalism to deal with (witness all the Absent Qualia and Inverted Spectrum objections). It is quite possible that functionalism is *compatible* with a correct theory of consciousness, but taken alone it is not that theory.

Functionalism gives very little insight into why consciousness exists. It is one thing to baldly state "subjective experience arises in virtue of our functional organization"; it is another to give an account of *why*. Functionalism, in this way, is not unlike its predecessor, the Identity Theory, which stated "mental states are brain states". The flat statement alone gives us no insight into *how* it could be true. There was, of course, some kernel of truth to the Identity Theory: this was that conscious mental states are indeed *supervenient* on brain states (that is, reproducing a brain state will reproduce the same corresponding mental state). In a similar way, Functionalism contains a kernel of truth: it is very plausible that conscious mental states supervene on certain functional states. By going from the loose notion of "physical state" to the more specific notion of functional state, functionalism as brought us closer to the goal, but it certainly has not achieved it. (Functionalism seems to do an excellent job of capturing non-phenomenological mental states, such as propositional attitudes, but that is not our concern here.)

So both Functionalism and the Identity Theory pass over the key question, of just *why* we are conscious in the first place. And of course, the fact that it fails on (C1) implies automatic failure on (C3) — without giving an account of the existence of subjectivity, it cannot give an account of coherence. Nevertheless, I believe that functionalism is a valuable theory, of which many

3. It should be noted that the implausibility of Functional Zombies does not necessarily imply the impossibility of Computational Zombies. It just might be possible to be a functionalist without believing that computational specifications can ever capture the appropriate causal structure. (Harnad (1989), for instance, believes that the requisite function is inherently analog.) Church's Thesis diminishes the plausibility of such a position -- it seems likely that a non-computational causal account always supervenes on a lower-level computational account -- and in practice functionalism and computationalism tend to stand and fall together.

aspects are worth retaining. It is simply that something needs to be added. The account that I will give below might be characterized as “functionalism plus something”. (Those third-person partisans who do not believe that the Mind-Body problem is a real problem will of course not be buying any of this; I hope, nevertheless, that they are still with me.)

To satisfy the Coherence Test, and thus be a candidate for the title of “completed theory of mind”, a theory must provide three things, on my estimation:

(1) A *metaphysical* account of why subjective experience is possible. This will almost certainly include some new metaphysical construct, over and above physical laws.

(2) A *functional* account of why we think and claim that we are conscious. This will presumably be in the idiom of cognitive science.

(3) An account of functional-metaphysical coherence, which shows how (1) and (2) cohere with each other.

As we have seen, theories that consciousness is biochemical pass the buck on (2) and fail miserably on (3). Epiphenomenalism makes a start on (1), but as it stands has no chance on (3). And functionalism, as currently conceived, does not deal at all with (1), and therefore with (3). I hope that I have demonstrated just how far short currently conceived theories fall of yielding a solution to the Mind-Body Problem. The mode will now switch from pessimism to optimism. The three requirements above sound daunting, but I believe they might be surprisingly easy to satisfy, without bringing in too many unfamiliar concepts. In the next section, I will briefly outline the beginnings of a speculative theory which has a chance of satisfying (1), (2) and (3) above. This theory alone will not be a solution to the Mind-Body Problem, but I hope that it shows what a solution might look like.

Mind, pattern, information

The theory I am going to suggest is an example of what is known as a “double aspect” theory. That is, it holds that first-person mental states and third-person mental states are two different aspects of the same thing. double aspect theories have been attractive to others (e.g. Nagel 1986), but they have usually foundered because we cannot say exactly what those aspects might *be*. The simple idea “double aspect theory” explains very little. But before I go on to elaborate on this particular theory, it might be useful to say a couple of words about why double aspect theories are so attractive. In short, this is because double aspect theories are at the same time almost epiphenomenalism and almost identity theories, combining the virtues of both with the vices of neither. The theory I will propose is almost an identity theory (by this term I include both “Brain State” and “Functional State” Identity Theories, though the proposed theory is nearer to the latter than the former), in that it holds that first-person and third-person states are the same thing — but different aspects of the same thing, a crucial difference. (You might say: an identity theory which takes subjective states seriously.) It is almost a version of epiphenomenalism, as we can imagine the subjective aspects “hanging off” the non-subjective aspects, allowing the complete autonomy of the physical — while at the same time allowing that subjective states can be causally efficacious, as they are but another aspect of the objective states. (You might say: an epiphenomenalism where mind can matter.) The account which follows will necessarily be brief, speculative, poorly-defined and incomplete. Nevertheless, I hope that it offers a glimpse of what might be the correct direction for the understanding of consciousness.

