Consciousness Studies: Research Prospects in the 'Cradle of Human Consciousness'

Michael Mark Pitman

Introduction

Consciousness seems, at one and the same time, to be both the most obvious characteristic of human existence and yet also the most mysterious. If you were not now conscious, you would be asleep or in a coma, and certainly not aware of the meaning conveyed by the words on this page. Yet we live in a world dominated by a belief in physicalism—roughly, the doctrine that the ultimate constituents of the universe, and thus the 'building blocks' of all entities both conscious (us) and non-conscious (this page, your chair), are purely physical particles—the kinds of 'stuff' studied by physicists. From the point of view of physicalism, it seems thoroughly surprising that there are objects (creatures) in this world that experience qualitative states—states that it *feels like something* to be in. We have made great progress in understanding how certain organisations of matter can 'give rise to' what we call life. After all, we can describe the characteristics and behaviour of living systems in the unproblematically objective terms of a third-person observer, or as Thomas Nagel has put it, from a 'view from nowhere'. Consciousness, however, seems essentially private and subjective. The only conscious experiences we ever have direct access to are our own, and there seems to be an unbridgeable gap between that first-person knowledge of consciousness and the prospect of ever examining a system or creature from the outside in a manner which will identify and adequately characterise any conscious states that it may have.

Consciousness as a subject of investigation has indeed had something of a torrid time within the interdisciplinary field of cognitive science. Cognitive science roughly straddles the disciplines of psychology, philosophy, linguistics, artificial intelligence and neuroscience (Thagard 1996). Each

discipline has, for various and interconnected reasons, historically neglected the field of consciousness studies until quite recently. In psychology, the most obvious explanation for this neglect lies in the dominance of behaviourism for much of the twentieth century. In philosophy, and specifically the philosophy of mind, the seeming intractability of the 'puzzles' surrounding consciousness has seen decades of work on the so-called mind-body problem pay little attention to what, for most non-philosophers (and philosophers when they are off duty), is the most striking characteristic of all things mental. Artificial intelligence (AI) is dedicated to understanding the mind by modelling it (or, for some, replicating it) in machines that most would agree are not conscious at all. And neuroscience, with its 'outsider' view of the brain and its functions, has seemed equally inclined to explain human behaviour with minimal reference to subjective 'inner' experiences of the brains under investigation. Indeed, some philosophers (e.g. Churchland & Churchland 1997) who are particularly keen on both AI and neuroscience seem determined to see conscious phenomena eliminated from any 'respectable' scientific view of the world!

Consciousness studies have, however, recently exploded onto the scene within cognitive science. The reasons for this resurgence of interest are no doubt numerous, but they include the fact that respected, Nobel-prize winning scientists such as Francis Crick (1994) and Gerald Edelman (1989; 1992; Edelman & Tononi 2000) have turned their expertise towards the study of consciousness, and that journals dedicated to the study of consciousness (*Journal of Consciousness Studies, Consciousness and Cognition, Psyche*) have made a significant impact on the academic scene, providing a place for discussion and publication of research for people of all disciplines interested in this most remarkable mental phenomenon.

It is not clear that the same can be said within the South African academic and research community. A casual analysis of articles published in 2000 in the *South African Journal of Psychology*, the *South African Journal of Philosophy* and *Philosophical Papers*, for example, reveals only one article falling in the field of consciousness studies out of 68 articles examined. This article (Gennaro 2000) is, further, the work of an American academic. Of course, these journals do not exhaust the interdisciplinary scope of consciousness studies, nor need local authors submit work in this area to these local journals. On the other hand, the result does suggest that consciousness studies represent a relatively small 'blip' on the local academic radar, at least within psychology and philosophy.

It could be speculated that this is a justifiable reflection of a certain theory-application orientation within academia more appropriate to the pressing social challenges and needs that a developing, African country like South Africa faces. However, if the anecdotal evidence is correct, it does suggest that there are opportunities for research and theorising that are going amiss.

It was therefore with both excitement and sadness that I encountered a paper on consciousness by the late David Brooks (2000) entitled 'How to Solve the Hard Problem: A Predictable Inexplicability' published posthumously in the e-journal, *Psyche*. My excitement was prompted by the interest and challenges that his paper raised. My sadness arose from the fact that I never met Brooks, and that he is no longer here to pursue what, as will be suggested below, is a promising attempt to grapple with the puzzle of consciousness.

The purpose of the remainder of this paper is to briefly outline Brooks' account of consciousness and provide a critical but constructive response to its central claims. The goal will be to delineate a project that is similar (in significant ways) to Brooks' own and which can be linked to other contemporary research within the field of consciousness studies, thus outlining a possible course for further contributions from the South African academy.

Evolutionary Predictability and For-the-organism Properties

Brooks (2000)[1] sets up his account of consciousness as an attempt to address what Chalmers (1995) has called the 'hard problem' of consciousness:

> It is widely agreed that experience arises from a physical basis, but we have no good explanation of why and how it so arises. Why should physical processing give rise to a rich inner life at all? It seems objectively unreasonable that it should and yet it does (Chalmers 1995:201).

