How to Solve the Hard Problem: A Predictable Inexplicability

David Brooks
University of Cape Town
SOUTH AFRICA

Copyright (c) David Brooks 1999

PSYCHE, 5(36), December 1999

KEYWORDS: consciousness, qualia, zombies, physicalism, dualism, epiphenomenalism, explanation.

ABSTRACT: Qualitative states are no threat to physicalism. They have a causal effect upon the world in virtue of their qualitative nature. This effect is exploited in biological mechanisms for representing the world. Representation requires differential responsiveness to different perceived properties of things. Qualia are taken to be tagged properties of internal representation models. These properties are properties for-the-organism. Such for-the-organism properties are to be expected in beings which perceive the world and interact with it intelligently. Consciousness presents a problem for science. Human beings (and probably some animals) are conscious of the world and of themselves. In so far as science has the ambition of explaining everything consciousness is another unexplained phenomenon. However some claim that it is distinctive and different in kind from other problems which science hopes to solve using methods which have been successful up until now. It may indeed be so different that we have to adopt a dualistic metaphysics and accept that there is more to the world than physics knows. In this paper I intend to outline how the physicalist should fight back.

1. The Problem

David Chalmers (Chalmers, 1995) distinguishes what he calls the easy problems presented by consciousness from the hard problem. Among the easy problems Chalmers lists our ability to describe and conceptualise the world around us, the ability to attend to one thing rather than another, the difference between wakefulness and sleep, and the ability to report on our own mental states. These are "easy" problems because we can understand what it would be to solve them. The hard problem concerns qualitative states. Frank Jackson's story (Jackson, 1982) about Mary, a future neurophysiologist, puts the point well. Mary knows all there is to be known about the physical side of colour
perception but has always seen in black and white because she has been brought up in a black and white room. Mary knows all that science can ever hope to teach us but she still does not know what it is like to see red as opposed to blue. One of the serious problems confronting science is the question of how "raw feels", "Qualia" etc. fit into our overall picture of the world. Chalmers puts it thus:

> 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, p.201)

Chalmers can be seen as asking two separate questions: the "why" question and the "how" question. The "why" question asks why our physics should have given rise to qualitative states when it seems that the job of dealing with the world intelligently could be done perfectly adequately without them. The "how" question asks how it is that the trick is done given the perverse fact that intelligent beings in this world do possess the extra baggage of qualitative states. This paper is primarily concerned with the "Why" question.

Chalmers suggests that in the face of the hard problem we should be open to the possibility of a large scale metaphysical dualism. Physicalism may not be up to the task. There is a particular thought experiment which lends great credence to his suggestion. This is the apparent logical possibility of zombies. Philosophical zombies behave exactly like intelligent human beings but do not have qualitative states.

The being underlying Chalmers' account of the hard problem is a zombie twin of a real person. Let the real person be John and his double be Zohn (zombie John). Zohn is made of flesh and blood and is neurophysiologically identical with John. John and Zohn are atom for atom the same, but John has qualitative states while Zohn does not. If Zohn is logically possible this implies first of all that the qualitative is not supervenient upon the physical. If phenomenon P is supervenient upon a substructure S then it is not logically possible for two beings to have the same substructure S and yet differ with respect to P. John and Zohn by definition share their substructure yet differ with respect to their mental states. Furthermore, since John and Zohn are physically just the same, but different in qualia, it follows that qualia must be nonphysical. So if zombies are logically possible, physicalism is false.

2. The Physicalist Response

Many people find the thought experiment convincing and feel that Zohn is logically possible, even though this means that supervenience has to go. The physicalist cannot accept the possibility of Zohn. I wish to argue that this is a principled refusal, and not a defensive knee jerk. The physicalist should not be fazed by the peculiarities of consciousness. I will claim that the distinctive features of qualitative states are not only
not incompatible with physicalism, they are to be expected. It is not just that John has qualia, it is that it would be surprising if he did not. This is because any device engineered in accordance with the laws of physics which could perform the tasks which John can perform would also have to have qualia. Qualia are not a metaphysical extra but a physical necessity.