The two different “aspects” that I propose are pattern and information. Wherever they occur, pattern and information occur together. All information is carried by some pattern in the physical world; all patterns carry some information. I speculate that they should be regarded as two different aspects of the same thing, which for want of a better term we may call “pattern/information."

My proposal is that third-person (objectively understandable) mental events are *patterns* in the brain, while the corresponding subjective (first-person) mental events are *information*. On the above view of pattern/information, this implies that first-person and third-person mental events are indeed two different aspects of the same thing, the underlying pattern/information. (This has the advantage of not expanding our ontology *too* far.)

The idea that third-person mental events can be regarded as patterns should not be too controversial. A commitment very much like this is already made by functionalism, though it is not always emphasized. When we give a functional account of a system, we are necessarily abstracting away from many superficial details, and highlighting particular patterns in the physical substrate. When the functionalist says that a mental state is a functional state, she is committing herself to the notion that a mental state is an *abstraction* — and every abstraction is an abstraction of a *pattern*. (Of course, not every pattern is a *functional* pattern, but these are a very important subset.)

It is rarely made clear just what ontological claims a functionalist would wish to make, though. Are they committed to patterns as part of their natural ontology? They might not necessarily have to take this step, as if we are considering mental events only from the third-person viewpoint, then it is not clear that we have to admit them into our ontology anymore than we have to admit centers of gravity — both might be “convenient fictions”, in Dennett’s terminology. But I am prepared to bite the bullet, and reify patterns. Given that we believe that the mind and conscious experience are part of the basic ontology of things, then some concession is necessarily — we have to posit some metaphysical construction of the kind mentioned in the previous section. Admitting patterns into our ontology is a surprisingly painless way of doing this — and once it’s done, we get minds for free.

Of course, we’ve done a little more than reify pattern: we’ve reified pattern/information. I claim that this is a reasonable step, given that they always go together. Put it this way: if there’s any chance that a double aspect theory might be the right theory — and as we’ve seen and will see, there is very much that’s attractive about them — then I can’t imagine two more appropriate or natural candidates for the “aspects” than pattern and information. To say that “mental events arise from a double aspect” is to say very little, unless we want to admit the mental as a primitive member of our ontology. Instead, by positing pattern/information, we have a much more natural primitive in our ontology, from which the double-aspect nature of mentality is inherited.

Once we’ve made this posit, then conscious mentality falls out. Third-person mental events are patterns in the brain: the corresponding conscious mental events are the information that these patterns carry. *Qualia are just information.* (A nice way of putting this is “information is what pattern is like from the inside". I’m not quite sure what it means, but it certainly sounds good.) Anyway, conscious mentality arises from the one big pattern which I *am*. That pattern, at any given time, carries a lot of information — that information is my conscious experience. (I should be more careful, I know. I don’t necessarily want to commit myself to the existence of a single individual that the term “I” refers to. But the picture is much the same whether one is committed to that or not.) Incidentally, given any physical substrate there are many different ways of abstracting patterns from it. In my brain, there will be some patterns that don’t correspond to mental states at all, and some which correspond to unconscious states. This is not a problem for the theory; there will be more on this later.