This characterisation of the 'hard problem' highlights two features of debates over consciousness. First, there is a conviction amongst many contributors to the debate (Brooks 2000; Chalmers 1995; Searle 1992; 1997)

[1] All further references to Brooks refer to this article.

that the problem arises precisely because we are searching for an explanation of certain *facts* about consciousness, not for accounts of human behaviour that seemingly *explain away* the existence and/or importance of conscious states (e.g. Churchland & Churchland 1997; Dennett 1991). Second, as Brooks notes, the so-called 'hard problem' involves two kinds of questions about this 'rich inner life'—a 'why' question about the need for qualitative states in intelligent (physical) creatures like ourselves, and a 'how' question about the way in which the these qualitative states in fact come into existence[2]. Brooks' account is concerned with the 'why' question, as is the balance of this paper.

The pertinence of the 'why' question can be brought into focus by considering the logical or conceptual possibility of Zombies—creatures physically and behaviourally identical to ourselves except for the crucial fact that they lack qualitative states. That is, there is nothing that it is like to be a zombie, and zombies are able to negotiate the world around them in much the same way as we do despite this absence of qualitative experiences. If the logical possibility of zombies is granted, then we appear to be faced with a choice: either physicalism is false, since we can have atom-for-atom identical creatures without identity in terms of their mental (especially conscious) states; or we accept that the qualitative aspect of mental states is epiphenomenal (causally redundant) since creatures very much like ourselves can achieve all that we can achieve in the absence of such qualitative phenomena.

Brooks' strategy is to grasp the nettle and argue against the logical possibility of zombies so as to preserve both physicalism and the place of qualitative states within the causal order of things. What is offered is an evolutionary/ engineering account of qualitative states (or qualia) committed to the following claim: any device engineered in accordance with the laws of physics which could perform all the tasks of which a human is capable would also *have to have* qualia. In short, zombies are not possible because qualia are *necessary* for the possibility of certain modes of interacting with and behaving in the world that characterise human cognition and behaviour.

One ingenious example offered by Brooks (2000) against the plausibility of zombies involves behaviour in which the role of the sensuous nature of our qualitative states seems incontrovertibly central to our

[2] For an extended debate and exploration of the 'hard problem', see contributions to Shear (1997).

explanations of such behaviour[3]. Consider John and his zombie twin, Zohn, each returning home from work to relax after a stressful day. Each makes a stop at their favourite Chinese takeaway on the way home to order a tangy Shanghai Steak supper (the prospect of another frozen convenience dinner not being commensurate with their need to relax). Before sitting down to their meal, each chooses a CD (say Schubert's 'Death and the Maiden') and turns up the volume on their music system. Subtleties and trivialities aside, Brooks' contention is that whilst John's behaviour is clearly explicable in terms of his preferences between different sensuous, qualitative states (Chinese over bland TV dinners; Schubert over techno-rave), Zohn's behaviour is thoroughly puzzling. With no qualitative states to refer to in explaining Zohn's choices, we are left to marvel at why Zohn did not take the most quick and easy route to relaxation (straight home, quick TV dinner, flick on the radio) rather than go to all the effort that he did. In general then, zombies seem implausible as behavioural 'twins' of humans because we cannot imagine them behaving in ways that we do when the sensuous nature of our qualitative states—specifically, the character of certain sensuous states rather than others—plays a distinctive role in the explanation of our behaviour. In so far as we can imagine zombies, these would be creatures that go through life making choices like our own (what taste to savour, what music to enjoy, what work of art to place on the wall) for no apparent good reason.

This tactic of finding an ineliminable causal role for qualitative states in the explanation of human behaviour is the core of Brooks' strategy. However, Brooks rightly wants to argue that qualitative states play an even more significant role in the causation and explanation of human behaviour than merely determining our more luxurious wants and preferences. Specifically, Brooks wants to argue that, from an engineering or design perspective consistent with the laws of physics, creatures with our *intellectual, cognitive* and *behavioural* capacities should be expected (or predicted) to have qualitative states like our own. We should not, that is, be too quick to assume that the task of dealing intelligently with the world (as we do) can be achieved in the absence of qualitative states. The way to solve the hard problem is to thus produce an account of the predictability of qualitative states in a world governed by the laws of physics.

[3] The example has been developed and embellished, but is based on Brooks' (2000) original example.

The core of Brooks' account involves the idea that humans interact intelligently with the world on the basis of what have been called 'mental models' (see Johnson-Laird 1983; McGinn 1989). The idea is that we interact with the world intelligently on the basis of isomorphic representational models of entities, properties and relations in the world. Brooks' claim is that any system capable of building up a picture or model of the world as sophisticated as ours, and with comparable discriminatory powers, will have to have, as a matter of what he calls *engineering necessity*, (1) perceptual equipment similar to ours, such as edge enhancers; and (2) *quality spaces* distinguished by the distinctive qualitative states (the qualia of colour, sound, taste, etc.) associated with different sensory systems.