The first move the physicalist must make is to claim that qualitative states are physical states which have a genuine causal role in the generation of John's behaviour. My physicalist response to Chalmers' thought experiment rejects the supposition that qualitative states are mere epiphenomena.

If zombies are logically possible then physicalism is false in this world. That is, our physics does not logically imply qualia. But in Chalmers' thought experiment the world also seems to be, in a sense, physically closed. After all Zohn behaves exactly like John. It follows that either qualia are epiphenomenal or behaviour is overdetermined. If qualia are epiphenomenal, then in addition to the physical goings-on there is a parallel psychological world of qualia which run in perfect synch with the physical world but which have no causal effects. If behaviour is overdetermined, then John's actions are caused by his qualitative states in addition to being caused by his physical states. He would have done what he did even without qualia, as shown by the fact that Zohn behaves just like him.

Overdetermination is not a serious contender. On the one hand Occamist considerations argue against the metaphysical extravagance of doubling the number of causes. On the other hand overdetermination implies that the mental cause could operate in the absence of the physical cause and there is no evidence of this.<1>

The alternative is epiphenomenalism, which I wish to reject on the grounds that qualia are part of the causal order of the world. When John acts very often his actions are determined by the phenomenal character of his qualitative states. Why did he put that disc on his player? Because he wanted to hear its brassy sound. The desire for that phenomenal sensation moves him to the player and the pleasure caused by those qualia when he receives them account for the ecstatic expression on his face. Aesthetic considerations can play a considerable role in our lives and qualitative states form an intrinsic part of our aesthetic appreciation. Without qualia we might be able to appreciate the elegance of mathematical proofs but music and painting would be meaningless.

What, then, are we to make of Zohn? Lacking John's qualia, he seems to be an incomprehensible and possibly dishonest zombie who wastes great amounts of money buying expensive foods he cannot taste and going to operas he cannot hear and who, worst of all, spends hours talking about things he's never experienced. (Zohn even does these things in private, when no one can be impressed by him!)

If we reject this bizarre possibility, it must be because we accept that qualia are truly part of the causal order. This means that a creature which lacked qualia would have to be blind and deaf, and could not behave just like one who can see and hear. If we accept that
Qualia are part of the causal nexus, so that they cause and are caused by physical events, then we seem to be left with two options - physicalism (and the logical impossibility of Chalmers' zombies) or overdetermination. I have already given reasons to be dubious about overdetermination. In the last resort, it is no better than epiphenomenalism. Both allow blind Zohn to behave just like sighted John, which is absurd.

My argument so far has been that Chalmers' thought experiment is incompatible with physicalism. Two beings could not be physically identical and psychologically different unless we accept one or other form of dualism. The dualist supposition allows for the possibility of Zohn at the cost of making him incomprehensible. It conflicts with our belief in the causal efficacy of qualitative states. What is more, there are independent reasons for the superior plausibility of physicalism. So, probably, physicalism is true and Chalmers' zombies are impossible. But why should they be impossible when they seem so easily describable and imaginable?

My provisional answer to Chalmers' "Why" question is that we should not be too quick to assume that the job of intelligently dealing with the world can be done without qualitative states. True intelligence by contrast requires a being to act upon information received from the world in a way which serves its interests. Intelligent deliberation on the basis of perceptual information derived directly from the world may only be possible for creatures with qualitative states. Another piece of fine tuning to be found in this universe in accordance with the weak anthropic principle may be that intelligent life can only arise in a universe whose physics permits qualitative states. In the alternative world where physics is just like ours only it does not give rise to qualia, Zohn could not have evolved. Moreover when Zohn is constructed out of twin-atoms to be twin-physically identical to John, he turns out to be insentient and lacks John's capacity for dealing with the world intelligently.

