www.imprint-academic.com/rama

V. S. Ramachandran and William Hirstein

Three Laws of Qualia

What Neurology Tells Us about the Biological Functions of Consciousness, Qualia and the Self

Neurological syndromes in which consciousness seems to malfunction, such as temporal lobe epilepsy, visual scotomas, Charles Bonnet syndrome, and synesthesia offer valuable clues about the normal functions of consciousness and 'qualia'. An investigation into these syndromes reveals, we argue, that qualia are different from other brain states in that they possess three functional characteristics, which we state in the form of 'three laws of qualia' based on a loose analogy with Newton's three laws of classical mechanics. First, they are irrevocable: I cannot simply decide to start seeing the sunset as green, or feel pain as if it were an itch; second, qualia do not always produce the same behaviour: given a set of qualia, we can choose from a potentially infinite set of possible behaviours to execute; and third, qualia endure in short-term memory, as opposed to non-conscious brain states involved in the on-line guidance of behaviour in real time. We suggest that qualia have evolved these and other attributes (e.g. they are 'filled in') because of their role in facilitating non-automatic, decision-based action. We also suggest that the apparent epistemic barrier to knowing what qualia another person is experiencing can be overcome simply by using a 'bridge' of neurons; and we offer a hypothesis about the relation between qualia and one's sense of self.

Introduction

Nothing is more chastening to human vanity than the realization that the richness of our mental life — all our thoughts, feelings, emotions, even what we regard as our intimate self — arises exclusively from the activity of little wisps of protoplasm in the brain. The distinction between mind and body, illusion and reality, substance and spirit has been a major preoccupation of both eastern and western thought for millenia (Aristotle, 1961; Descartes, 1986; Fodor, 1975; Dennett, 1978; Searle, 1980). And although these distinctions have generated an endless number of debates among philosophers, little of lasting value seems to have emerged. As Sutherland (1989) has said, 'Consciousness is a subject on which much has been written but little is known.' Our primary goal in this paper is to forge a fresh approach to the problem, by treating it not as a philosophical, logical, or conceptual issue, but rather as an empirical problem. Our focus is on showing the *form* a scientific theory of consciousness might take, something which is independent of the truth of all of the more detailed claims and suggestions we will make. Our essay will consist of two sections. In part one, which

philosophers can profitably skip, we describe some thought experiments to illustrate the problem of qualia, since in our experience, most neuroscientists and even most psychologists dispute the very existence of the problem. In part two, we offer numerous examples from neurology and perceptual psychology that, together with a new theoretical framework we offer, will help eventually solve the problem of consciousness. Our theory should be seen as complementing rather than replacing a host of other recent biological approaches to the problem such as those of Crick and Koch (1992), Pat Churchland (1986), Baars (1988), Edelman (1989), Llinás (Llinás & Paré, 1991), Plum (Plum & Posner, 1980), Bogen (1995a,b), Gazzaniga (1993), Humphrey (1993), Damasio (1994) and Kinsbourne (1995).

Much of our discussion will focus on the notion of qualia. It is our contention, however, that the problem of the self and the problem of qualia are really just two sides of the same coin. In part, our argument is that the self is indeed something that arises from brain activity of a certain kind and in certain brain areas, and that this activity is also closely tied to functions related to qualia. In contrast to the idea that qualia are private, subjective, and unsharable properties belonging exclusively to a private self, we suggest two thought experiments to show that there is no insurmountable barrier to sharing them. We then explore various issues involved in how qualia are generated and managed by neural systems, and by examining pathological and experimental cases that clarify these functions, we propose at the same time to clarify the nature of the self. We conclude that the self, or the thing that leads to the illusion of a unitary, enduring self, is neither a separable subject of consciousness nor a homunculus, but it can be mapped anatomically to limbic and other associated structures which 'drive' frontal executive processes. This view contrasts sharply with the widely held view that consciousness is based on the frontal processes themselves.

Part I: Epistemological Prolegomena

The qualia problem

We will illustrate the problem of giving an account of conscious experience, referred to by philosophers as the problem of *qualia*,[1] with two simple thought experiments.

First, imagine that you are a future superscientist with a complete[2] knowledge of the workings of the brain. Unfortunately however, you are a rod monochromat: you don't have any cone receptors in your eyes to delineate the different colours; you are colour blind. For the sake of argument, however, let's also assume that the central processing mechanisms for colour in your brain are intact, they haven't withered away. This is not an illogical assumption; it's fanciful perhaps, but not illogical.

---

[1] Qualia are the 'raw feels' of conscious experience: the painfulness of pain, the redness of red. Qualia give human conscious experience the particular character that it has. For instance, imagine a red square; that conscious experience has (at least) two qualia: a colour quale, responsible for your sensation of redness, and a shape quale, responsible for the square appearance of the imagined object.

[2] The assumption that anyone could ever have a complete knowledge of the brain is questionable, depending of course on what one means by 'complete'. All we mean by this is that the super- scientist's theory has no obvious explanatory gaps in it, and that it allows him to predict behaviour with an extremely high level of accuracy. This example borrows liberally from Jackson's ingenious 'Mary' scenario (Jackson, 1986).

You, the superscientist, study the brain of X, a normal colour perceiver, as he verbally identifies colours he is shown. You've become very interested in this curious phenomenon people call colour; they look at objects and describe them as red or green or blue, but the objects often all look like shades of grey to you. You point a spectrometer at the surface of one of the objects and it says that light with a wavelength of 600nm is emanating from the object, but you have no idea what *colour* this might correspond to, or indeed what people mean when they say 'colour'. Intrigued, you study the pigments of the eye and so on and eventually you come up with a complete description of the laws of wavelength processing. Your theory allows you to trace out the entire sequence of neural events starting from the receptors all the way into the brain until you monitor the neural activity that generates the word 'red'. Now, once you have completely understood the laws of colour vision (or more strictly, the laws of wavelength processing), and you are able to predict correctly which colour word X will utter when you present him with a certain light stimulus, *you have no reason to doubt the completeness of your account*.

One day you come up with a complete diagram. You show it to X and say, 'This is what's going on in your brain.' To which he replies, 'Sure that's what's going on, but I see red, where is the red in this diagram?' 'What *is* that?' you ask. 'That's part of the actual experience of the colour which it seems I can never convey to you,' he says. This is the alleged epistemological barrier which you confront in trying to understand X's experience. Our thought experiment is also useful in that it allows us to put forward a clear *definition* of qualia: they are that aspect of X's brain state that seems to make your scientific description incomplete *from X's point of view*.

Second, imagine there is a species of electric fish in the Amazon which is very intelligent, in fact as intelligent and sophisticated as us. But it has something we lack: the ability to sense electrical fields, using special organs in its skin. You can study the neurophysiology of this fish and figure out how the electrical organs on the sides of its body transduce electrical current, how this is conveyed to the brain, what part of the brain analyses this information, how it uses this information to dodge predators, find prey, and so on. If the electric fish could talk, however, it would say, 'Fine, but you'll never know what it *feels* like to sense electricity.' These two thought experiments exemplify the problem of qualia. They are vaguely similar to Nagel's 'what is it like to be a bat' problem ('You'll never know what it's like to *be* a bat', Nagel, 1974), except that our examples are better, for the following reason. In the Nagel version, it's the whole bat experience, the qualia produced by the bat's radar system *along with everything else* in its conscious mental life, which Nagel claims we cannot know. But this misses the point. Most people would agree that you couldn't know what it is like to *be* a bat unless you are a bat — after all, the bat's mental life is so completely, utterly different. In our electric fish example, however, we are deliberately introducing a creature which is similar to us in every respect, except that it has *one* type of qualia that we lack. And the point is, even though your description of the fish is complete scientifically, it will always be missing something, namely the actual experience of electrical qualia. This seems to suggest that there is an epistemological barrier between us and the fish. What we have said so far isn't new, except that we have come up with a thought experiment which very clearly states the problem of why qualia are thought to be essentially private. It also makes it clear that the problem of qualia is not necessarily a scientific problem, because your *scientific* description is complete. It's just

that the description is incomplete epistemologically because the experience of electric current is something you never will know.

This is what philosophers have assumed for centuries, that there is a barrier which you simply cannot get across. But is this really true? We think not; it's not as though there is this great vertical divide in nature between mind and matter, substance and spirit. We will argue that this barrier is only apparent,[3] and that it arises due to *language*. In fact, this barrier is the same barrier that emerges when there is *any translation*. The language of nerve impulses (which neurons use to communicate among themselves) is one language; a spoken natural language such as English is a different language. The problem is that X can tell you about his qualia only by using an intermediate, spoken language (when he says, 'Yes but there's still the experience of red which you are missing'), and the experience itself is lost in the translation. You are just looking at a bunch of neurons and how they're firing and how they're responding when X says 'red', but what X is calling the subjective sensation of qualia is supposed to be private forever and ever. We would argue, however, that it's only private *so long as he uses spoken language* as an intermediary. If you, the colour blind superscientist, avoid that and take a cable made of neurons from X's area V4 (Zeki, 1993) and connect it directly to the same area in your brain, then perhaps you'll see colour after all (recall that the higher-level visual processing structures are intact in your brain). The connection has to bypass your eyes, since you don't have the right cone cells, and go straight to the neurons in your brain *without an intermediate translation*. When X says 'red', it doesn't make any sense to you, because 'red' is a translation, and you don't understand colour language, because you never had the relevant physiology and training which would allow you to understand it. But if you skip the translation and use a cable of neurons, so that the nerve impulses themselves go directly to the area, then perhaps you'll say, 'Oh my God, I see what you mean.' The possibility of this demolishes the philosophers' argument (Kripke, 1980; Searle, 1980; 1992) that there is a barrier which is insurmountable. Notice that the same point applies to any instruments I might use to detect activity in your brain — the instrument's output is a sort of translation of the events it is actually detecting.

In principle, then, you *can* experience another creature's qualia, for example even the electric fish's. It's not inconceivable that you could find out what that part of the brain is doing in the fish and that you could somehow graft it onto the relevant parts of your brain with all the associated connections, and that you would then start experiencing the fish's electrical qualia.[4] Now we could get into the philosophical debate over whether you need to be a *fish* to experience it, or whether as a human being you could experience it, but we've already made the distinction between the entire experience of being a fish, and the qualia themselves, which are just part of that experience. Thus qualia are not the private property of a particular self; other selves can experience a creature's qualia.

---

[3] This idea emerged in discussions with F.H.C. Crick. See the acknowledgements at the end of this paper.

[4] The same thought experiment can be performed within a single subject. Anaesthetize the corpus callosum of a human at birth, expose the right brain alone to colours, then at age twenty-one de-anaesthetize the callosum, in order to see if the left brain then begins to experience the right brain's qualia.

What are qualia for?