Philosophers (at least since Locke) have distinguished between the primary and secondary qualities of the objects of our experience. In the visual modality, shape is a primary quality—a perceived feature of an object that inheres objectively in the object, independently of any perception thereof—whereas colour is a secondary quality—a perceived feature of an object that, although it covaries with objective features of the object (wavelengths of light reflected by the object's surface), is subjective in so far as (a) our colour categorisations depend on objects affecting human consciousness in some way (e.g. producing the experience of red); and (b) interrelations between colours reflect peculiarities in our own perceptual equipment (which wavelengths we are sensitive to and can discriminate between) rather than differences in objects 'out there' in the world.

Brooks' first contention is that for systems like us to transform two dimensional retinal images into (accurate) three dimensional representational models conveying information about primary qualities such as spatial dimension and relative position *requires* representations of objects in a quality space (the 'space' of colour experience) with a range of variations (different colour qualia, such as red, blue, yellow, shades of white through grey to black) that reflect or model dimensions and positions of objects in the world by way of relative similarities and dissimilarities within that quality space. More simply, to model 3-D shape (and position) requires edge enhancers of some kind, and this is to be achieved by differences in quality (distinct colour qualia and shades thereof) within a quality space. Of course, Brooks does not need to argue that our particular colour qualia are necessary to achieve this edge-enhancing function (echolocation seems to serve bats pretty well when it comes to 3-D modelling), but only that *some* distinctive quality space with a requisite range of variations is necessary t model the world in 3-D as well as

we do (echolocation may be fine for navigation, and it works in the absence of light, but colour modelling probably enables modelling and behavioural repertoires that echolocation does not).

There is thus an adaptive advantage to be gained by modelling important primary properties of the world around us by way of qualitative variations within an experienced quality space. Brooks' second contention is that there is further adaptive advantage to be gained by the association of distinctive quality spaces with different perceptual modalities—colour with sight, sound with hearing, taste/smell with the olfactory senses, 'feel' with touch and proprioreception. The advantage to be gained is simply that this allows for (a) easy discrimination between sources of perceptual input, and (b) easy direction of sensory and attentional resources to the sources of salient input (e.g. turning one's ear towards a suspicious sound rather than one's eyes, nose or hand).

Systems with edge-enhancers and quality spaces do not, however, strike Brooks as sufficient to secure the necessity of qualitative states. He cites the example of the filming and broadcast of a television show: at every stage of this process, from the light entering the camera lenses through the radio waves sent by the transmitting broadcaster to the reception and transformation of these waves into patterns of pixel activation on a receiving television screen, there is an isomorphic model of the original 'scene' that preserves or models all the requisite spatial relations. Yet the system is completely 'blind' to this accurate modelling. What is needed, argues Brooks, is to add the self-awareness that characterises our own perceptual and intellectual activities. For a system with edge-enhancers and quality spaces to behave as we do requires awareness *within the system* of the representational medium *as being medium of representation* (and this, needless to say, the 'broadcast system' lacks). That is, the system requires an awareness that certain properties of its internal states are properties of its own subjective perceptual states (rather than of the world itself) that act as labels, codes or filing tags within a representational medium whose function is to model the world outside. These 'tags' are such that they are regularly associated with a distinctive class of external things, and such that the system is readily able to distinguish one tag from another. It is only when the system becomes aware of these tags *as being* tags that these qualitative features acquire all the features we associate with qualia. These properties of such systems Brooks calls 'For-the-organism' properties.

To clarify this argument, let us briefly consider Brooks' (2000) reference to research on image rotation (Kosslyn & Pomerantz 1981). This research has found that the response time of a subject asked to describe (or decide) how an object would look from a different angle or perspective *depends on* the angle through which the object would have to be rotated to see it from that point of view. This finding has been used to support the idea that the response times vary because the subjects in fact rotate a mental image or model of the object in producing their response. The research not only supports the 'mental modelling' approach mentioned earlier, but helps illustrate Brooks' point about awareness of tags. The research subject, it seems, is able to mentally rotate the requisite object image (in part) *because* they are aware of the differential tagging of the object and its features and that of the background against which it is rotated. It is the awareness of these discriminating For-the-organism properties and their mapping onto the object in question that enables the intelligent response of the subject to the task. The subject, in interaction with the object, picks out qualia as the requisite For-the-organism properties, and the resultant manipulations/ calculations are carried out in virtue of these For-the-organism properties.

Brooks, however, leaves the task of establishing the necessity of these For-the-organism properties as an open-ended strategy to be pursued rather than a thesis already established. Specifically, Brooks leaves us with the task of pin-pointing some intellectuo-perceptual capacity *C* such that (1) we (humans) have *C* and (2) the possession of *C* underpins our ability to report on and take pleasure in our qualitative states. Furthermore, any creature (or system) without qualitative states would lack *C*, thus establishing the necessity of qualitative states. That is, any device *D* engineered, by evolution or otherwise, in accordance with the laws of physics *must* have qualitative states if it is to have the capacity *C*. Identifying *C* will thus secure the place of qualitative states (and specifically, of the *qualitative aspect* of qualitative states) as having a distinctive causal and functional role within the cognitive and behavioural activities of creatures like ourselves—and secure this in a way entirely compatible with physicalism.