As I see it, the first step on the way to solving the hard problem is to accept the identity thesis and accept that qualia are physical brainstates. Brain states are able to cause the types of behaviour that they do because they are qualia. If they were not qualia they would not be so causally efficacious. My argument for this identification has been by showing that physicalism is superior to its dualist alternatives once we accept that qualia have a significant causal role. There are further reasons for accepting the identification. The most compelling of these derive from science. Qualitative states as we experience them are not homogenous sensory atoms, but have a structure.

As I look at my bookshelf, I see the spines of my books as a series of adjacent rectangles. If I walk over to the shelf and pull on one of the rectangles, out will come a single book. The variegated visual field presented by the bookcase seems to have played a significant causal role in this simple everyday action. If the earliest explorers of the brain had found little model bookshelves in people's heads, which appeared at such crucial book-selecting moments but disappeared at other times, we would have had no difficulty in supposing that these evanescent, intracranial structures were qualitative states and we would have been quite prepared to accept that they were physical. The strange thing is that scientists are now discovering structures like this in the brain, but failing to convince the sceptics.
For example, Kosslyn and Koenig have a marvellous picture of the neurones of a monkey which is looking at a dart board. (Kosslyn & Koenig, 1992, p.68). The monkey's brain contains a picture of the dartboard, with lines radiating from the centre of a circle in a very literal sense. Kosslyn and Koenig did of course have to go through various steps to make this structure directly visible, but this does not cast doubt on its reality. The best and most economical account of what is going on in the brain may call for us to suppose that certain structures are qualitative states of which the brain is phenomenologically aware. In the rest of the paper I intend to make this scenario more convincing.

3. The Strategy: Qualia are to be Expected

So far, I have not provided a conclusive argument for physicalism. I have tried to show that physicalism is at the least as plausible as the alternatives, and that there is a prima facie case for the identification of qualitative states with physical states. If my case is not yet proved beyond all doubt, still this is the way to go. If the hard problem can be solved physicalistically then the first step is to argue that, at the least, token qualia are token brain states.

The next step is to show that the puzzling features of qualia are quite compatible with physicalism. I then want to go one step further and argue that such features are predictable and only to be expected. Brains are representing devices which perceive the world and think about it with a view to acting in accordance with their goals. Physical devices which can perform this function are only able to do so as a result of certain of their inner perceptual states possessing what I shall call for-the-organism properties. For-the-organism properties are things which a physicalist should anticipate arising, and they should also be expected to possess the peculiarities which disturb us about qualia. In particular we should expect for-the-organism properties to resist explanation by physics, in exactly the way qualia are thought to resist physical explanation.

This outlines my overall strategy in approaching the hard problem. If physicalists can predict that properties with these peculiarities will arise in certain circumstances, then these peculiarities should not be considered a threat to physicalism.

My argument that there are a priori reasons for expecting qualia to exist does not rely on the standard sorts of a priori reasoning. I am not putting forward a conceptual truth or a mathematical proof. It is rather that qualia are what we might call an engineering necessity. God or the angels might be able to interact intelligently with the world without being subject to the laws of physics. In particular, they might not learn of things through perception as we know it. If so, they need not have qualia.

An example of an engineering imperative might be; "Anything which flies by utilising Bernoulli's principle must have some sort of wing." Rockets do not utilise Bernoulli's principle, but birds, helicopters and flying fish do. A wing is an extended structure the upper surface of which is more extensive than the lower surface. When it moves through
air, the air travels faster over the upper surface, causing a thrust upwards in terms of Bernoulli's principle.

Helicopter rotors thus count as wings. The idea that flight requires wings is a posteriori, in the sense that Daedalus would never have thought of constructing wings if he had not seen birds. However, once we have understood the engineering problem, perhaps through seeing a single solution to it, we can predict that anything which solves the problem will have a distinctive feature. If it flies by utilising Bernoulli's principle then it must have some sort of wing. It is in this sense that I am claiming that any intelligent, physical perceiver must have qualia.