So far we've talked about the epistemology of qualia and we've suggested that there is no barrier, and that you can in principle experience someone else's qualia by using a bridge of neurons — this problem may simply be a translation problem. We now want to address the question of why qualia evolved. Many others have raised this question before and come up with a wide range of different answers. One could also put on the sceptic's hat and say, 'Since you have already shown that the scientific description is complete without qualia, it is meaningless to ask why it evolved or what its function is. Doing so would entail converting a closed system — the physical universe — into an open one, and that would be a logical fallacy.' We could, however, temporarily set aside scepticism[5] and instead search for a reply to the questions 'Why did qualia emerge in evolution; or, why did some brain events come to have qualia?' Is it a particular *style* of information processing that produces qualia, or is it a particular neural locus, or perhaps only some types of neurons are associated with qualia? Crick (1996; Crick & Koch, 1992) has made the ingenious suggestion that the neural locus of qualia is a set of neurons in the lower layers of the primary sensory areas, because these are the ones that project to the frontal lobes. His approach has galvanized the entire scientific community (cf. Horgan, 1994) and has served as a catalyst for those seeking biological explanations for qualia. Similarly, people have suggested that it's the synchronization of oscillations that leads to conscious awareness (Paré & Llinás, 1995; Purpura & Schiff, 1997). This seems somewhat *ad hoc*, however — why this rather than something else? These approaches are attractive, if only for one

---

[5] Epiphenomenalism cannot be rejected on strictly logical grounds and can be defended on grounds of parsimony; we may not *need* qualia for a complete description of the way the brain works. Since when, however, has Occam's razor been useful for scientific *discovery*? In fact, all of science begins with a bold conjecture of what *might* be true. The discovery of relativity, for example, was not the product of applying Occam's razor to our knowledge of the universe at that time. The discovery came from rejecting Occam's razor and asking what if some deeper generalization were true, which is not required by the available data, but which makes unexpected predictions (which later turn out to be parsimonious after all). It is ironic that most scientific discoveries come not from brandishing (or sharpening) Occam's razor — despite the view to the contrary held by the great majority of scientists and philosophers — but from generating seemingly *ad hoc* and ontologically promiscuous conjectures which are *not* called for by the current data.

For the same reason, we are sympathetic to Penrose's (1994) view that some hitherto undiscovered physical principles may be required for explaining conscious experience. Although his particular theory may turn out to be wrong (see, e.g., Grush and Churchland, 1995), we would argue that his idea should not be rejected on the grounds of parsimony alone. The fact that nothing we know about consciousness demands the postulation of new physical principles is not a sound argument against seeking such principles.

In general then, although philosophical scepticism may be logically justified (just as we cannot prove with complete logical certainty that we are not dreaming, or that your 'red' is not my 'green'), it is misplaced in the scientific realm, where one is concerned most often with what is likely to be true 'beyond reasonable doubt' — rather than with absolute certainty. Unless we set aside such misgivings one is trapped in an intellectual stalemate. In this respect we are in complete agreement with Crick and Koch (1992).

Another famous sceptic's challenge (also known as Molyneux's question) is 'Can a person blind from birth ever experience visual qualia?' Although this is often posed as a conceptual dilemma, we believe that it can be solved empirically by simply delivering localized transcranial magnetic stimulation to visuotopic V1 in blind human volunteers, to see whether it evokes completely novel, yet visuotopically organized visual qualia. (There is a paper by Ramachandran, Cobb & Hirstein, on this topic in preparation.)

reason — that reductionism has been the single most successful strategy in science. Unfortunately however, it is not always easy to know *a priori* what the appropriate *level* of reductionism is for a given scientific problem (Churchland, 1996). Elucidation of the role of the double helix in heredity turned out to be the most important scientific discovery in this century (Medawar, 1969), because Crick and Watson had the foresight and genius to realize that the *molecular* level was the appropriate one. Had they chosen the quantum level, they would have failed! In a similar vein, we wouldn't expect an exhaustive description of the molecular structure of a mousetrap to reveal its function. Nor would a parthenogenetic (asexual) Martian scientist understand how the testicles worked by simply studying their structure, *unless* he knew about sex! And yet this is precisely the strategy adopted by the vast majority of neuroscientists trying to understand the functions of the brain.

Part II: The Biological Functions and Neural Basis of Qualia

In this essay we would like to try something different. We will deliberately begin at a 'higher' level of analysis, and use simple introspection as a strategy for elucidating the biological functions of consciousness. Toward this end we will first present some simple demonstrations of the 'filling in' of the natural blind spot of the eye (Ramachandran, 1992) and argue that this can provide some strong hints about the functions of qualia. Following these demonstrations we will examine a number of neurological syndromes in which qualia seem to *malfunction*, which raises the possibility that far from being a holistic property of the entire brain, qualia are indeed associated with the activity of a small subset of neural structures, as suggested by Crick (1994; 1996). We do not claim to have solved the problem of qualia, but at the very least the examples and thought experiments should provide food for thought.

First, consider the well-known example of the blind spot corresponding to the optic disc — the place where the optic nerve exits the back of the eyeball. To demonstrate the blind spot to yourself, shut your right eye and hold Figure 1 about 10 inches away from your face while looking at the small fixation star on the right. Now move the page toward or away from your eye very slowly, and you will find that there is a critical distance at which the spot on the left completely disappears. Notice, however, that when the spot disappears, it does not leave a gap or a dark hole behind in the visual field. Indeed, the entire field looks homogeneous, and the region corresponding to the blind spot is 'filled in' with the same texture as the background. Sir David Brewster, who discovered filling in, believed it was evidence for a benevolent deity (1832): 'The Divine Artificer has not thus left his works imperfect . . . the spot, in place of being black has always the same colour as the ground.' Curiously, Sir David was not troubled by the question of why the Divine Artificer should have created an imperfect eye to begin with!

Now close your right eye and aim the blindspot of your left eye at the middle of your extended finger. The middle of the finger *should* disappear, and yet the finger looks continuous. In other words, the qualia are such that you do not merely *deduce* intellectually that the finger is continuous — 'after all, my blind spot is there' — you literally *see* the missing piece of your finger. A dramatic demonstration of this phenomenon is the following: if you show someone a donut shape so that the donut is 'around' the blind spot, say a yellow donut, and if the inner diameter of the donut is

Figure 1. The eye's natural blind spot
Close your right eye and fixate the star with your left eye. Slowly move the page back and forth about ten inches from your eye until the dark circle on the left disappears.

Figure 2a. Filling in
Cover your right eye and fixate your left eye on the small white cross. Move the figure back and forth until your blind spot encompasses the centre of the ring on the left. Visual processes fill in the centre of the ring so that it looks like a solid disc.

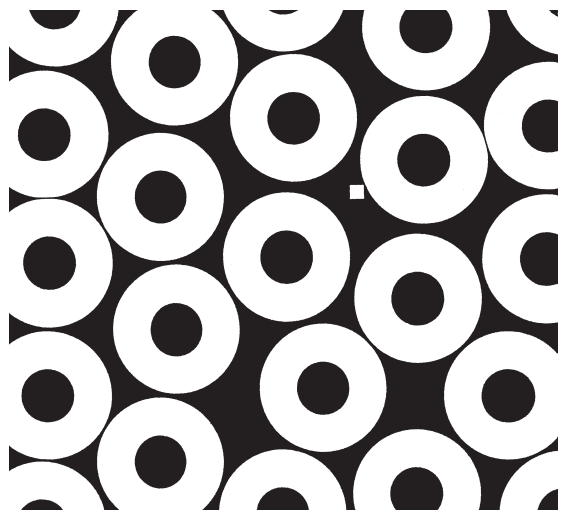

Figure 2b. Salience of filled-in objects
Cover you right eye and fixate your left eye on the small white square. Move the figure back and forth until your blind spot encompasses the centre of the ring on the left of the square. The solid filled-in disc will perceptually 'pop out' from the other rings.

slightly smaller than the blind spot, the donut will look like a complete, homogeneous disk. In fact, the size of the donut can be such that you're actually seeing three times as much yellow now as you did before (see Figure 2a), which in turn means that your brain actually 'filled in' your blind spot with qualia. The reason we emphasize this is that there are some who have argued you simply ignore the blind spot and don't notice what's going on (Dennett, 1991), so that there really is no filling in. But this can't be right, because if you show someone several rings, one of which alone is concentric with the blind spot, that single one will look like a disc and will actually 'pop out' perceptually (see Figure 2b). How can something you are ignoring pop out at you? This means that not only does the blindspot have qualia associated with it, but that the qualia can provide 'sensory support'and therefore are being filled in preattentively, so to speak.

As we have emphasized in previous papers (Ramachandran, 1993; 1995a,b; Churchland and Ramachandran, 1993) we use the phrase 'filling in' in a somewhat metaphorical sense. We certainly do not wish to imply that there is a pixel-by-pixel rendering of the visual image on some internal neural screen, which would defeat the whole purpose of vision (and would imply a 'Cartesian theatre', an idea which Dennett has brilliantly demolished). We disagree, however, with Dennett's specific claim that there is no 'neural machinery' corresponding to the blind spot. (There is, in fact, a patch of cortex corresponding to each eye's blind spot that receives input from the other eye as well as the region surrounding the blind spot in the same eye; Fiorini *et al.*, 1992; see below.) What we mean by 'filling in' is simply this: that one quite literally sees visual stimuli (e.g. patterns and colours) as arising from a region of the visual field where there is actually no visual input. This is a purely descriptive, theory-neutral definition of filling in and one does not have to invoke — or debunk — homunculi watching screens to accept it. We would argue that the visual system fills in not for the benefit of a homunculus but in order to make some aspects of the information explicit for the next level of processing (Ramachandran, 1993). In the last section we will argue that filling in is just one example of a general coherencing of consciousness, which perceptual systems undertake in order to prepare representations to interact with limbic executive structures, an interaction from which both the experience of qualia and intentionality emerge.

Now consider a related example. Suppose I put one finger in front of another finger and look at the two fingers. Of course I see the occluded finger as continuous. I *know* it's continuous. I *sort of* see it as continuous. But if you ask me, do you *literally see* the missing piece of finger, I would say 'no' — for all I know, someone could have actually sliced two pieces of finger and put them on either side of the finger in front to fool me. I don't literally *see* that missing part.

Compare these two cases, the blind spot and the occluded finger, which are in fact quite similar in that they are both cases where there is missing information which the brain supplies. What is the difference, however? What difference does it make to you, the conscious person, that the representation of the yellow donut now has qualia in the middle and that the representation of the occluded finger part does not? The difference, we suggest, is that *you cannot change your mind* about the yellow in the middle of the donut. In other words, you can't think '*Maybe* it's yellow, oh well, maybe it's pink, maybe it's blue. You can't think 'Well, it's *probably* yellow, but who knows, it may be pink.' No, it's shouting at you 'I am yellow', with an explicit representation of

Figure 3
The irrevocability of shape qualia

Once you see the dalmation dog in the picture on the left, it is impossible to go back to the state of not seeing it.

yellowness in its centre. In other words, the filled-in yellow is not revocable, not changeable by you. In the case of the occluded finger, however, you can think 'there's a high probability that there is a finger there, but some malicious scientist could have pasted two half-fingers on either side of it', or, 'there could be a little Martian sitting there for all I know'. These scenarios are highly improbable, but not inconceivable. Another way, then, to capture the difference between the two types of cases is that I could choose to assume that there is something else behind the occluding finger, but that I cannot do that with the filled-in region of the blind spot.