If this is an accurate characterisation of the project and challenge with which Brooks has left us, then it is to the critical assessment of this project that we must turn before charting, if possible, a way in which to take it forward.

The Limits of Physicalism, Science and Evolutionary Explanations

In critically evaluating Brook's (2000) suggested strategy for dealing with the hard problem of consciousness, three areas of contention will be highlighted: (1) the place of physicalist assumptions in articulating the puzzle(s) of consciousness; (2) the usefulness and the limits of evolutionary explanations; and (3) the role of self-awareness within Brooks' account.

As noted at the outset, the assumption of physicalism is widespread, and nowhere more so than within the philosophy of mind, where much (if not most) of the last three decades of work in this field has been centred on the project of 'naturalising the mind', in the sense of providing an account of mind that secures its respectability and place within a monistic, physical universe. Yet physicalism is an *assumption*—or rather a set of assumptions—that is as regularly left undefined or underspecified as it is assumed (Crane & Mellor 1991). There are those (of whom Crane & Mellor 1991; 1995) are a most notable example) who are inclined to argue that when the assumption of physicalism is unpacked, and specifically unpacked in such a way as to render the mental domain uniquely distinct, mysterious and/or problematic in some way, the plausibility of physicalism comes seriously under question (see also Pitman 1998).

Although highly pertinent to the current debate, the substance of these arguments against the 'obviousness' of physicalism cannot be dealt with within the scope of this paper. What should be noted is that one need not be a committed physicalist to find consciousness puzzling. Consciousness is puzzling and mysterious (and fascinating) because it is *the* quintessentially subjective phenomenon in the universe that, for this reason, seems to resist description and explanation from 'the outside'. It is further puzzling because we have a tendency to think that, when neuroscience fills in the gaps in its understanding of the brain mechanisms associated with consciousness, there will be no causal or functional role left for the qualitative or phenomenal aspects of these conscious states: each brain state will follow smoothly on from the preceding one with no apparent need to refer to other causal factors, not least to qualia and the like. There is a strain of physicalism in this tendency, but it is not central to our present concerns.

One last point is that physicalism is a particular brand of what were referred to earlier as 'naturalisation projects' within the philosophy of mind. The guiding motivation behind such naturalisation projects is that if the mind,

or various aspects of the mind, cannot be naturalised in some way (usually by way of providing at least sufficient non-mental conditions for the occurrence of these phenomena), then some dire consequence will follow (Pitman 1998; Stich & Laurence 1994). The usual candidates for such dire consequences are that the mind is really illusory or irreal, or more often that the mind (or rather its features, such as consciousness) is epiphenomenal. This approach seems to me mistaken. The brain is the most complex object that we know of in the universe (Edelman & Tononi 2000), and to blithely assume that we will come to understand all its mysteries, and worse, that failure to do so would render some of its most remarkable characteristics (such as the capacity to 'produce' consciousness) epiphenomenal, seems both arrogant and defeatist.

The point of these observations is not to castigate Brooks or anyone else for their seeing a pressing need in pursuing physicalist or naturalising agendas. Rather, they are presented as a means to suggesting that philosophers and cognitive scientists should be more cautious in how they frame their puzzles and their criteria of success, especially when it comes to the mind. *Trying to find* necessary and/or sufficient conditions of a certain kind for consciousness is a worthwhile project, then, as long as we don't attach too high a cost to the possible failure of our endeavour. Physicalist naturalisation projects are worthy paths of inquiry, but they should not be viewed as the only game in town.

The fault, therefore, in Brooks' account that we might endeavour to avoid is his evident commitment to physicalism. Whether or not physicalists can sleep easy at night (Brooks' phrase) over the problem of consciousness, we can simply concern ourselves head on with the attempt to answer the hard question as to why (and, if possible, how) consciousness arose in creatures like ourselves. And in pursuit of this goal, we should be as willing to question, at any time, the background assumptions within which we are operating (e.g. physicalism) as we are to critically question the plausibility of any account of consciousness on offer.

The second point of contention concerns the limits of evolutionary explanations, both in general and with regard to consciousness in particular. Evolutionary explanations are, of course, perfectly suited and appropriate to answering 'why' questions about the origins and functions of features that we find in living creatures like ourselves. To assign a function to consciousness that both explains human behaviour and outlines the adaptive advantages conferred on creatures endowed with such functional capacities would indeed go a long way (if not all the way) to explaining why consciousness emerged in

our universe in creatures like ourselves. Brooks' positive account is an admirable instance of just such an attempted explanation, and provides considerable 'ammunition' to be used against those inclined to explain consciousness, or its causal efficacy, away.

However, evolutionary explanations are ill suited to the task of establishing the *necessity* of some or other function being achieved in a certain way (see Flanagan & Polger 1995; Polger & Flanagan 1996). The easiest way to see this is by considering the nature of adaptive advantage and the role of chance within evolution. Assigning and explaining function by way of adaptive advantage depends on the idea that characteristics emerging by descent-with-modification result in differences in relative environmental fitness, survival and procreative success. From the point of view of evolution, it matters not so much how a function is performed as simply that it does manifest in a breeding population. For this reason, evolutionary accounts of function cannot in general rule out the possibility of other means to performing that function. Evolutionary necessity, if there is such a thing, is thus a highly contingent affair tied to the peculiarities of particular species in particular environments with particular histories of descent. It is not well suited to the task of being used in arguments against the (logical or metaphysical) possibility of a certain function being carried out in different ways by different or even similar creatures that have, for example, radically different histories.