4. Perception and Perceivers

When it comes to perception, sight is the sense that is best understood. When we see, light strikes the retina which is approximately two-dimensional. The information presented at the eye in the form of light is transformed into a pattern of nerve impulses, which travel via the optic nerve to the brain's visual centre. What we eventually see is a three-dimensional array of coloured objects. Our experienced visual field is isomorphic with the world outside. Sitting at the dining room table I see the shape of a wooden pepper grinder. Stretching out my hand, I can feel the same shape with my eyes closed. There is evidence that there is something isomorphic to the pepper grinder in my brain. Brain states corresponding to conscious states can contain within themselves in physically describable terms some of the striking features of a qualitative state. I have already mentioned the representation of the dartboard which is in principle visible in the brain of the monkey perceiving it. It is quite easy, with regard to spatial properties, to see how a brainstate can be isomorphic with external things. If such brain states are qualia, then there is no reason why the qualia should not also have properties isomorphic with the external object.

Shape is a primary quality. We believe that the shape of the pepper grinder is a real aspect of it. Two different senses agree on that lathe-turned near-cylinder. For that shape to be extracted from the 2-D pattern of light falling on the retina, various operations have to be performed upon the information travelling from the retina to the visual centre. Edges, for example, are detected and enhanced.<2>

While we believe that we see and feel the real shape of a thing, colours are thought to be less objective. They are secondary qualities and are defined with reference to the observer. The pepper grinder is brown because it looks brown to a normal observer under normal circumstances. The colours that we see can be seen as real qualities of things. They indicate the extent to which the surfaces of things reflect and absorb light. I can see the brown wooden grinder standing on the brown wooden table because the grinder and the table are made from different woods, whose surfaces interact differently with light. In my visual field, the shape of the pepper grinder is defined by the enhanced edge where two different colours brown meet. The colours that I see represent various different
properties of incoming light in virtue of its intensity, hue and saturation. These properties of incoming light are due to the light having been reflected from a particular type of surface. However, the classification of surfaces in terms of colour does not pick out a real world property of surfaces. Many very different things, for example, the sky, blue paint, and beetle's wings, can look blue for different reasons (Hardin, 1988).

Colours are subjective in so far as the classification of things by colour depends on the fact that these things all affect human consciousness in the same way. They are subjective in another way. This is that the qualitative interrelations between colours reflect peculiarities in our perceptual equipment as well as objective differences between things out there.

One way of classifying colours is in terms of colour space. A quality space can be defined in psycho-physics on the basis of judgements of the form "This is more similar to this than that". So orange is more similar to red than to green and the smell of lilies is more similar to the smell of roses than to the smell of garlic. Given a complete mapping of quality space, individual shades of colour or smells can be defined by their position in quality space. The structure of these quality spaces reflects peculiarities of our perceptual equipment. Those who are red-green colour blind have a diminished colour space compared to normal people because they lack certain receptors. Their colour space fails to distinguish red from green or blue from yellow in consequence. So too colours can be arranged in a circle in colour space because "one endpoint of the spectrum stimulates just one type of receptor" (Clark 1993, p.154). Sounds, by contrast, cannot be arranged thus because we do not have specialised receptors which pick up only one particular type of sound frequency.

The qualitative states associated with different sensory systems are qualitatively different and form different quality spaces. Sounds are different from smells and from colours. This differentiation of different quality spaces with respect to the different sensory organs may confer adaptive advantages. For example, we turn our ears not our eyes in the direction of a puzzling sound.

This brief overview gives an account of how evolution has "designed" the human body to solve an engineering problem. The task is to build a perceptual system that will deliver the most useful and informative picture of the world to a creature with the sensory organs and interests of our remote ancestors. Given the constraints of the laws of nature it may well be that the best or the only way of doing the job is to transduce incoming information into another medium where it can be processed into an analogue model of the external world. This model is constructed in an analogue space which is variable along different dimensions to reflect variations in the properties of things outside. The need to differentiate the different organs which "source" the different types of sensory information means that the analogue space is constructed from different quality spaces peculiar to different sense organs.
The argument then is that any machine which can build up a picture of the world as sophisticated as ours and with comparable discriminatory powers will have to have perceptual equipment similar to ours. It will need both edge enhancers and quality spaces.