Why we think we’re conscious

The last few paragraphs may strike the reader as wanton ontological extravagance. But apart from any inherent elegance the theory might have, it is also motivated by the fact that it can pass the Coherence Test. In fact, the above is at least in part motivated by an functional account of why we think we are conscious, and why consciousness seems strange. Logically, the functional account ought to come first, but it makes sense to present the metaphysical theory up front. But what follows can be read independently of what went before. This will be a purely functional account, in the tradition of cognitive science. So at least temporarily, all the metaphysical baggage may be thrown away.

Very, very briefly, here is what I believe to be the correct account of why we think we are conscious, and why it seems like a mystery. The basic notion is that of *pattern processing*. This is one of the things that the brain does best. It can take raw physical data, usually from the environment but even from the brain itself, and extract patterns from these. In particular, it can discriminate on the basis of patterns. The original patterns are in the environment, but they are transformed on their path through neural circuits, until they are represented as quite different patterns in the cerebral cortex. This process can also be represented as *information flow* (not surprisingly), from the environment into the brain. The key point is that once the information flow has reached the central processing portions for the brain, further brain function is *not* sensitive to the precise original raw data, but only to the pattern (to the information!) which is embodied in the neural structure.

Consider colour-perception, for instance. Originally, a spectral envelope of light-wavelengths impinges upon our eyes. Immediately, some distinctions are collapsed, and some pattern is processed. Three different kinds of cones abstract out information about how much light is present in various overlapping wavelength-ranges. This information travels down the optic nerve (as a physical pattern, of course), where it gets further transformed by neural processing into an abstraction about how much intensity is present on what we call the red-green, yellow-blue, and achromatic scales. What happens after this is poorly-understood, but there is no doubt that by the time the central processing region is reached, the pattern is very much transformed, and the information that remains is only an abstraction of certain aspects of the original data.

Anyway, here is why colour-perception seems strange. In terms of further processing, we are sensitive *not* to the original data, not even directly to the physical structure of the neural system, but only to the *patterns* which the system embodies, to the information it contains. It is a matter of *access*. When our linguistic system (to be homuncular about things) wants to make verbal reports, it cannot get access to the original data; it does not even have direct access to neural structure. It is sensitive only to pattern. Thus, we *know* that we can make distinctions between certain wavelength distributions, but we don't know how we do it. We've lost access to the original wavelengths — we certainly can't say "yes, that patch is saturated with 500-600 nm reflections". And we don't have access to our neural structure, so we can't say "yes, that's a 50 Hz spiking frequency". It's a distinction that we're able to make, but only on the basis of *pattern*. We can merely say "Yes, that looks different from that." When asked "How are they different?", all we can say is "Well, that one's...*red*, and that one's *green*". We have access to nothing more — we can simply make raw distinctions based on pattern — and it seems very strange.

So this is why conscious experience seems strange. We are able to make distinctions, but we have direct access neither to the sources of those distinctions, or to how we make the distinctions. The distinctions are based purely on the information that is processed. Incidentally, it seems that the more abstract the information-processing — that is, the more that distinctions are collapsed, and information recoded — the stranger the conscious experience seems. Shape-perception, for instance, strikes us as relatively non-strange; the visual system is extremely good at preserving shape information through its neural pathways. Colour and taste are strange indeed, and the processing of both seems to involve a considerable amount of recoding.

The story for "internal perception" is exactly the same. When we reflect on our thoughts, information makes its way from one part of the brain to another, and perhaps eventually to our speech center. It is to only certain abstract features of brain structure that the process is sensitive. (One might imagine that if somehow reflection could be sensitive to every last detail of brain structure, it would seem very different.) Again, we can perceive only via pattern, via information. The brute, seemingly so non-concrete distinctions thus entailed are extremely difficult for us to understand, and to articulate. That is why consciousness seems strange, and that is why the debate over the Mind-Body Problem has raged for thousands of years.

The above account is far too brief, and is almost certainly wrong in many important aspects. Nevertheless, I believe there is a kernel of truth in there. Even if the above account needs to be thoroughly revised, I think this fact will remain: the facts that we think we are conscious, we claim to be conscious, and consciousness seems strange all hold true in virtue of the fact that the brain is a sophisticated pattern-processor (is a sophisticated information-processor).