It might be countered that evolutionary arguments *are* at least suited to providing statistical arguments in favour of the necessity of a function being carried out in a certain way, given the time (and genetic and environmental diversity) 'available' for evolution to have come up with suitable alternatives. No doubt there is something to this point, but it brings us to the issue of chance in evolution. The difficulty of providing evolutionary explanations of anything derives not only from lack of the definitive evidence required to establish such explanations, but also because chance has such a significant possible impact on the course of evolutionary history. Many would be familiar with the common speculation that the 'age of the mammals' might never have come about had it not been for the catastrophic (meteoritic) extinction of the dinosaurs. But for that cosmic interference, the phylogenetic scale might look rather different today. Similarly, in relation to our case of the possibility of zombies, Polger and Flanagan (1996) ask us to imagine the scenario in which our ancestors cohabited the earth with zombie counterparts who were functionally identical. By way of a freak volcanic eruption, all these zombies

were eliminated, leaving our own conscious forebears alone with their now distinctive, but only accidentally unique functional capabilities. Given such potential for chance interference in the otherwise 'neat' evolution of function by selective descent-with-modification, evolutionary explanations are again unsuited to producing accounts of the necessity of performing functions in particular ways.

Polger and Flanagan's (1996) claim is thus that evolution alone cannot rule out the possibility of conscious inessentialism (i.e. the claim that consciousness is not essential to our human cognitive and behavioural capacities) that the case of zombies was designed to illustrate. To this extent, we should be wary of Brooks' attempts to establish the necessity of consciousness to the performance of certain intellectuo-perceptual operations that we humans are capable of. To ask that consciousness be necessary is to ask too much of our evolutionary account of human behaviour.

Does this mean that evolutionary explanations cannot be used to answer the hard 'why' question? Certainly not. First, it simply means that our sights should be set lower—we should attempt to establish that, as a matter of contingent fact, consciousness evolved in humans because it enabled certain intellectuo-perceptual capacities that conferred selective and adaptive advantages on our ancestors. That would surely be an answer to the 'why' question that is as naturalistic as one could desire.

But there is a second point to be held in mind—namely that we might be able to establish (or at least argue for) the 'contingent necessity' of consciousness relative to certain functions of which we are capable as humans. That is, whatever the logical or metaphysical necessity of consciousness to certain capacities that humans happen to have, we might still hope to eliminate the possibility of zombies in our world by arguing for the necessity of consciousness to certain intellectuo-perceptual capacities in creatures with a biological/physiological makeup like ours. Our ideal would thus be a scenario in which we can plausibly assert that creatures with brains like ours could not achieve what we achieve except in so far as such creatures have conscious, qualitative states. We might not attain this ideal of 'contingent necessity', but it is worth aiming for, and it still leaves us with the possibility of obtaining a 'purely' contingent evolutionary explanation that is still sufficient to our needs. Brooks' account thus essentially points us in the right direction, only with slightly too much confidence in what we will be able to achieve by following his strategy.

An example of this kind of approach to understanding the significance of consciousness can be found in what Baars (1997) calls 'contrastive phenomenology'. Focussing on humans (both normal and abnormal), the idea is that

> by contrasting similar conscious and unconscious mental representations in beliefs, perception, selective attention, imagery and the like, *we can gather a set of sound empirical constraints on the distinctive properties of consciousness as such* (Baars 1997:187; e.i.o.).

Baars' strategy is thus less concerned with establishing the necessity of consciousness for any given task or capacity as it is with better understanding the nature and role of consciousness in human performances of a range of tasks and operations. Because the analysis is contrastive in nature, the outcomes of such investigations are supposed to provide a clear account of what it is that consciousness adds to our mental and behavioural capacities that is not there when we operate with similar capacities that do not involve consciousness. Clearly, a contrastive account of what consciousness *adds* to our mental and behavioural capabilities would go a long way, in evolutionary terms, to explaining *why* creatures like ourselves are conscious (for more on contrastive analysis and Baars' Global Workspace Theory, see Baars 1996).

The third point of contention is more concerned with the details of Brooks' (2000) argument. As we have seen, Brooks introduces self-awareness at a seemingly crucial point in his argument where he introduces the notion of For-the-organism properties. It is the awareness of qualia as tags within a representational medium, or awareness of these tags as tags for the organism, that Brooks thinks is the key to such cognitive capacities as image rotation and, by generalisation, mental manipulations, operations and calculations generally. Whatever the virtues of these insights, one might legitimately raise the concern that Brooks has too quickly introduced a form of conscious awareness that is itself distinctively human[4]. Cognisant of the possibility of 'blind' systems equipped with edge-enhancers and quality spaces (e.g. the TV

[4] In Edelman's terms, as will be discussed, the kind of self-awareness that Brooks introduces requires an ability to explicitly reconstruct a conscious mental scene, and this capacity only comes with the kind of higher-order consciousness found in humans and some of the higher apes.

broadcast), Brooks introduces a feature of self-awareness which itself remains unexplained within the account, and which leaves 'why' questions about the significance of phenomenal states in non-human conscious animals beyond the scope of the account. If the accusation is correct, then we need to seriously question whether Brooks has pointed to a way in which the hard problem can be resolved with sufficiently generality to deal with these non-human varieties of consciousness.