5. Self-Awareness

The dedicated dualist may grant this but go on to claim that these quality spaces are blind. Of course a machine which can distinguish not only blue and yellow but also subtly different shades of brown will have to have some analogue of colour space! However the different positions in this machine's colour space are not rich qualitative experiences like ours. They are simply different causal potentialities. This part of the machine represents a soldier wearing a red coat. Of course there is no red patch in the machine, there is simply a distinctive pattern of silicone chip spiking which is caused by red objects and only by red objects. It looks red neither to us nor to the machine.

The dualist's sceptical response may be appropriate to a certain class of representational devices. Consider a television broadcasting system. A quiz show is filmed in the studio. The sounds and colours are captured by cameras and recorders and then transformed into radiowaves and broadcast to outlying receivers which transform the waves into electrical currents and then into cathode rays which recreate the original shapes and colours on the screen. At every stage in this process the system contains an isomorphic representation of the original quiz show. However, every stage of the system is blind. Even the final screen display in glorious technicolor is not coloured for the machine. It simply displays the information in a format which is coloured for humans. A device which did no more than build up a complicated model of the world might be no more than a sophisticated mirror or television screen. Qualitative states may need more than this.<4>

There are two important ways in which we differ from a mirror or television receiver. We are able to respond to the world intelligently and we are self-aware. We also have qualia. My contention is that an intelligent and self-aware perceiver must have qualitative states. I have briefly described how a model of the world is assembled within us. Aspects of the model represent the world because they are isomorphic to it. The model is presented in a representational medium constructed from the different quality spaces. The first part of my argument will be that anything that is able to represent the world as well as we can must, as a matter of engineering, have a perceptual system like ours at some level of abstraction. To be able to distinguish subtle shades of brown for example it must, perhaps, have a quality space which is as discriminative as ours. So too an internal model must be modelled in some or other medium of representation. Not just anything can be a representational medium. In particular a representational medium must be differentially responsive to differences in things represented.<5>

Consider the exercise of rotating a mental image. Researchers have shown that if someone is shown a complicated structure like a machine part and asked about how it would look from another perspective then the time taken to respond depends on the angle.
through which the machine part would have to be turned to see it from that point of view. (Kosslyn & Pomerantz, 1981) This lends support to the idea that the time taken to respond depends on the time taken actually rotating an image of the machine part in one's head. Certainly it subjectively feels as if I am rotating a mental image! But how can this rotation be performed if the machine part in the image is not different from the imaged background against which it is rotated? The machine part image must be tagged differently for the image rotator from the way in which the background is tagged. These differences in tagging correspond to qualitative differences in the medium of representation.

When such a representational medium is the means whereby a self-aware being perceives and responds to the world it can itself become an object of awareness for the perceiver. The self-aware perceiver can come to realise that some properties are properties of subjective perceptual states. It is true of course that a perceiving organism is primarily aware of the world. It is normally assumed that the properties the world is perceived as having are properties of the world. Only very sophisticated organisms like philosophers fail to be naive realists! The ability to be aware of the representational medium as a representational medium requires both that there be a representing medium of the sort described and that the organism is self-aware in the sense that it can be aware of its own mental states. It is with this sort of self-awareness that qualia become fully conscious. Qualia are inner labels, codes or filing tags. Tags of this sort must on the one hand be regularly associated with a distinctive class of external things. They must also be such that the organism can distinguish one from the other. When the organism becomes aware of them as tags they finally acquire all the characteristics we associate with qualia.

6. The Predictability of Qualia

In arguing that it is predictable that organisms of a particular degree of sophistication will have qualitative states I am arguing that certain modes of intelligent action are only possible if the actor has inner mental states with For-the-organism labels. Any simulacrum of the agent which does not deliver these internal tagging properties will lack certain abilities. With us the sort of activities which seem to demand qualitative states for their explanation are ones where we are motivated by or cognizant of the sensuous nature of our qualia, as in choosing some music or reporting on our feelings. However the equipment which enables us to choose music and report upon our sensory states probably confers a range of other abilities with a more obvious adaptive value. Evolution does not throw in optional extras for nothing.