Thus the crucial difference between a qualia-laden percept and one that doesn't have qualia is that the qualia-laden percept is irrevocable, whereas the one which lacks qualia is flexible; you can choose any one of a number of different 'pretend' inputs using top-down imagery. Once a qualia-laden percept has been created, you're stuck with it. A good example of this is that high-contrast photo of the dalmatian dog (Figure 3). Initially, as you look it, it's all fragments, then suddenly everything clicks and you see the dog, you've got the dog qualia. The next time you see it, there's no way you can avoid it, and *not* see the dog. Indeed, we have recently shown that neurons in the brain have permanently altered their connections once you have seen the dog (Tovee *et al.*, 1996).

Three laws of qualia

We now describe three laws of qualia (with apologies to Sir Isaac Newton) which we hope will serve as guideposts for future inquiry. The examples we have just described demonstrate an important feature of qualia: if something is revocable, it isn't a quale (or has only weak qualia associated with it). To put it less strongly, there is a link between the strength or vividness of a quale and the degree of its irrevocability, i.e., this may be quantitative, rather than a qualitative distinction. However, although something's being irrevocable may be necessary, it is certainly not *sufficient* for the presence of qualia. Why? Well, imagine that I shine a light into the eye of someone who is in a coma. If the coma is not too deep, the patient's pupil will constrict, even though she will have no subjective awareness of any qualia caused by the light. The entire reflex arc is irrevocable, and yet there are no qualia associated with it. You

can't change your mind about it, you can't do anything about it, just like you couldn't do anything about the yellow filling in your blind spot in the donut example. So why is it that only the latter has qualia? The key difference, we submit, is that there are no qualia in the case of the pupil's constriction because *there is only one output available*. But in the case of the yellow, even though the representation which was created is irrevocable, *what you can do with the representation is open-ended*; you have the luxury of choice. This is the second important feature of qualia: sensations which are qualia-laden afford the luxury of choice. So now we have identified *two* functional features of qualia: irrevocability on the input side, and flexibility on the output side.

There is a third important feature of qualia. In order to make decisions on the basis of a qualia-laden representation, the representation needs to exist long enough for executive processes to work with it. Your brain needs to hold the representation in an intermediate buffer, in other words, in 'working memory'. Again this condition is not enough in itself, because there could be other reasons why a neural system needs to hold some information in a buffer where qualia are not involved (e.g. spinal cord 'memory'). Typically in these cases, however, there is only one output possible, in which case the second important feature of qualia would be missing, on our scheme. There is some physiological evidence for such a connection between qualia and memory. Goodale has reported a certain type of 'blindsight' patient who can correctly rotate an envelope to post it in a horizontal or a vertical slot, even though he does not consciously perceive the slot's orientation and cannot tell you whether the slot is vertical or horizontal (Milner & Goodale, 1995). But if the room lights are switched off just before he puts the letter in, 'he' forgets the orientation of the slot almost immediately and is unable to get the letter in. This suggests that the unconscious 'dorsal stream' visual system which discerns orientation and affects arm movements accordingly is not only devoid of qualia but also does not have memory; it is the 'ventral stream' visual system that is conscious and has memory. We would maintain that *the reason the qualia-laden ventral system has memory is because it is involved in making choices* based on perceptual representations. In contrast, the system without qualia engages in continuous real-time processing running in a tightly closed loop and consequently doesn't need memory — it is not involved in the making of choices.

This suggests a testable prediction: in patients with blindsight, and in Goodale's visual zombie, if you give the patient a *choice*, the system should go haywire. Not only should it not have short-term memory as Goodale showed, but also it should be incapable of making choices. For example if the person is asked to mail a letter and shown *two* orthogonal slots simultaneously, he should fail, being unable to *choose* between the two (or alternatively, the system might always go for the first one it detects). This is consistent with the Crick-Koch view that the neurons which project to the frontal lobes are the qualia neurons because, obviously, the frontal lobes are important for the *execution* of choices. We would argue, however, that what we think of as the choice itself is really the work of a *limbic executive* system consisting of the amygdala, anterior cingulate cortex, and other areas, and that the frontal lobes are needed only for fully working out the long-range implications and possible alternatives which the decision entails, and for dealing with complications arising as the decision is executed (more on this in the final section).

Let's extend the account to the qualia associated with pain. Say you prick somebody with a pin. It's well known that there are two components: there is an immediate

withdrawal, involving no qualia, followed a couple of seconds later by the experience of pain qualia. This dissociation is itself striking evidence for our view because the non-qualia-laden pathway is irrevocable, but has a fixed output (withdrawal) and therefore doesn't have qualia in our scheme. The pain you experience, on the other hand, is irrevocable, and what you do about it is flexible. You can put some medication on it, or you can run away from whatever caused it. This is a nice example because it's a case of the same stimulus producing two different streams of processing, one involving qualia and the other not.

Bistable percepts

Let's take bistable figures; how would our account apply to them? Here, the sensory stimulus can specify two qualia with equal certainty, so the output system can only choose between those two in creating an intermediate-level representation (Figure 4). Once you settle on an interpretation, however, it clicks and if it's revocable it's only in favour of a *single* other percept. You can only see that famous ambiguous figure as a duck or a rabbit, for instance. But when you finally do see it, the *implications* are infinite — this fulfills our criterion about output flexibility. In the spinal cord on the other hand there are neural circuits that display a type of bistability, but the implications are finite. So for qualia to exist you need potentially infinite implications, but a stable, finite, irrevocable representation as a starting point. But if the starting point is revocable, then the representation will not have strong, vivid qualia. Good examples of this are something seen behind an occluder, or imagining that there is a monkey sitting on that chair. These do not have strong qualia, for good reason, because if they did you wouldn't be able to survive long, given the way your cognitive system is structured. As Shakespeare said: 'You cannot cloy the hungry edge of appetite by bare imagination of a feast.' Very fortunate, for otherwise you wouldn't go eat, you would just generate the qualia associated with satiety in your head. In a similar vein, one could argue that a mutant creature that could imagine having orgasms is unlikely

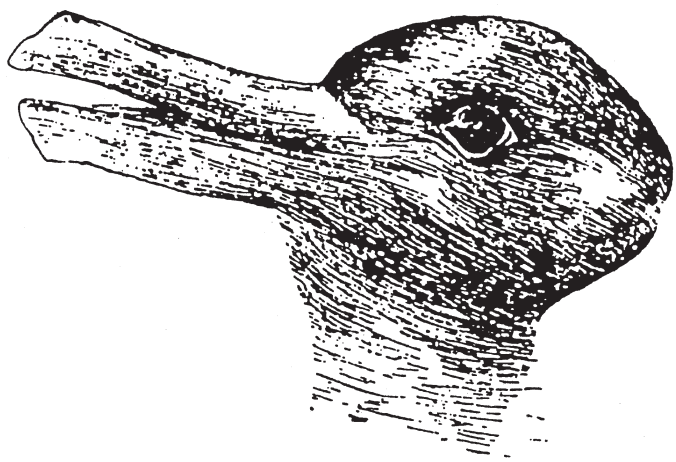

Figure 4
Bistable drawings

'Ambiguous figures' such as this one are designed to allow two possible interpretations. Such figures offer a sort of limited revocability: one set of shape qualia is revocable only in favour of the other.

to pass on its genes to the next generation. *Therefore (real perceptual) qualia are protected; they are partially insulated from top-down influences*.

At the same time, however, you occasionally need to run a virtual reality simulation using less vivid qualia generated from memory representations in order to make appropriate decisions in the absence of the objects which normally provoke those qualia. The memories one normally evokes in this case are not fully laden with qualia; they have qualia which are just vivid enough to allow you to run the simulation. If they possessed full-strength qualia, again, that would be dangerous; indeed that's called a hallucination. Presumably that's what happens in temporal lobe seizures; some mechanism has gone awry, and the virtual reality simulation has now become like real sensory input. The simulation loses its revocability and generates pathological qualia.

Why don't these internally generated images, or beliefs for that matter, have strong qualia? We can explain that. Percepts need to have qualia because they are driving ongoing, decision-laden behaviour. You can't afford the luxury of hesitating over the percept itself, however. The stimulus ensemble determines it, and you don't have time to say, 'Maybe it determines something else.' You need to 'plant a flag' and say 'This is it.' Beliefs and internal images on the other hand should not be qualia-laden, because they should not be confused with real perception; you need to be constantly aware of their tentative nature. And by virtue of their tentative status beliefs lack strong qualia — they are indefinitely revocable. So you believe — and you can imagine — that under the table there is a cat because you see a tail sticking out, but there *could* be a pig under the table with a transplanted cat's tail. You must be willing to entertain that hypothesis, however implausible, because every now and then you might be surprised.

What is the computational advantage to making qualia irrevocable? One answer is stability. If you constantly change your mind about qualia, then the number of potential outputs will literally be infinite; there will be nothing constraining your behaviour. At some point you need to say 'this is it' and plant a flag on it, and *it's that planting of the flag that we call qualia*. The perceptual system follows a rationale something like this: given the available information, it is 90% certain that the object perceived is red. Therefore for the sake of argument, I'll assume that it is red and act accordingly, because if I keep saying 'maybe it's not red', I won't be able to take the next step. In other words, if I treated percepts like beliefs, I would be blind. *Qualia are irrevocable in order to eliminate hesitation and to confer certainty to decisions*.

Charles Bonnet syndrome

This system can break down, however. For example, consider the curious neurological disorder known as Charles Bonnet syndrome. Patients with this disorder typically have damage to the retina, to the optic nerve, optic radiations, or sometimes even to area 17, producing blindness in either a large portion or in the entire visual field. But remarkably, instead of seeing nothing, they experience vivid visual hallucinations. Typically these are 'formed' hallucinations rather than abstract patterns; i.e., the patients claim to see little circus animals, or Lilliputian beings walking around. No adequate explanation of the syndrome has been proposed to date, although the hallucinations are sometimes referred to as 'release hallucinations' in the older clinical literature.

We recently had the opportunity to examine two patients with this syndrome, both of whom present certain novel features, which may help to elucidate the neural mechanisms underlying this disorder. These patients had a sharply circumscribed region in the visual field where they were completely blind; i.e., they had a blind spot, or scotoma. The remarkable thing is that their hallucinations are confined entirely to the blind region. For example, patient MB had a left paracentral scotoma, about the size of her palm (held out at arm's length), caused probably by damage to area 17 and the optic radiations, as a result of laser surgery to destroy an arteriovenous malformation. She was of course completely blind in this region, and yet as often as twenty or thirty times a day she would experience the most vivid hallucinations *confined entirely to the blind spot*. Surprisingly, these were static, outline drawings, like cartoon drawings, filled in with colour, but having no depth or motion.