The objection highlights a concern that Brooks may have tried to achieve too much too quickly. Conscious states are characteristic of many creatures other than humans, and from the perspective of evolution, there are thus many potential consciousness-function links that we could explore to establish the importance and predictability of qualitative conscious states before we arrive at the relatively sophisticated functions of the human mind. We can, that is, distinguish between primary consciousness (that we share with many other creatures) and higher-order consciousness (of which we share some aspects with other primates, and other aspects of which we seem to possess uniquely) (Edelman 1989; 1992; Edelman & Tononi 2000), which brings us to an examination of some contemporary research that holds promise for taking Brooks' strategy forward.

Primary and Higher-order Consciousness

That Brooks' may have moved too quickly to a focus on human consciousness is most evident in the concern expressed in his account for being able to fully account for qualia as these are understood within the traditional philosophical debates over consciousness. Possible reasons for this error are not difficult to find. Edelman (1989), in the process of developing a theory of consciousness grounded in neuroscientific research rather than philosophical argumentation, nevertheless reflects on important philosophical issues raised in the study of consciousness. One of these issues is precisely that it is only humans who have qualia, not because other creatures do not possess primary consciousness and its attendant phenomenal states, but because they do not possess the requisite higher-order conscious capacities, and specifically linguistic capacities, to be able to *report* on these phenomenal states. In Edelman's (1989:166f) terms, a creature may have phenomenal experience,

> but it is not *subjective* experience—there is no subject or person to

> make discriminations or reports of that phenomenal experience over time... [W]e cannot call such states qualia in the sense that we can call our own states qualia[5].

In this sense, Brooks (2000) is correct to place emphasis on self-awareness when it comes to explaining qualia. But the criticism stands, because to move too quickly to this focus on distinctively human consciousness is to gloss over other manifestations of conscious states with qualitative aspects that, from an evolutionary perspective, would presumably have preceded human consciousness, and indeed form the basis for human higher-order consciousness. If, therefore, the task is to generate an evolutionary account of qualitative states (even an account of qualia *per se*), it seems far more suitable to start by explaining the causal and functional role of qualitative/ phenomenal states first before moving onto the more complex forms of consciousness of which humans are capable.

This is Edelman's (1989; 1992; Edelman & Tononi 2000) starting point. Edelman and Tononi (2000) describe primary consciousness as the capacity to form a mental scene, and in most creatures (other than humans) this capacity is accompanied by limited semantic or symbolic capabilities and no true linguistic ability. Higher-order consciousness, they claim, involves capacities that flourish in humans and that *presuppose the coexistence of primary consciousness*. Higher-order consciousness

> is accompanied by a sense of self and the ability in the waking state explicitly to construct past and future [mental] scenes. It requires, at a minimum, a semantic capability and, in its most developed form, a linguistic capability (Edelman & Tononi 2000:102).

The significance of this distinction to both the study of consciousness generally, and to an evaluation of the promise of Brooks' account, becomes evident when one examines Edelman's (1989) theory of the evolution of primary consciousness. Edelman (1989) claims that primary consciousness

[5] It should be noted that Edelman's use of 'subjective' is thus different to the sense of 'subjective' used elsewhere in this paper. Following Nagel (1974), the subjectivity of conscious experience refers to there being something that is it like to be a creature with those experiences. It does not imply, as it does for Edelman, a sense or experience of a self as subject of those experiences.

arose as a result of two developments in the evolution of brain systems: (1) the development of special memory repertoires composed of neuronal groups dedicated to storing past matchings of value and perceptual categorisation; and (2) the development of reentrant signalling between these special memory repertoires and neuronal groups currently devoted to sampling of the sensory environment for perceptual categorisation in all sensory modalities.

Three points of clarification are required before proceeding to Edelman's (1989) explanation of primary consciousness itself. First, 'value' in this context is probably best understood as 'salience for the organism', and is determined by (largely proprioreceptive) homeostatic mechanisms in the brain concerned with the immediate needs for survival of the individual creature. Second, 'perceptual categorisation' in this context is taken to involve categorisation at the level of neuronal groups (i.e. patterns of neural activation in relevant sensory cortices), such that environmental stimuli will fall in the same perceptual category if they trigger the same (or a significantly related) neuronal group. Value-perceptual categorisation pairings are thus to be understood as pairings of environmental stimuli (via exteroceptive signals) with 'salience to the organism' values determined by homeostatic (interoceptive) signals from within the organism. To construct a crude example, lion-shape pattern of stimulation would be paired with fear-and-flight salience. Both systems of value-perceptual categorisation memory and of current perceptual categorisation can (and probably do) operate in the absence of primary consciousness.