Further, this account should not surprise us. Things *have* to be this way. Any being which "perceives" the world must do so in virtue of pattern/information processing. Once the processing has reached the heart of the cognitive system, further processing can *only* be sensitive to the information distinctions embodied there. When the being tries to articulate the nature of its perception, it will be reduced to talk of "qualitative distinctions". Similarly, if a being has any reflective access to its internal processing, it will only be in terms of high-level patterns. We should *expect* such beings to be baffled by this "consciousness". The Surprise Principle is vindicated: Consciousness should surprise us, claims about consciousness should not.

The above account, of course, did not rely on any *ontological* commitment to pattern and information — it merely used them as "convenient fictions." This was a completely functional account. If I were a third-person aficionado, who believed that the Mind-Body Problem is only a pseudo-problem, I might stop now, and say "See! Consciousness is just an illusion." But I am not, and this account was intended not to explain away consciousness but to cohere with it.

So now, we may run the Coherence Test. We *claim* that we are conscious in virtue of the brain's ability as a pattern-processor (as an information-processor). We in fact *are* conscious in virtue of the patterns (information) embodied by the brain. These explanations seem to cohere rather well. Of course, there is no pattern-processing without patterns. It is precisely the same patterns which are at the crux of the pattern processing (those at the heart of our processing system, from which all distinctions are made), which are also the patterns which *are* our mental states. (I occasionally use "pattern" as a shorthand for "pattern/ information". No great harm is done.) One could not ask for better coherence than this. The reason we *talk* about mental states is that our processing is sensitive to precisely those patterns which *are* our mental states.

Another, slightly over-simplified way to put this, is: (1) The brain perceives itself as pattern. (2) The mind (that we experience) is pattern. Therefore (3) The brain perceives itself as the mind.

Patterns in pattern-processors

On this account, there are two criteria for being a conscious entity: a metaphysical (first-person) and a functional (third-person) criterion. The metaphysical criterion is that one must be a pattern. All patterns exist, presumably. It would seem strange to reify some patterns but not others. But not all patterns are conscious. To be a conscious pattern, one must be part of the right kind of pattern processor, bearing an appropriate relation to it.

An interesting and difficult question concerns the status of patterns which are not part of pattern processors. Such patterns may still carry information; is there anything it is like to *be* them? My answer is a tentative "yes". But it would not be anything like being a human being, for such patterns are not *conscious*, in the appropriate functional sense. They are not parts of pattern-processors, so they cannot be aware. This incidentally seems to demonstrate that we should separate the purely first-person notion of "be-ability" from the partially third-person notion of "consciousness". Consciousness itself is dependent on certain functional criteria, but it seems implausible that such third-person criteria should impose restrictions on metaphysical be-ability. Being the number 5 might be like *something*; if only like being asleep, without the excitement of dreaming.

On the other hand, there are pattern processors which are not embodied in human brains. Might such processors (or the patterns therein) have subjective experience? A connectionist network, for instance, is a pattern processor *par excellence*. A typical feed-forward network might have served as a perfect simplified example in the above account of the kind of pattern/information-processing that goes on in the human brain. Networks are sensitive to certain patterns in environmental data; they recode these patterns as they pass down the information-processing chain, until output is sensitive only to a certain kind of pattern in the inputs. Do connectionist networks have qualia? My answer is "yes!" This may seem very counter-intuitive at first, but I believe that upon reflection it becomes plausible. There is no principled distinction between the kind of pattern-processing performed by such a network and that performed by a brain, except one of complexity. And there is no reason to believe that complexity should make such a great difference to the existence of qualia. Certainly human qualia will be more complex and interesting, reflecting the more complex processing; but then, even considering our processing of "redness", there is no evidence that much more complex processing goes into this than goes into a typical large connectionist network. Pattern-processing leads to qualia in humans, and there seems no reason to deny that pattern-processing leads to qualia in networks also. (Of course networks cannot as yet *reflect* on the fact that they have qualia, let alone *talk* about it, but this is not the point.) [4]