We suggest that the hallucinations associated with Charles Bonnet syndrome arise because of the massive feedback projections (Ramachandran, 1993) that are known to exist from higher cortical areas to visual areas that precede them in the hierarchy; for example from area 17 to the LGN, or from IT and MT to areas 17 and 18 (Zeki, 1978; van Essen, 1979; Churchland *et al.*, 1994). When a normal person imagines something, such as a rose, we usually assume that some sort of activity is evoked in the higher centres such as the temporal lobes, where the memory of this rose is stored in the form of altered synaptic weights (and perhaps new synaptic connections). So when you imagine a rose, one expects activity in the temporal lobes. But there is a great deal of evidence now to suggest that in addition to the expected activity in IT, there is also activity in area 17, as though somehow this information was being projected back onto your 'neural screen' corresponding to area 17 (Cohen *et al.*, 1996; Farah, 1989). It's as though, to enable you to make certain fine spatial discriminations, your brain needs to run a sort of virtual reality simulation, and for some reason this requires the participation of area 17. (In particular, discrimination of topological features of the image, for example, may require that it be represented again in area 17).

However, when a normal person imagines a rose, she does not literally hallucinate a rose; what she experiences is typically a faint, ghostlike impression of one. Why? One possibility is that the normal person, unlike the Charles Bonnet patient, has real visual input coming in from the retina and optic nerve. This is true, by the way, even when the eyes are closed, because there is always spontaneous activity in the retina, which may function to provide a null signal informing the higher centers that there is no rose here, and this prevents her from literally hallucinating the rose. (Indeed, this may be one reason why spontaneous activity in the peripheral receptors and nerves evolved in the first place.) Again, all this is very fortunate, otherwise your mind would be constantly flooded with internally generated hallucinations, and if you begin confusing internal images with reality, you will be quickly led astray.

In the Charles Bonnet patient the visual input is completely missing, therefore the internally generated images which are sent back to V1, or perhaps V2 (areas 17 and 18), assume a degree of vividness and clarity not seen in normal people. This explains why the images are confined entirely to the scotoma, why they are so extremely vivid (one patient told us that the colours 'look more real than real colours'), and why they have the irrevocable quality of genuine, stimulus-evoked qualia. In other words, ordinarily your top-down imagery will produce only weak images because there is com-

peting real visual input (or spontaneous activity), but when the input goes away, then you start confusing your internal images with external reality.

It is not clear why in the case of MB the images lacked depth and motion. One possibility is that for some reason the feedback information arises only from the ventral stream (the IT-V4 pathway), which is concerned primarily with colour and form, and there was no feedback from the dorsal stream and MT which would have conferred the appropriate spatial attributes, such a depth and motion, to the image.

Perhaps a more important general implication of this syndrome which has been overlooked in the past is that it is strong evidence for the idea that vision is not the one-way cascade or flow of information which it is often thought to be. For example, one simple-minded view of vision (Marr, 1982) holds that visual processing is sequential, modular, and hierarchical: each box computes something and sends it to the subsequent box, a model proposed frequently by AI researchers. This is clearly not how human vision works (Edelman, 1989); instead, there seems to be a constant echo-like back-and-forth reverberation between different sensory areas within the visual hierarchy and indeed (as we shall see) even across modalities. To deliberately overstate the case, it's as though when you look at even the simplest visual scene, you generate an endless number of hallucinations and pick the one hallucination which most accurately matches the current input — i.e., the input seems to *select* from an endless number of hallucinations.[6] There may even be several iterations of this going on, involving the massive back-projections — a sort of constant *questioning*, as in a game of twenty questions, until you eventually home in on the closest approximation to reality (a *partially* constrained hallucination of this sort is, of course, the basis of the well-known Rorschach ink blot test). Thus what you finally see is the result of a compromise between top-down processes and bottom-up processes, a very different view from the conventional one in which vision is seen as involving a hierarchical upward march of information; a bucket brigade.

Synesthesia

A second illustration of breakdown in the functions of qualia is provided by the extraordinary phenomenon of synesthesia, where sensations evoked through one modality produce vivid qualia normally associated with another modality. Many of these cases tend to be a bit dubious — the claims of 'seeing' a sound or 'tasting' a colour turn out to be mere metaphors. However, we recently examined a patient who had relatively normal vision up until the age of seven, then suffered progressive deterioration in his sight due to *retinitis pigmentosa*, until finally at the age of forty he became

---

[6] This is analogous to the way in which the immune system works. When I inject you with killed or denatured smallpox virus (antigens), they generate antibodies and lymphocytes that are *specific* to smallpox. It was once believed by medical scientists (and is still believed by many laypeople and philosophers) that upon entering your blood, the smallpox antigens — a protein molecule — *instruct* the formation of specific antibodies by acting as a template. We know now that this view, while intuitively plausible or obvious, is wrong. In fact, your body has antibody-producing cells for *every conceivable antigen*; even 'martian' antigens, so to speak. What the antigenic challenge (smallpox, for example) does is simply to select the appropriate clone of cells causing them to multiply and produce the specific anti-smallpox antibody. This is a useful analogy, but there is of course a difference: the random gene shuffling that leads to a multiplicity of antibodies has already been accomplished in the fetus, and no longer goes on in the adult. In the case of perception, on the other hand, the random combinations are tried out online even as you watch the stimulus.

completely blind. After about two or three years, he began experiencing visual hallucinations similar to those experienced by Charles Bonnet patients. For example, he would see little spots of red light which initially lacked depth, but which coalesced over time to form the clear visual impression of a face, including depth and shading. More interestingly however, this patient began to notice that whenever he palpated objects while negotiating the visual environment, or held an object in his hand, or even just read braille, this would conjure up the most vivid visual images, sometimes in the form of unformed flashes, sometimes a movement or 'pulsation' of pre-existing hallucinations, or sometimes the actual shape of the object he was palpating (e.g. a corner). These images were highly intrusive, and actually interfered with his braille reading and object palpation. We suggest that in this patient, as indeed in normal people, palpating an object evokes visual memories of that object, as a result of a previously established Hebbian association.[7] Of course, when a normal person closes his eyes and palpates a ruler, he doesn't hallucinate a ruler, even though he will typically visualize it. The reason, again, is because of the presence of normal, countermanding visual input in the form of spontaneous activity from the retina and visual pathways. But when this information is removed, as with the Charles Bonnet patient, our patient begins hallucinating. This can be verified by directly recording evoked potentials from his visual cortex while he is palpating objects (Cobb *et al*., in preparation).

Finally, this line of speculation is also consistent with what we have observed in amputees with phantom limbs. After amputation, many of these patients experience a vivid phantom arm, and while most of them are able to 'move' their phantom, a subset of them find that the phantom is in a fixed position, i.e., their phantom is paralysed. But what would happen if one were to somehow create the visual illusion that the phantom had come back, and could move? To do this, we placed a vertical mirror on the table in front of the patient in the sagittal plane. The patient then puts his normal (say) right hand on the right side of the mirror and 'puts' his phantom left hand on the left side of the mirror. He then looks at the mirror reflection of his right hand, and moves his right hand around until its reflection is exactly superimposed on the felt position of the phantom limb. If he now starts making movements with his right hand, he gets the distinct *visual* illusion that his phantom hand is moving. Remarkably, this also seems to produce vivid sensations seeming to come from joints and muscles in the phantom limb, i.e., the patient experiences a curious form of synesthesia.

Such effects do not occur in normal individuals, supporting our conjecture that the presence of real (somatosensory) input somehow prevents such synesthesia. In a normal person, even though there is a visual impression that their left hand is moving (when they are actually looking at the mirror image of their right hand) this is contradicted by somatic sensations which inform the brain that the left hand is not in fact

---

[7] A second possibility is 'remapping'. We have previously shown that upon amputation of an arm in a human patient the brain area corresponding to the missing hand gets 'invaded' by sensory input from the face. Consequently, touching the face evokes sensations in the missing phantom hand (Ramachandran *et al.*, 1995).

In a similar vein, when the visual areas — either cortical or subcortical — are deprived of input it is not inconceivable that input from the somatosensory area 'invades' the vacated territory so that touching stimuli begins to evoke visual sensations. The two hypotheses, haptically-induced visual imagery vs 'remapping' can be distinguished by measuring the latency of evoked MEG responses (Cobb *et al.*, in preparation).

moving. The fact that this does not happen in the phantom limb patient may imply that the visual signals are causing activation to travel back all the way to the primary somatosensory areas concerned with proprioception. Again, this can be tested using imaging techniques.

Filling in the blind spot

Is there an absolute, qualitative distinction between qualia-laden percepts and those which are not; between perception and conception? Let us illustrate this point with three thought experiments. Consider the obvious phenomenological distinction between the region corresponding to my blind spot, where I can't see anything, and another sort of 'blind spot': the region behind my head, where I also can't see anything. In other words, each of us actually has three blind spots, one in the field of view of each eye, and a third behind our heads, which is much larger. Now, ordinarily you don't walk around experiencing an enormous gap behind your head, and therefore you might be tempted to jump to the conclusion that you are in some sense filling in the gap. But obviously, you don't: there simply is no visual neural representation in the brain corresponding to this area behind your head. You fill it in only in the trite sense that, for example if you are standing in a bathroom with wallpaper in front of you, you assume that the wallpaper continues behind your head. But the important point to emphasize here is, even though you assume (imagine, believe) that there is wallpaper behind your head, you don't literally see it. In other words, any 'filling in' is purely metaphorical and does not fulfill our criterion of being irrevocable. In this fundamental sense there is an important distinction between filling in of the blind spot, and our failure to notice the presence of a big gap behind your head (even though it is the conceptual similarity between these two cases that has misled many psychologists and philosophers to conclude that the eye's blind spot is not filled in). Put very simply, this means that in the case of the blind spot, as we said earlier, you can't change your mind about areas which have been filled in, whereas in the region behind your head, you are free to think, 'In all likelihood there is wallpaper there, but who knows, maybe there is an elephant there.'

It would appear then, that filling in of the blind spot is fundamentally different, both phenomenologically and in terms of what the neurons are doing, from your failure to notice the gap behind your head.[8] But the question remains, is the distinction between what is going on behind your head and the blind spot qualitative or quantitative, and is the dividing line completely arbitrary (cf. 'Is a man bald if he has only three hairs on his head?')? To answer this, let us consider the following thought experiment. Imagine we continue evolving in such a way that our eyes migrate toward the sides of our heads, while at the same time preserving the binocular visual field. The fields of view of the two eyes encroach further and further behind our heads

---

[8] Gattass *et al.* (1992) showed that there is a patch of neurons in area 17 corresponding to the blind spot. The neurons in this patch fire when there are two bars on either side of the blind spot, creating an irrevocable representation in area 17. That is about as close as you can get to arguing that there is a neural mechanism for filling in. To argue otherwise is pedantic.

The converse of qualia-laden filling in would be qualia-less 'repression' or inhibition of irrelevant, confusing, or destabilizing information that would otherwise clutter up consciousness and 'distract' executive structures (Ramachandran, 1995b). Analogously, one might leave non-urgent mail sitting in one's mailbox lest it clutter up one's desktop and distract one from more pressing matters.