Third, the concept of reentry and reentrant pathways requires brief explanation. Edelman and Tononi (2000:48) describe reentry as a process of cyclical signalling:

> ...the ongoing, recursive interchange of parallel signals between reciprocally connected areas of the brain, an interchange that continually coordinates the activities of these areas' [neuronal] maps to each other in space and time. This interchange, unlike feedback, involves many parallel pathways and has no specific instructive error function associated with it. Instead, it alters selective events and correlations of signals among areas and is essential for the synchronisation and coordination of the areas mutual functions.

Reentry thus involves massive numbers of parallel, reciprocal pathways between different areas in the brain which, when operating, synchronise (or

even activate) the functioning of diffuse but related areas and systems within the brain. For example, reentrant signalling between the visual cortex and the part of the motor cortex dealing with eye movement would assist in the synchronisation of such movements in relation to what is being processed and attended to within the visual cortex.

The key to Edelman's account of primary consciousness thus lies in reentrant pathways and activity linking the special memory repertoire and the areas associated with each exteroceptive sensory modality charged with perceptual categorisation. The account has two key facets. First, these reentrant pathways allow the animal to coordinate activation of past value-category pairings when the requisite neural group for that category is activated in the perceptual categorisation area of the brain. This enables the real time interaction between memories of past value-category associations and current perceptual categorisations *before* the value-determining parts of the nervous system have a chance to change these memories based on current homeostatic status. That is, the organism is enabled to maintain a degree of independence between acquired memories and current, ongoing perceptual categorisations, such that these memories can influence the relative salience-to-the-organism of particular perceived events, and assist in the choice of goals and actions. Such a mechanism is the fundamental basis of learning.

The second facet of the account is that this discrimination takes place in parallel across *all perceptual modalities*. That is, the reentrant signalling enables a unification and synchronisation of 'processing' in all modalities, with each modality/ area able to simultaneously activate associated value-category pairings within the memory system. Taken together, these two facets reveal the significance of the unified conscious phenomenal experience that thus arises for the organism. In the absence of these mechanisms of primary consciousness, salience in the array of signals arriving simultaneously at the various sensory modalities and perceptual categorisation areas would be determined *almost entirely* by the dominance of one external event over another in each of the parallel sensory channels (e.g. the loudness of a sound, size or proximity of an object, strength of a smell), rather than by the adaptive values of the animal. Primary consciousness thus 'allow[s] an animal ... to direct attention to particular events in a selective fashion that serves its own adaptive needs' (Edelman 1989:98).

This empirically based account of primary consciousness seems to me highly compatible with the spirit of Brooks' (2000) proposed strategy for solving the hard problem, and bodes well for further empirical and

philosophical work in this area. First, Edelman's (1989) explanation of the significance of the phenomenal or qualitative aspect of primary consciousness also seems to draw on a view of these qualitative features as tags for the organism. Where the explanation seems to differ from Brooks' account is that the significance of these tags lies more at the level of tagging (or representing in real time) remembered associations of value or salience for the organism. Edelman's theory does not argue against the significance of quality spaces and edge-enhancing mechanisms, but it suggests that these are not sufficient to require consciousness. Primary consciousness becomes necessary in linking perceptual categorisations to memories of value-category pairings, and it is this feature of consciousness that has such great adaptive advantage for the organism.

Second, Edelman's theory also makes much of the adaptive value that is gained by having a mechanism that allows for appropriate direction of attentional resources. Again, however, his theory differs from Brooks' account by moving beyond mere perceptual-sensory considerations (presumably coordination of sensory receptors with perceptual categorisation activity in the cortex could, on this account, be achieved 'blindly'). Yet the adaptive advantage gained is similar. Primary consciousness enables a more adaptive use of attentional resources by (1) integrating and coordinating the perceptual categorisation activities in all modalities (i.e. integrating what would otherwise be parallel processes) into a unified mental scene, and (2) by efficiently attaching salience-for-the-organism values to categorisations within this integrated mental scene such that attention can be directed, both within and across sensory modalities, to environmental stimuli that have most adaptive significance for the organism. In a sense, primary consciousness is thus a vehicle to freeing an organism from 'the prison of the present', as well as making the organism less of a slave to the intensity of sensory-perceptual input. Brooks was thus correct in his rough outline of the adaptive significance of integrated mental scenes constituted by quality spaces associated with each sensory modality. What was missing was an account of how salience-for-the-organism enters the picture. On Edelman's theory, it is the connection between perception/categorisation and salience values that holds the key to understanding the evolutionary advantage of primary consciousness.

Unanswered Questions: Future Directions

Where does this leave those of us interested in the interdisciplinary study of

consciousness? In the area of philosophy, it seems that the above integration requires more detailed development into a systematic attempt to address the hard problem of consciousness. This will, at the very least, involve (a) deciding *how much* of the hard problem can be solved by way of such an account of the evolutionary significance of primary consciousness (which, we should remember, requires basic phenomenal or qualitative states rather than fully fledged qualia); and then (b) developing the theory, in conjunction with an account of higher-order consciousness, to provide an explanation of the evolutionary significance of qualia themselves, as well as the other features of higher-order consciousness described by Edelman (1989; Edelman & Tononi 2000). Of key significance will be the evolution of language, an account of its relations to and dependence on primary consciousness, and exploration of the cognitive, perceptual and behavioural capabilities that language confers on creatures endowed with linguistic higher-order consciousness[6].