Connectionist networks help illustrate another point: that patterns may supervene on functional descriptions. The relevant pattern in networks are *patterns of activation* over a number of units, where units and their activations are of course functionally specifiable. It is may not be the case that *all* patterns supervene on functional descriptions (that is, are specifiable relative to functional descriptions, and duplicating the functional system in another substrate allows duplicating the pattern), but it seems plausible that all patterns that represent first-person mental states should. The plausibility lies in the fact that patterns need to be *processed* for us to be aware of them, and processing is a functional notion. The current account is thus compatible with functionalism of a certain variety. You might say that it takes functionalism as a starting point, and adds what is necessary to deal with the problems of consciousness.[5]

Loose ends

Moving away from the more cognitive aspects to the more metaphysical aspects, we are yet to explicitly answer our original question (4c): how can consciousness have any bearing on what we say about consciousness? On the present account, this is easy. Conscious experience is identified with the "information" aspect of certain pattern/information states. This information can certainly make a difference. It is just another aspect of the pattern, and there is no question that the pattern plays a causal role. Changing the information changes the pattern, and changing the pattern changes many consequences. In fact, as we have seen, our claims about consciousness reflect precisely that pattern/information in which our mental states consist. Consciousness is thus causally efficacious. At the same time, it is quite possible to analyze third-person mental states and thus behaviour without ever needing to invoke consciousness, if we so desire. It is possible to cast everything in terms of pattern, without invoking information.

This answer corresponds to none of the answers (A), (B) or (C) given originally. Our options were in fact not quite so limited as they seemed. This again shows the advantages of a double aspect theory — an identity theory which takes consciousness seriously.

---

4. Then there is that old favourite, the thermostat. Do thermostats have qualia? It is not totally clear that thermostats do the right kind of pattern-processing, but if they do, their qualia are remarkably simple. For thermostats process all physical inputs down to three states: too hot, too cold, and just right. It's not clear just what three-valued qualia would be like. But if we are ever to understand qualia, this is perhaps a good test case.

5. No theory of mind worth its salt would pass up the chance to attack Searle's "Chinese Room" problem. For our theory, the explanation is straightforward. When we have a homunculus manipulating symbols on paper, there are two quite distinct sets of patterns: patterns carried by the head of the homunculus, and patterns carried by the complex system of symbols on pattern. The patterns in the paper may well support their own mind. Even if we assume the homunculus could internalize the rules (which would require vastly more memory than any human has), then there would still be two sets of quite distinct patterns, both of which are present in the homunculus's head. It is a fundamental fact that there can be many different patterns present in a given substrate. Thus, in a single head, there might be two quite distinct phenomenologies, without any overlapping first-person mental states.

It might be noted that this is not quite a double aspect theory in the traditional sense, where the two aspects are the physical and the mental. This theory holds instead that the two aspects are first-person and third-person mental states — both of them mental, though in different senses. We are able to use the resources of theories like functionalism to get us past the first stage, from the physical to third-person mental states, which may be functional states, computational states, or some other patterns. All that is left is Jackendoff's "Mind-Mind" Problem — the bridge from third-person patterns (the "Computational Mind") to first-person mental states (the "Phenomenological Mind"). This is precisely what our posited pattern/information duality achieves.

Finally, it must be glaringly obvious that I have said very little about a very important topic: that is, the relationship between pattern and information. Is this merely a primitive posited relation, or can it somehow be explicated? Some might claim that a given pattern can be *interpreted* as carrying any information you like. I believe that this is using a different sense of the word "information" to mine. The kind of information I have dealt with here is not "information *that*" — it does not need to *refer*. The kind of information I believe in is *intrinsic*. In this sense, there is a *natural mapping* from pattern to information.