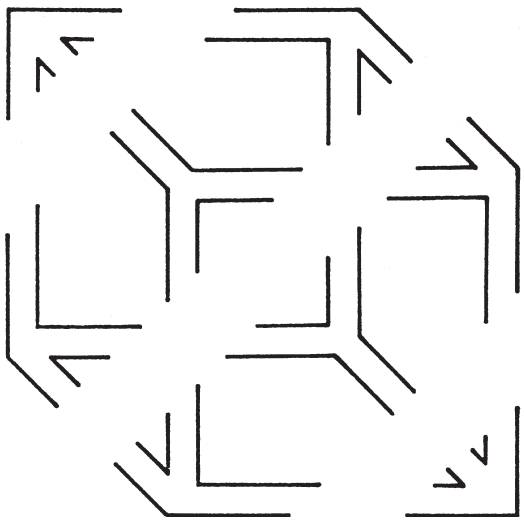

Figure 5
Intermediate case

There is a strong impression that there is a complete cube underneath the three slates, but is this due to genuine filling-in, or to conceptual 'amodal completion'?

(After Kanizsa, 1979, and Bregman, 1981.)

until they are almost touching. At that point let's assume you have a blind spot behind your head (between your eyes) which is identical in size to the blind spot which is in front of you. The question then arises: Would the completion of objects across the blind spot behind your head be true filling in of qualia, as with the real blind spot, or would it still be conceptual, revocable imagery or guesswork of the kind that you and I experience for the region behind the head? The answer to this question, we suggest, is that there will be a definite point when the images become irrevocable, and when representations are created, or at least recreated and fed back to the early visual areas, and at that point it becomes functionally equivalent to the blind spot. If this account is true, there is indeed a fundamental qualitative change, both in the phenomenology and in the corresponding information-processing strategies in the nervous system that are used to create the representations.

Thus, even though blind-spot completion and completion behind the head can be regarded as two ends of a continuum, evolution has seen fit to partition this continuum in order to adequately 'prepare' the completed data for subsequent processing in the case of blind spot completion. We suspect that the motive behind the partition has to do with balancing the need to reduce the workload of higher-level processes by passing them definite, perspicuous, gap-free representations on the one hand, with the need to avoid error on the other. In the case of the eye's blind spot, the chance that something significant is lurking there is small enough that it pays simply to treat the chance as zero. In the case of the blind area behind my head, however, the odds of something important being there are high enough that it would be dangerous to fill in this area with wallpaper or whatever pattern is in front of the eyes.

The second experiment might again be used to undermine the case for a strong qualitative distinction between qualia-laden percepts and conceptual representations, however. Let us go back to the example of the finger occluded by another finger. We argued there that the region behind the occluder is at least partially revocable. However, consider the following intermediate case: a cat behind a picket fence. Or even better, a cube hidden by three slats (Figure 5). It is very hard not to see a cube in this figure. Here you have an intermediate case where the representation seems to be filled

in and yet not filled in. However, the existence of such intermediate cases should not forbid us from arguing that there may be separate neural mechanisms at the two ends of the spectrum.

It is very unlikely that the visual system has evolved dedicated neural machinery for the specific purpose of filling in the blind spot. What we are seeing here, instead, may be a manifestation of a very general visual process — one that we may call surface interpolation (Ramachandran, 1992; 1993; 1995b). It is very likely that the process may have much in common with — and may involve some of the same neural machinery as — the sort of filling in one sees in the example of the occluded finger (which Kanizsa (1979) termed 'amodal completion'). There are nevertheless important differences between completion across the blind spot and amodal completion (e.g. the occluded finger example), which implies that, although the two processes are similar, they are not identical (contrary to the views of Durgin *et al.*, 1995). The most important difference, of course, is that filling in across the blind spot is modal, whereas filling in behind occluders is amodal. What this means is simply that in one case you literally see the filled-in sections, in the other case you don't. (This distinction will not appeal to behaviourists but should be obvious to anyone who has carefully observed such stimuli and is not wholly devoid of common sense.)

A second difference between genuine filling in and conceptual or amodal completion is that the corner of a square or the arc of a circle will get completed amodally behind an occluder but will not get completed modally across the blind spot (Ramachandran, 1992; 1993). In fact, subjects sometimes report the corner or arc being completed amodally behind an 'imaginary' occluder corresponding to the blind spot; the occluder is usually reported to resemble an opaque smudged 'cloud'.

In spite of the differences, it is very likely that the two completion processes share some neural activity *up to a certain stage* in visual processing. Evidence for this comes from the work of Gattass *et al.* (1992). They found that neurons in the patch of area 17 corresponding to (say) the left eye's blind spot respond not only to the right eye (as expected) but also to two collinear line segments lying on either side of the left eye's blind spot — as though they were filling in this segment. Intriguingly, they also noted that similar effects could sometimes be seen in the rest of the normal visual field if a small occluder was used instead of the blind spot. The implication is that, at least in the early stages of processing, both modal completion across the blind spot (i.e. the filling in of qualia) and amodal completion behind occluders may be based on similar neural mechanisms. But if so, why is there such a compelling phenomenological difference between the two? One possibility is that the presence of the occluder itself might be signalled by a different set of neurons which vetoes the modal completion process. This makes good functional sense, for if you were to hallucinate something in front of the occluder you might be tempted to grab it!

Consider a third example: the peculiar mental diplopia or 'multilayer' qualia associated with locating objects in a mirror. Assume you're looking into the rearview mirror of your car, when suddenly you see the reflection of a red car zooming towards you from behind. You accelerate rather than brake, even though, optically, the image is in front of you and expanding. It is as if, when you look at the rearview mirror, you are dealing with bilayered qualia. There is a sense in which you continue to localize the image in front of you, and there is a sense in which you localize it behind you. This raises an interesting question, namely, does the 'location quale' represent the object

as being in front of you or behind you? (Qualia represent an object as being red or square, but they also represent it as being in a certain location, egocentrically specified.)

Now imagine that, instead of a rearview mirror, there is a small window in front of you, and through that window you see a missile being hurled at you. Now of course, you duck backwards. Even though the two situations are exactly equivalent optically — there is an expanding retinal image — in the former case you accelerate forward because somehow, at some level, your visual system performs the appropriate transformation.

Another possibility is that when you look into the mirror you accelerate forward, not because the location qualia are now actually behind you, but because you've learned a reflex avoidance manoeuver, using the dorsal stream system alone. (So, in this situation, the input is irrevocable, and the output is also not open-ended, i.e. it's a single behaviour.) On the other hand, the high-level revocable aspect of the experience — where you think, 'Hey this is a mirror, so the object must be behind me,' doesn't have qualia either — it is more conceptual in nature. But at the critical *intermediate* level, which is still qualia-laden, the object is still represented as being in front of you. You look at the red object, and it's clearly in front of you, and if a fly appears on the mirror, it is right next to the red object. You certainly don't experience it behind your head. So what initially seems to be a disturbing borderline case, in fact can be readily explained in terms of our overall conceptual scheme. But even so, the example is thought-provoking and it leads to experimental questions,[9] such as: If someone were to hurl a missile at you from behind, as you watched in a mirror, would you duck forward or backward?

A fourth example of 'bilayered' (or bistable, really) qualia is shown in Figure 6 (below). What you see initially is a grey rectangle occluded partially by an opaque white square with Swiss-cheese like holes in it. Obviously the grey of the rectangle is not seen where it is occluded, but with a bit of practice you can get yourself to see this as a *transparent* grey film stuck *in front* of the white square with holes. When you see it this way, the film does have qualia because you 'choose' to see it in front and to flag it with the appropriate qualia — preparing it for further processing, as it were.

Anosognosia, schizophrenia, and other delusional states

Notice that in the 'cognitive' realm this sort of completion or filling in is not unlike the confabulations that right hemisphere stroke patients generate to 'deny' that they are paralysed — an anomaly is simply explained away (Ramachandran, 1995c). Some process located in the left hemisphere fills in gaps and smooths over contradictions in the patient's belief system (e.g., the contradiction between 'I can use both arms' and 'I can't see my left arm moving'). We have suggested elsewhere that such psychological defences evolved mainly to stabilize behaviour (they prevent your having to orient to every kind of anomaly that threatened the *status quo*) and should be seen as part of a general strategy for the 'coherencing' of consciousness: they help

---

[9] Intriguingly, we have recently described a new neurological sign of right hemisphere disease, which we call the looking glass syndrome, in which patients, while looking at a mirror, will reach for objects 'inside' the mirror, and assert that the object is inside or behind the mirror, even though they realize they are looking into a mirror (Ramachandran *et al.*, 1997).

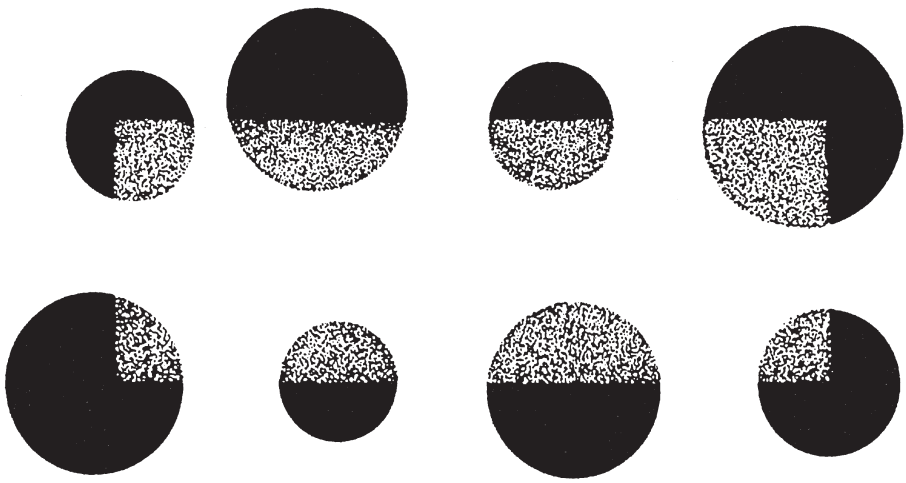

Figure 6. Bilayered qualia
Look at the figure and try to see the grey areas as part of a single translucent rectangle. This represents a higher level of filling in located somewhere between the filling in of the blind spot and amodal filling in. (After Kanizsa, 1979)

avoid indecisive vacillation and serve to optimize resource allocation, and to facilitate rapid, effective action. Similarly, perceptual filling in occurs to keep conscious qualia coherent, perspicuous, and distraction-free — it is another example of a general strategy of coherencing consciousness.

The cognitive styles of the two hemispheres might be fundamentally different; when faced with an 'anomaly' or discrepancy in sensory input, the left hemisphere tries to 'smooth over' the discrepancy (employing denial, repression, or confabulation) in the interest of preserving stability, whereas an 'anomaly detector' in the right hemisphere tends to orient to the discrepancy and generate a paradigm shift in the brain's representation of the situation (Ramachandran, 1995c).

The dialectic between the opposing tendencies of the two hemispheres that we are proposing also bears a tantalizing resemblance to what physicists refer to as the 'edge of chaos' in dynamical systems: the emergence of 'complexity' at the boundary between stability and chaos. Chaos arises in deterministic systems that show a highly sensitive dependence on initial conditions. This is not unlike the sensitivity to perturbation (or 'anomalies') that we have postulated for the cognitive style of the right hemisphere. In marked contrast, the left hemisphere is relatively *insensitive* to change and tries to preserve stability. Interesting or complex types of behaviour, on the other hand, seem to emerge spontaneously at the boundary between the two — a place where there is just enough novelty to keep things interesting and predictable but also just enough stability to avoid complete anarchy and instability. And it is precisely these little eddies of complexity at the border zone that may correspond roughly to what we call human caprice, innovation and creativity.