In the field of (evolutionary) neuroscience, a wealth of empirical and theoretical work lies in wait for the would-be researcher. Edelman himself is insistent that his model of primary consciousness (Edelman 1989), and his more recent 'dynamic core hypothesis' (Edelman & Tononi 2000) that builds on this model, are testable and *refutable* scientific hypotheses that require investigation and confirmation rather than *a priori* defences. Sadly, the research technologies involved in much of this research (PET and functional MRI scanning) are extremely costly for a society and academic environment like our own. But with the appropriate level of interest and international collaboration, South African researchers would do well to contribute to this exciting and boundary-pushing area of research. We do, after all, live in the 'Cradle of Humankind', and it would be fitting for us to contribute something to understanding the evolutionary 'birth' of human consciousness so as to complement our considerable contributions to, for example, paleoanthropology and the fossil record.

School of Human and Community Development
University of the Witwatersrand

[6] A current project of my own involves investigating the role of higher-order consciousness and its associated representational and linguistic aspects in generating the behavioural capacities that we typically associate with free will.

References

Baars, Bernard J 1996. *In the Theatre of Consciousness: The Workspace of the Mind.* New York: Oxford University Press

Baars, Bernard J 1997. Contrastive Phenomenology: A thoroughly Empirical Approach to Consciousness. In N Block, O Flanagan & G Guzeldere (eds): *The Nature of Consciousness: Philosophical Debates.* Cambridge: MIT.

Brooks, David 2000. How to Solve the Hard Problem: A Predictable Inexplicability. *Psyche* 6. http://psyche.cs.monash.edu.au/v6/psyche-6-04-brooks.html

Chalmers, David 1995. Facing up to the Problem of Consciousness. *Journal of Consciousness Studies* 2:200-209.

Churchland, PM & PS Churchland 1997. *On the Contrary: Critical Essays, 1987-1997.* Cambridge, Mass.: MIT Press.

Crane, T & DH Mellor 1991. There is no Question of Physicalism. In Mellor, DH (ed): *Matters of Metaphysics.* Cambridge: Cambridge University Press.

Crane, T & DH Mellor 1995. 'Postscript' to 'There is no question of physicalism'. In Moser, PK & JD Trout (eds) *Contemporary Metaphysics: A Reader.* London: Routledge.

Crick, Francis 1994. *The Astonishing Hypothesis: The Scientific Search for the Soul.* London: Simon & Schuster.

Dennett, Daniel C 1991. *Consciousness Explained.* Boston: Little, Brown & Co.

Edelman, Gerald M 1989. *The Remembered Present: A Biological Theory of Consciousness.* New York: Basic Books.

Edelman, Gerald M 1992. *Bright Air, Brilliant Fire: On the Matter of Mind.* London: Allen Lane/Penguin.

Edelman, Gerald M & Giulio Tononi 2000. *Consciousness: How Matter becomes Imagination.* London: Allen Lane/Penguin.

Flanagan, Owen & Thomas Polger 1996. Zombies and the Function of Consciousness. *Journal of Consciousness Studies* 2:313-321.

Gennaro, RJ 2000. Fiction, Pleasurable Tragedy, and the HOT Theory of Consciousness. *Philosophical Papers* 29:107-120.

Johnson-Laird, Philip 1983. *Mental Models.* Cambridge: Cambridge University Press.

Kosslyn, S & J Pomerantz 1981. Imagery, Propositions and the Form of Internal Representations. In Block, Ned (ed): *Readings in the Philosophy of Psychology* 2. Cambridge, Mass.: Harvard University Press.

McGinn, Colin 1989. *Mental Content*. Oxford: Basil Blackwell.

Nagel, Thomas 1974. What is it Like to be a Bat? *Philosophical Review, 83*:435-450.

Pitman, Michael M 1998. Mental Causes, Supervenience and Psychological Explanation: A Defence of the Causal and Explanatory Relevance of Content. Unpublished MPhil dissertation, University of Cambridge, U.K.

Polger, Thomas & Owen Flanagan 1996. Explaining the evolution of consciousness: the other Hard Problem. Paper presented at the Tucson II Conference *Towards a Science of Consciousness*. Paper available at: http://www.duke.edu/~twp2/tucson2.html

Searle, John R 1992. *The Rediscovery of Mind*. Cambridge, Mass.: MIT Press.

Searle, John R 1997. *The Mystery of Consciousness*. London: Granta Books.

Shear, Jonathan (ed) 1997. *Explaining Consciousness—the 'Hard Problem'*. Cambridge, Mass.: MIT Press.

Stich, SP & S Laurence 1994. Intentionality and Naturalism. *Midwest Studies in Philosophy* 19:122-142.

Thagard, Paul 1996. *Mind: Introduction to Cognitive Science*. Cambridge, Mass.: MIT Press.