Exactly what the nature of this mapping is is very difficult to say. This is why the current theory is not, as yet, a solution to the Mind-Body Problem. It may be the case that the pattern/information relation is a brute fact, in much the same way that the laws of physics are brute facts. We have a certain amount of knowledge about their nomological connection, via our subjective experience. This may or may not be enough to understand all the facts perfectly. If it is not, then it may be the case that we can never know what precisely what it is like to be an alien being (or a connectionist network), as we will not understand precisely the nature of the information that their patterns carry. The outlook is not necessarily so pessimistic, though, and perhaps a better analysis of the nature of pattern and information will tell us everything there is to know. (If pressed for an informal opinion, I would guess that at least nine tenths of the pattern/information relationship will be understandable analytically, with the last tenth up for grabs. If we are not able to understand *precisely* what it is like to be a bat, we will be able to come very close.)

This account has not removed the Mind-Body Problem. Nevertheless, my minimal hope is that it shows what the beginnings of a solution might look like. My maximal hope is that it has removed much of the confusion surrounding the problem, and localized the mystery at one key, primitive locus: the relationship between pattern and information.

References

Block, N. J. (1980). Are Absent Qualia Impossible? *Philosophical Review*, 89: 257-274.

Churchland, P. M. (1985). Reduction, Qualia and the Direct Introspection of Brain States. *Journal of Philosophy*, 82: 8-28.

Clark, A. (1989). *Microcognition.*. Cambridge, MA: MIT Press.

Dennett, D. C. (1982). How to Study Consciousness Empirically: or, Nothing Comes To Mind. *Synthese*, 53: 159-180.

Dennett, D. C. (1985). A Cognitive Theory of Consciousness. In *Brainstorms*, pp. 149-173. Cambridge, MA: MIT Press.

Dennett, D. C. (1988). Quining Qualia. In A. Marcel and E. Bisiach (eds.), *Consciousness in Contemporary Science*. Oxford: Oxford University Press.

Elitzur, A. (1989). Consciousness and the Incompleteness of the Physical Explanation of Behaviour. *Journal Of Mind and Behavior*.

Fodor, J. A. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.

Foss, J. (1989). On the Logic of What it is Like to Be a Conscious Subject. *Australasian Journal of Philosophy*, 67: 205-220.

Harnad, S.R. (1989). Minds, Machines and Searle. *Journal of Theoretical and Experimental Artificial Intelligence* 1: 5-26.

Hofstadter, D. R. (1979). *Gödel, Escher, Bach*. New York: Basic Books.

Horgan, T. (1984). Functionalism, Qualia, and the Inverted Spectrum. *Philosophy and Phenomenological Research*, 44: 453-469.

Jackendoff, R. S. (1987). *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.

Jackson, F. (1982). Epiphenomenal Qualia. *Journal of Philosophy*, 32: 127-136.

Johnson-Laird, P. N. (1983). A Computational Analysis of Consciousness. *Cognition and Brain Theory*, 6: 499-508.

Lewis, D. (1972). Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy*, 50: 249-258.

Lycan, W. G. (1987). *Consciousness*. Cambridge, MA: MIT Press.

McGinn, C. (1989). Can We Solve The Mind-Body Problem? *Mind*, 98: 349-366.

Nagel, T. (1974). What is It Like To Be a Bat?, *Philosophical Quarterly*, 83: 435-450.

Nagel, T. (1986). *The View From Nowhere*. Oxford: Oxford University Press.

Parfit, D. A. (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Rumelhart, D. E., McClelland, J.R., and the PDP Research Group (1986). *Parallel Distributed Processing*. Cambridge, MA: MIT Press.

Searle, J.R. (1980). Minds, Brains and Programs. *Behavioral and Brain Sciences*, 3: 417-424.

Searle, J.R. (1989). Consciousness, Unconsciousness, and Intentionality. *Philosophical Topics*, 17: 193-209.

Shoemaker, S. (1975). Functionalism and Qualia, *Philosophical Studies*, 27: 271-315.

Watkins, M. (1989). The Knowledge Argument Against the Knowledge Argument. *Analysis*, 49:158-160.