There is a similarity between anosognosics and schizophrenics who have 'positive' symptoms. In the former, we have argued, there is a failure to register a mismatch between expectation and current sensory input — leading to hallucinations (e.g. 'I can see my arm moving') as well as delusions ('My left arm is fine') and memory dis-

tortions ('I know that I am paralysed now; therefore I never denied that I was paralysed'). We suggest that such a failure results from damage to an anomaly or mismatch detector in the right frontoparietal cortex of anosognosics but it is entirely possible that some similar pathology may also underlie schizophrenia. Indeed, Frith & Dolan (1997) have recently performed an ingenious experiment using the same mirror box we use on our phantom limb patients (Ramachandran & Rogers-Ramachandran, 1996) to demonstrate that a mismatch between vision and proprioception results in right frontal activation (during a PET scan), independently of whether the mismatch occurred on the right or left side of the body!

Surprisingly, there is no neurological syndrome in which one sees exactly the same types of 'positive symptoms' — i.e. combinations of hallucinations and delusions — that occur in schizophrenia (Frith & Dolan, 1997). We would venture to predict, however, that if someone developed Charles Bonnet syndrome (from ocular pathology) *combined with* a right frontoparietal lesion (causing a failure to register a mismatch between fantasy and reality) then you would come pretty close to a neurological equivalent of schizophrenia, for such a patient would take his hallucinations quite literally — not recognizing them to be illusory. Surprisingly, we have recently seen a phantom limb patient who precisely fits this general description. He lost his left arm in a car accident in which he also suffered bilateral frontal damage. While most people who lose an arm experience the illusion of a persisting arm, i.e., a phantom limb, they obviously do not literally see the arm or believe that the arm still exists. Our patient (DS), on the other hand, insisted that his arm was still there and had not been lost, even though he was quite lucid mentally in other domains (Hirstein and Ramachandran, 1997).

Qualia of percepts vs. qualia of beliefs

Beliefs are also associated with 'partial qualia' and conscious awareness, once they are made explicit in 'working memory'. In the absence of sensory support, however, the qualia associated with beliefs are fleeting and less robust than the real qualia-laden percepts associated with sensory stimuli. Therefore the distinction between qualia associated with *percepts* and those associated with explicit (or occurrent) *beliefs* may be quantitative rather than qualitative. Tacit beliefs, on the other hand, are completely qualia-free.

As an analogy, consider the distinction between 'knowing' and 'remembering' or between 'procedural' and 'episodic' memory in humans. We know that episodic memories are partially qualia-laden, whereas skills are not (Tulving, 1983). However, when a bee does a waggle dance, it is communicating an episodic memory. Why does this not qualify as an episodic memory analogous to human episodic memory? To argue that the bee is not conscious, and true episodic memories are conscious, would be circular, and does not answer the question. The problem is readily solved in our scheme, however, since in the bee, the alleged episodic memory is available for the production of only one (or two) outputs, and hence the bee lacks the second of our defining features of consciousness: flexibility of output.[10]

---

[10] One prediction here is that a non-conscious zombie, such as Milner's visual zombie, or perhaps certain sleepwalkers should not have episodic memories.

This is one advantage that our scheme has over other theories of consciousness: it allows us to unambiguously answer such questions as, Is a sleepwalker conscious? Is the spinal cord of a paraplegic conscious? Is a bee conscious? Is an ant conscious when it detects pheromones? In each of these cases, instead of the vague assertion that one is dealing with various degrees of consciousness, which is the standard answer, one should simply apply the three criteria we have specified. For example, Can a sleepwalker make choices? Does he have short-term memory? Does a patient with akinetic mutism have short-term memory? Can the bee use its waggle dance for more than one output?, etc., thereby avoiding endless semantic quibbles over the exact meaning of the word 'consciousness'.

The importance of the temporal lobes for consciousness and qualia

'Does any of this yield clues as to where in the brain might qualia might be?', you ask. It is ironic that people have often thought that the seat of consciousness is the frontal lobes, because nothing dramatic happens to consciousness if you damage the frontal lobes. We suggest instead that most of the action is in the temporal lobes. Admittedly this allows us only a fourfold reduction in the problem space, since the brain has only four lobes; but at the very least it may help us narrow down the problem by allowing us to focus on specific neural structures and their functions. In particular, we suggest, one needs the amygdala and other parts of the temporal lobes for seeing the *significance* of things to the organism. Without this structure you are like Searle's Chinese room (Searle, 1980): capable of giving a single correct output in response to a demand, but with no ability to sense *the meaning* of what you are doing or saying.[11]

Our claim that qualia are based primarily in the temporal lobes is consistent with the idea put forward by Jackendoff (1987) and Crick (1996) that qualia and consciousness are associated not with the *early* stages of perceptual processing (at the level of the retina, for instance), where (in our scheme) obviously multiple choices are not possible. Nor are they associated with the *final* stages of perceptual processing and behaviour planning, where behavioural programs are executed. Rather, they are associated with the *intermediate* stages of processing. The temporal lobes are in fact the interface between perception and action.

Another piece of evidence for the idea that the temporal lobes are the neural locus of consciousness and qualia is that the brain lesions which produce the most profound disturbances in consciousness are those which generate temporal lobe seizures. Researchers who electrically stimulate the temporal lobes of epileptics prior to performing lobectomies have found the temporal lobes to be the best place for producing conscious experiences in their subjects (Penfield & Perot, 1963; Gloor *et al.*, 1982; Gloor, 1992; Bancaud *et al.*, 1976). Stimulating primary sensory areas, such as the visual cortex, can produce strange, unformed qualia, such as phosphene flashes, but only, we suspect because the events set in place by the stimulation eventually follow the natural course of processing into the temporal lobes, and produce (weak) effects there. Stimulating the amygdala is the surest way to 'replay' a full, vivid experience, such as an autobiographical memory complete with intense emotions, or a vivid hal-

---

[11] This reminds one of the old quip in which one behaviourist zombie turns to his mate after passionate lovemaking and says, 'I know it was good for you, but was it good for me?', a question which encapsulates the entire Searle/Dennett debate.

lucination (Gloor, 1992). The seizures which temporal lobe epilepsy (TLE) sufferers endure are associated not only with alterations in consciousness in the sense of personal identity, personal destiny, and personality, but also with vivid, qualia-laden hallucinations such as smells and sounds (MacLean, 1990; Bear, 1979; Waxman & Geschwind, 1975; Gloor, 1992; Bancaud *et al*., 1994). If these are mere memories as they are sometimes claimed to be, why would the person say 'I literally feel like I'm reliving it'? What characterizes these seizures is the *vividness* of the qualia they produce. So the smells, the pains, the tastes and the emotional feelings, all of which are generated in the temporal lobes, suggest that they are in fact the seat of consciousness.

Another reason for choosing the temporal lobes — especially the left temporal lobe — as the main player in generating conscious experience is that this is where much of language — especially semantics — is represented. If I see an apple, it is the activity in the temporal lobes that allows me to *apprehend all its implications* almost simultaneously. Recognition of it as a fruit of a certain type occurs in IT (infero- temporal cortex), the amygdala gauges its significance for my well-being, and Wernicke's and other areas alert me to all the nuances of meaning that the mental image — including the word 'apple' — evokes; I can eat the apple, I can smell it, I can bake a pie, remove its pith, plant its seeds, use it to 'keep the Doctor away', tempt Eve, and on and on.

If one enumerates all of the attributes that we usually associate with the words 'consciousness' or 'awareness', each of them, you will notice, has a correlate in temporal lobe seizures:

(1) *Sensory Qualia — the raw feel of sensations, such as colour or pain*. TLE: Vivid visual and auditory hallucinations; the patient always notices that these look and feel like the real thing — they do not merely have the fleeting qualia of memories (Penfield & Jasper, 1954).

(2) *The attachment of emotional significance and value labels to objects and events*. TLE (especially seizures involving the amygdala): The patient may see cosmic significance in everything around him (Waxman and Geschwind, 1975), or feel intense fear (Strauss *et al.*, 1982). Conversely, bilateral damage to the amygdala may lead to a loss of emotion and empathy, or to the 'psychic blindness' and unthinking, automatic behaviour characteristic of the Kluver-Bucy syndrome (Lilly *et al.*, 1983). It is a moot point whether such a person would have any visual qualia. (One could regard the zombie-like behaviour of Goodale's patient as an extreme example of this.)

(3) *Body image — the sense of being corporeal and of occupying a specific location in space*.TLE: Autoscopic hallucinations (Devinsky *et al.*, 1989), 'out of body' experiences. Also, the temporal lobes and the limbic system receive a more massive projection form the viscera than any other part of the brain. The construction of a body image is one of the foundations of our sense of self but, as we will show in the next section, the body image is a merely a temporary construct, and in the next section we will describe experiments that clearly demonstrate its transitory nature.

(4) *Convictions of truth or falsehood*. TLE: An absolute sense of omnipotence or omniscience (Bear, 1979; Trimble, 1992). It seems ironic that our convictions about the absolute truth or falsity of a thought should depend not so much on the

propositional *language* system but on much more primitive limbic structures which add a form of emotional qualia to thoughts, giving them a 'ring of truth'. This would explain why the more dogmatic assertions of priests as well as scientists are so notoriously resistant to correction through intellectual reasoning!

(5) *Unity — the sense of being a single person despite experiencing a lifetime of diverse sensory impressions*. TLE: Synesthesia; doubling of consciousness; multiplication of personal identity, e.g. in Capgras syndrome (which we have argued is due primarily to a temporal lobe lesion, see Hirstein & Ramachandran, 1997) and other reduplicative paramnesias, the patient may come to regard himself as more than one person. Similarly, multiple personality disorder (MPD) is often seen in association with TLE (Schenk & Bear, 1981; Ahern *et al.*, 1993).

(6) *Free will — the sense of being able to make a decision or control one's movements*. TLE: Even though the ability to engage in long range-planning is lost mainly in frontal disease, it is damage to the cingulate (which is part of the limbic system) that often results in something like 'disorders of *the will*' (e.g. the alien hand syndrome (Goldberg *et al.*, 1981), akinetic mutism: 'loss of will' (Nielson and Jacobs, 1951). Zombie-like automatisms are a frequent concomitant of TLE seizures, and also result from stimulation of the anterior cingulate gyrus (Bancaud *et al.*, 1976). It would be interesting to find out whether the patient can make actual *choices* during such states (we would argue that they cannot).

Furthermore, one frequently sees profound alterations in conscious experience — such as loss of contact with reality (de-realizations) and dream-like trance states during TLE seizures. While each of the disorders listed above can also be seen when other brain areas are damaged (e.g. body image distortions in parietal lobe syndromes), almost all of them can be seen in various combinations when the temporal lobes are damaged. Thus if there is a single brain region that can be regarded as critical for generating conscious experience, it would be the temporal lobes and various interconnected parts of the amygdala, the inferotemporal cortex, Wernicke's area and other associated structures (e.g. the cingulate gyrus). Remove these and you have the prototypical zombie of philosophers' thought experiments.