It might perhaps be that for-the-organism tags are needed for intelligent reflection on the world. Thinking might, as some suggest, be a matter of manipulating representational models of the world. In so far as such mental modelling involves mental imagery there is some plausibility in supposing that this requires qualitative states, i.e. sensorily derived inner states with for-the-organism properties.
Qualitative states, I hypothesise, serve some important evolutionary function which could not be carried out or could not be performed so well without them. Whatever this may be, I conclude that there is a particular intellectuo-perceptual capacity C such that we have C and the possession of C underpins our ability to report on and take pleasure in our qualitative states. The abilities in C depend crucially on qualitative states so that any physical creature without qualitative states would lack C. Qualia are necessary for C in the sense that any device D engineered in accordance with the laws of physics must have such states if it is to have capacity C. As tags for D qualia have For-D qualities. In D's intelligent and perceptual interactions with the world qualia are picked out by D and used in D's calculations in virtue of their For-D qualities. Their For-D qualities are determined by the fact that the different quality spaces in which they are arranged reflect both different properties in the outside world and peculiarities of D's perceptual equipment. Being For-D means that they have a distinctive causal and functional role within D's brain/mind/central processor in virtue of which D as an intelligent and self-aware being can discriminate and categorise in the first instance properties of external objects and secondarily properties of the representational medium in which those external objects are represented. If they did not generate For-D properties D could not function normally as an intelligent perceiver. A full understanding of the workings of D would reveal how the causal and functional roles of qualitative states generate For-D properties essential for D's cognitive processing. (This would provide the answer to Chalmers' "How" question.)

Such an understanding does not reveal how these For-D properties appear to D. For-D properties are experienced by D and to know how they are experienced one needs to actually have those experiences and be in that qualitative state. No mere propositional knowledge can deliver this experience. A full understanding of the workings of D would reveal that D's perceptual equipment can be expected to deliver For-D properties which are experienced in a distinctive manner by D. At the same time these qualitative states which have these For-D aspects are perfectly normal parts of the natural order. Knowing the way mechanisms like D work it can be predicted in accordance with physicalism that For-D properties will be generated. It can also be predicted both that the experiencing of these properties will be unique to D and that to know what it is like to have them one would have to be D. In so far as this means that For-D properties are in a sense inexplicable it also means that this inexplicability is predictable as a consequence of the way that D is engineered.

What I have presented so far is a programme which should allow the physicalist to sleep easily at night without worrying about qualia. It is not a mathematical demonstration that qualitative states are physical. Given the logical possibility of dualism such a proof is not available. Nor will it silence my opponents. But I hope that it shows the way forward. I have given some preliminary considerations in favour of the truth of the various contentions along the way. When these preliminaries are consolidated physicalism should be secure.

Editors Note
David Brooks died of a heart attack in October 1996. Aged 46, he was Associate Professor of Philosophy at the University of Cape Town. He was the author of *The Unity of the Mind*, published by St. Martin's Press in 1994, and of many articles in the philosophy of mind, metaphysics, applied ethics, aesthetics, and the philosophy of the social sciences. This article was submitted to PSYCHE on his behalf after his death. Undoubtedly Brooks would have revised it further had he lived, but the editors judged that the article deserves to see the light of day.]

**Notes**

<1>The behaviour of headless chickens can be explained physicalistically.

<2>For details see (Marr, 1982).

<3>I am taking these as given.

<4>It would not surprise me if they did not but this would require additional empirical knowledge and argument.

<5>I argue for this more fully in (Brooks, 1992).

<6>As I hint above it may well be that qualia exist far below us on the evolutionary scale where there is a much lesser degree of self-consciousness. I am arguing for the weaker thesis that creatures with our abilities which are not mere table-followers must have qualia.

<7>See e.g. (Craik, 1967) & (Johnson-Laird, 1983).

<8>See (Lewis, 1990).

**References**