A new illusion of decapitation

We will now describe an illusion which demonstrates how the body image — despite its apparent durability and permanence — is an entirely transitory internal construct that can be profoundly altered by the stimulus contingencies and correlations that one encounters. Consider the following two illusions, the 'phantom nose' and the 'phantom head' that we recently discovered in our laboratory. In the first experiment, the subject sits in a chair blindfolded, with an accomplice sitting at his right side, or in front of him, facing the same direction. The experimenter then stands near the subject, and with his left hand takes hold of the *subject's* left index finger and uses it to repeatedly and randomly tap and stroke the nose of the accomplice, while at the same time, using his right hand, he taps and strokes the subject's nose in precisely the same manner, and in perfect synchrony. After a few seconds of this procedure, the subject develops the uncanny illusion that his nose has either been dislocated, or has been stretched out several feet forwards or off to the side, demonstrating the striking plas-

ticity or malleability of our body image. The more random and unpredictable the tapping sequence the more striking the illusion. We suggest that the subject's brain regards it as highly improbable that the tapping sequence on his finger and the one on his nose are identical simply by chance and therefore 'assumes' that the nose has been displaced — applying a universal Bayesian logic that is common to all sensory systems. Interestingly, once the illusion is in place, if a drop of ice-cold water is now applied to the subject's nose, the cold is sometimes felt in the new location of the nose. The phantom nose illusion is a very striking one, and we were able to replicate it on twelve out of eighteen naive subjects.[12] Rather surprisingly, the illusion sometimes works even if the accomplice sits *facing* the subject; the logical absurdity of the situation seems not to veto the effect. This simple experiment demonstrates the single most important principle underlying the mechanisms of perception and conscious experience: that they may have evolved exclusively for *extracting statistical* regularities from the natural world.

In the second experiment we had a naive subject looking at his own reflection in a half-silvered mirror, and placed a dummy's head on the other side of the mirror, optically superimposed in exact registration on the subject's own reflection. The lights are switched off and the upper half of the dummy's face, including the nose, is illuminated with one spotlight and the lips alone of the subject are illuminated separately with a different light source. When the subject looks at the mask, he sees a combination of the top of the mask and, reflected in the glass, the bottom of his face. If the subject is asked to make large lip and tongue movements (and baring of the teeth), he develops the uncanny experience of being in direct control of the dummy's facial movements, as though his 'will' was manifesting itself through the dummy's mouth. It is as though the brain regards it highly improbable that the lips of the dummy should be so perfectly synchronized with his own motor speech commands, and therefore assumes that the subject's own free will has taken over the dummy.[13]

To test this objectively, we pinched the dummy's face and found that this evoked a striking increase in the subject's skin conductance response, whereas simply pinching the dummy without the initial lip movements evoked a much smaller response (Ramachandran *et al.*, in preparation). The extraordinary implication is that, using this relatively simple procedure, we had successfully 'decapitated' the subject, inducing the self to temporarily cast off its mortal coil to inhabit the dummy. The subject comes to experience the dummy's head as being his own to such an extent that it is now hooked up to his own limbic system and autonomic output. Even intermittent, unpredictable tactile stimulation (touch, cold, pain) delivered to the subject's face were occasionally referred to the dummy in a modality-specific manner (in a manner analogous to the referral of tactile stimulation to visually resurrected phantom limbs;

---

[12] It has not escaped our notice that if a willing accomplice were available, the effect could also be produced using other body parts.

[13] The sceptic could ask, How is this situation fundamentally different from ordering another human being — such as a valet — to perform an elaborate series of actions, or controlling a marionette on strings with one's fingers? The answer is that in the former case there is no perfect *temporal synchrony* between the orders issued and the actions performed by your subordinate; and in the latter case, even though there is some degree of synchrony, the movement trajectories and the body parts involved in the marionette are different from those of the puppeteer. This explains why the transfer of free will requires an experimental setup similar to the one we describe.

Ramachandran and Rogers-Ramachandran, 1996). The observation also lends credibility to the reports of the self temporarily deserting the body: out of the body experiences and 'autoscopic hallucinations' in parietal lobe syndrome and ketamine anesthesia.

Qualia and 'the self'

We have discussed qualia and the body image,[14] but what about the self? Even though the notion of a unitary, enduring self may turn out to be a form of adaptive self- deception or delusion (Ramachandran, 1995b) we must consider why the illusion arises. We also need to consider the question of who the so-called observer is in the two thought experiments we began with. Since qualia-laden percepts are generated *for* someone or something — presumably 'the self' — the problem of the self and the problem of qualia are really just two sides of the same coin.

One way to approach the question of how our account of qualia relates to the question of the self is to ask from a scientific point of view why something like filling in of the blind spot with qualia-laden representations occurs. The original motive many had for arguing that the blind spot is *not* filled in was that there is no one there to fill them in for — that no homunculus is there looking at them (Dennett, 1991). This is an argument against the following line of reasoning: 'If qualia are filled in, they must be filled in for some *viewer*, i.e., a homunculus.'

There is reason to think that the conclusion is false (i.e., there is no homunculus), it was argued, and hence reason to think that the antecedent is also false: qualia are not in fact filled in, and that the appearance that they are is an illusion (Dennett, 1991). Now, since we have argued that qualia are in fact filled in (Ramachandran, 1992; 1993; 1995a; Ramachandran & Gregory, 1991), does this mean that we believe they are filled in for a homunculus? Of course not, but the fallacy may not be in the *form* of the reasoning, just in the illegitimate specificity with which the conclusion is stated. The above argument is really a 'straw man'; the line of reasoning should run: 'If qualia are filled in, they are filled in for *something*.'

Now, what is the 'something' here? There exists in certain branches of psychology the notion of an executive, or a control process (McKay, 1969). These processes are generally taken to be frontal, or prefrontal, but we would like to suggest that the something which qualia are filled in for is a sort of executive process, but a limbic[15]

[14] Our 'phantom nose' effect is quite similar to one reported by Lackner (1988) except that the underlying principle is different. In Lackner's experiment, the subject sits blindfolded at a table, with his arm flexed at the elbow, holding the tip of his own nose. If the experimenter now applies a vibrator to the tendon of the biceps, the subject not only feels that his arm is extended — because of spurious signals from muscle stretch receptors — but also that his nose has actually lengthened. Lackner invokes Helmholtzian 'unconscious inference' as an explanation for this effect (I am holding my nose; my arm is extended; therefore my nose must be long). The illustration we have described, on the other hand, does not require a vibrator and seems to depend entirely on a Bayesian principle — the sheer statistical improbability of two tactile sequences being identical. (Indeed, our illusion cannot be produced if the subject simply holds the accomplice's nose.) Not all subjects experience this effect, but that it happens at all is astonishing: that a lifetime's evidence concerning your nose can be negated by just a few seconds of intermittent tactile imput.

[15] The limbic system includes the hypothalamic nucleii, amygdala insula, interstitial nuclei of the striae terminalis, fornix and fimbria, septum, mamillary bodies and cingulate gyrus, but the exact definition is not critical to our argument. The cholinergic lateral dorsal tegmental and pedunculopontine nuclei and the intralaminar thalamic nucleii that project to limbic structures may be an integral part of the

one, rather than a frontal one. This would be a process involved in connecting motivation and emotion with the choice of actions to perform, based on a certain definite incoming set of qualia — very much the sort of thing which the self was traditionally supposed to do. A control process is not something which has all the properties of a full human being, of course — it is not at all a homunculus. All the notion of a control process entails, as we are employing it, is that control processes are guided by some brain areas (i.e. perceptual areas and motivational areas) as they control the activities of other brain areas (i.e. motor and planning areas).

Seen this way, filling in is a kind of treating and preparing of qualia in order for them to interact properly with limbic executive structures. Qualia may need to be filled in before they causally interact with these structures because gaps interfere with the proper working of these executive structures. To speak metaphorically, perhaps the control structures are prone to be distracted by gaps in a way which greatly reduces their efficiency and their ability to select appropriate output. The processes involved in generating qualia smooth over anomalies in their product in the same way in which the president's advisers might remove any little confusions or fill in any gaps in the data they give him, inconsequential confusions and gaps which might unnecessarily distract his attention from the main message of the data, causing him to take longer to make a decision, or worse, to make the wrong decision.

Where in the limbic system are these control processes? Perhaps a system involving the amygdala and the anterior cingulate, given the amygdala's central role in emotion (LeDoux, 1992; Halgren, 1992), and the anterior cingulate's apparent executive role (Posner & Raichle, 1994; Devinsky *et al.*, 1995), and the connection between its damage and disorders of the will, such as akinetic mutism and alien hand syndrome. It is not difficult to see how such processes could give rise to the mythology of a self as an active presence in the brain — a 'ghost in the machine'.

Acknowledgments: We thank P.S. Churchland, F.H.C. Crick, R.L. Gregory, D.C. Dennett, M. Kinsbourne and J. Smythies for stimulating discussions, and the NIMH for support. The idea that the Charles Bonnet syndrome might arise from the activity of feedback pathways projecting back to areas 17 and 18 was first suggested by one of us (VSR) in two interviews (Grady, 1993; Nash, 1995). The notion that the epistemic barrier to sensing another person's qualia results entirely from a translation problem emerged from conversations (and correspondence) with F.H.C. Crick in 1984. Our ideas about bees emerged from discussions with M. Hauser.

References

Ahern, G.L., Herring, A.M., Tackenberg, J. *et al.* (1993), 'The association of multiple personality and temporolimbic epilepsy', *Archives of Neurology*, **50**, pp. 1020–5.

Aristotle (1961), *DeAnima* (Oxford: Clarendon Press).

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

Bancaud, J. *et al.* (1976), 'Manifestations comportmentales induites par la stimulation electrique du gyrus cingulaire anterieur chez l'homme', *Revue Neurologique*, **132**, pp. 705–24.

---

qualia-linked circuitry, but it remains to be seen whether they merely play a 'supportive' role for qualia (as indeed the liver and heart do!) or whether they are a part of the actual circuitry that embodies qualia; i.e. are they analogous to the power supply of a VCR or television set, or to the actual magnetic recording head and cathode ray tube? Hyperactivity of this system may contribute to peduncular hallucinosis. Also the doubling of dorsal tegmental and pedeuneulopontine cell numbers which is known to occur in schizophrenia may help explain hallucinations. (This idea emerged in conversation with John Smythies.)

Bancaud, J., Brunet-Bourgin, F., Chauvel, P. & Halgren, E. (1994), 'Anatomical origin of *deja vu* and vivid ''memories'' in human temporal lobe epilepsy', *Brain*, **117**, pp. 71–90.
Bear, D.M. (1979), 'Personality changes associated with neurologic lesions', in *Textbook of Outpatient Psychiatry*, ed. A. Lazare (Baltimore, MD: Williams and Wilkins Co.).
Bogen, J.E. (1995a), 'On the neurophysiology of consciousness: Part I. An overview', *Consciousness and Cognition*, **4**, pp. 52–62.
Bogen, J.E. (1995b), 'On the neurophysiology of consciousness: Part II. Constraining the semantic problem', *Consciousness and Cognition*, **4**, pp. 137–58.
Bregman, A. (1981), 'Asking the ''what for'' question', in *Perceptual Organization*, ed. M. Kubovy & J. Pomerantz (Hillsdale, NJ: Lawrence Erlbaum Associates).
Brewster, D. (1832), *Letters In Natural Magic* (London: John Murray).
Churchland, P.S. (1986), *Neurophilosophy* (Cambridge, MA: The MIT Press).
Churchland, P.S. (1996), 'The hornswoggle problem', *Journal of Consciousness Studies*, **3** (5–6), pp. 402–8.
Churchland, P.S. & Ramachandran, V.S. (1993), 'Filling in: Why Dennett is wrong', in *Dennett and His Critics: Demystifying Mind*, ed. B. Dahlbom (Oxford: Blackwell Scientific Press).
Churchland, P.S., Ramachandran, V.S. & Sejnowski, T.J. (1994), 'A critique of pure vision', in *Large-scale Neuronal Theories of the Brain*, ed. C. Koch & J.L. Davis (Cambridge, MA: The MIT Press).
Cobb S., Ramachandran, V.S. & Hirstein, W. (in preparation), 'Evoked potentials during synesthesia'.
Cohen, M.S., Kosslyn, S.M., Breiter, H.C. *et al.* (1996), 'Changes in cortical activity during mental rotation. A mapping study using functional MRI', *Brain*, **119**, pp. 89–100.
Crick, F. (1994), *The Astonishing Hypothesis: The Scientific Search for the Soul* (New York: Simon and Schuster).
Crick, F. (1996), 'Visual perception: rivalry and consciousness', *Nature*, **379**, pp. 485–6.
Crick, F. & Koch, C. (1992), 'The problem of consciousness', *Scientific American*, **267**, pp. 152–9.
Damasio, A.C. (1994), *Descartes' Error* (New York: Putnam).
Dennett, D.C. (1978), *Brainstorms* (Cambridge, MA: The MIT Press).
Dennett, D.C. (1991), *Consciousness Explained* (Boston, MA: Little, Brown and Co.).
Descartes, R. (1986), *Meditations on First Philosophy*, trans. J. Cottingham (Cambridge: Cambridge University Press).
Devinsky, O., Feldmann, E., Burrowes, K. & Broomfield, E. (1989), 'Autoscopic phenomena with seizures', *Archives of Neurology*, **46**, pp. 1080–8.
Devinsky, O., Morrell, MJ, Vogt, BA. (1995) 'Contribution of anterior cingulate cortex to behavior', *Brain*, **118**, pp. 279–306.
Durgin, F.H., Tripathy, S.P. & Levi, D.M. (1995), 'On the filling in of the visual blind spot: some rules of thumb', *Perception*, **24**, pp. 827–40.
Edelman, G. (1989), *The Remembered Present* (New York: Basic Books).
Farah, M.J. (1989), 'The neural basis of mental imagery', *Trends in Neurosciences*, **10**, pp. 395–9.
Fiorini, M., Rosa, M.G.P., Gattass, R. & Rocha-Miranda, C.E. (1992), 'Dynamic surrounds of receptive fields in primate striate cortex: A physiological basis', *Proceedings of the National Academy of Science* **89**, pp. 8547–51.
Fodor, J.A. (1975), *The Language of Thought* (Cambridge, MA: Harvard University Press).
Frith, C.D. & Dolan, R.J. (1997), 'Abnormal beliefs: Delusions and memory', Paper presented at the May, 1997, *Harvard Conference on Memory and Belief*.
Gattass, R., Fiorini, M., Rosa, M.P.G, Pinon, M.C.F., Sousa, A.P.B., Soares, J.G.M. (1992), 'Visual responses outside the classical receptive field RF in primate striate cortex: a possible correlate of perceptual completion', in *The Visual System from Genesis to Maturity*, ed. R. Lent (Boston, MA: Birkhauser).
Gazzaniga, M.S. (1993), 'Brain mechanisms and conscious experience', *Ciba Foundation Symposium*, **174**, pp. 247–57.
Gloor, P., Olivier, A., Quesney, L.F., Andermann, F., Horowitz, S. (1982), 'The role of the limbic system in experiential phenomena of temporal lobe epilepsy', *Annals of Neurology*, **12**, pp. 129–43.
Gloor, P. (1992), 'Amygdala and temporal lobe epilepsy', in *The Amygdala: Neurobiological Aspects of Emotion, Memory and Mental Dysfunction*, ed J.P. Aggleton (New York: Wiley-Liss).
Goldberg, G., Mayer, N. & Toglis, J.U. (1981), 'Medial frontal cortex and the alien hand sign', *Archives of Neurology*, **38**, pp. 683–6.
Grady, D. (1993), 'The vision thing: Mainly in the brain', *Discover*, June, pp. 57–66.
Grush, R. & Churchland, P.S. (1995), 'Gaps in Penrose's toilings', *Journal of Consciousness Studies*, **2** (1), pp. 10–29.
Halgren, E. (1992), 'Emotional neurophysiology of the amygdala within the context of human cognition', in *The Amygdala: Neurobiological Aspects of Emotion, Memory and Mental Dysfunction*, ed J.P. Aggleton (New York: Wiley-Liss).
Hirstein, W. & Ramachandran, V.S. (1997), 'Capgras syndrome: A novel probe for understanding the neural representation of the identity and familiarity of persons', *Proceedings of the Royal Society of London*, **264**, pp. 437–44.

Horgan, J. (1994), 'Can science explain consciousness?', *Scientific American*, **271**, pp. 88–94.
Humphrey, N. (1993), *A History of the Mind* (London: Vintage).
Jackendoff, R. (1987), *Consciousness and the Computational Mind* (Cambridge, MA: The MIT Press).
Jackson, F. (1986), 'What Mary did not know', *Journal of Philosophy*, **83**, pp. 291–5.
Kanizsa, G. (1979), *Organization In Vision* (New York: Praeger).
Kinsbourne, M. (1995), 'The intralaminar thalamic nucleii', *Consciousness and Cognition*, **4**, pp. 167–71.
Kripke, S.A. (1980), *Naming and Necessity* (Cambridge, MA: Harvard University Press).
Lackner, J.R. (1988), 'Some proprioceptive influences on perceptual representations', *Brain*, **111**, pp. 281–97.
LeDoux, J.E. (1992), 'Emotion and the amygdala', in *The Amygdala: Neurobiological Aspects of Emotion, Memory and Mental Dysfunction*, ed J.P. Aggleton (New York: Wiley-Liss).
Lilly, R., Cummings, J.L., Benson, D.F. & Frankel, M. (1983), 'The human Kluver-Bucy syndrome', *Neurology*, **33**, pp. 1141–5.
Llinás, R.R. & Paré, D. (1991), 'Of dreaming and wakefulness', *Neuroscience*, **44**, pp. 521–35.
MacLean, P.D. (1990), *The Triune Brain in Evolution* (New York: Plenum Press).
MacKay, D.M. (1969), *Information, Mechanism and Meaning* (Cambridge, MA: The MIT Press).
Marr, D. (1982), *Vision* (San Francisco: Freeman).
Medawar, P. (1969), *Induction and Intuition in Scientific Thought* (London: Methuen).
Milner, A.D. & Goodale, M.A. (1995), *The Visual Brain In Action* (Oxford: Oxford University Press).
Nagel, T. (1974), 'What is it like to be a bat?', *Philosophical Review*, **83**, pp. 435–50.
Nash, M. (1995), 'Glimpses of the mind', *Time*, pp. 44–52.
Nielson, J.M. & Jacobs, L.L. (1951), 'Bilateral lesions of the anterior cingulate gyri', *Bulletin of the Los Angeles Neurological Society*, **16**, pp. 231–4.
Paré, D. & Llinás, R. (1995), 'Conscious and preconscious processes as seen from the standpoint of sleep-waking cycle neurophysiology', *Neuropsychologia*, **33**, pp. 1155–68.
Penfield, W.P. & Jasper, H. (1954), *Epilepsy and the Functional Anatomy of the Human Brain* (Boston, MA: Little, Brown & Co.).
Penfield, W.P. & Perot, P. (1963), 'The brain's record of auditory and visual experience: a final summary and discussion', *Brain*, **86**, pp. 595–696.
Penrose, R. (1994), *Shadows of the Mind* (Oxford: Oxford University Press).
Plum, F. & Posner, J.B. (1980), *The Diagnosis of Stupor and Coma* (Philadelphia: F.A. Davis and Co.).
Posner, M.I. & Raichle, M.E. (1994), *Frames of Mind* (New York: Scientific American Library).
Purpura K.P. & Schiff, N.D. (1997), 'The thalamic intralaminar nuclei: a role in visual awareness', *The Neuroscientist*, **3**, pp. 8–15.
Ramachandran, V.S. (1992), 'Blind spots', *Scientific American*, **266**, pp. 85–91.
Ramachandran, V.S. (1993), 'Filling in gaps in logic: Some comments on Dennett', *Consciousness and Cognition*, **2**, pp. 165–8.
Ramachandran, V.S. (1995a), 'Filling in gaps in logic: Reply to Durgin *et al.*', *Perception*, **24**, pp. 41-845.
Ramachandran, V.S. (1995b), 'Perceptual correlates of neural plasticity', in *Early Vision and Beyond*, ed. T.V. Papathomas, C. Chubb, A. Gorea and E. Kowler (Cambridge, MA: The MIT Press).
Ramachandran, V.S. (1995c), 'Anosognosia in parietal lobe syndrome', *Consciousness and Cognition*, **4**, pp. 22–51.
Ramachandran, V.S. & Gregory, R.L. (1991), 'Perceptual filling in of artificially induced scotomas in human vision', *Nature*, **350**, pp. 699–702.
Ramachandran, V.S., Rogers-Ramachandran, D. & Cobb, S. (1995), 'Touching the phantom limb', *Nature*, **377**, pp. 489–90.
Ramachandran, V.S. & Rogers-Ramachandran, D. (1996), 'Synaesthesia in phantom limbs induced with mirrors', *Proceedings of the Royal Society of London*, **263**, pp. 377–86.
Ramachandran, V.S, Altschuler, E.L. & Hillyer, S. (1997), 'Mirror agnosia', *Proceedings of the Royal Society of London*, **264**, pp. 645–7.
Ramachandran, V.S., Hirstein, W. & Stoddard, R. (in preparation), 'The phantom head: An illusion of decapitation'.
Schenk, L. & Bear, D. (1981), 'Multiple personality and related dissociative phenomena in patients with temporal lobe epilepsy', *American Journal of Psychiatry*, **138**, pp. 1311–16.
Searle, John R. (1980), 'Minds, brains, and programs', *Behavioral and Brain Sciences*, **3**, pp. 417–58.
Searle, John R. (1992), *The Rediscovery of the Mind* (Cambridge, MA: The MIT Press).
Strauss, E., Risser, A. & Jones, M.W. (1982), 'Fear responses in patients with epilepsy', *Archives of Neurology*, **39**, pp. 626–30.
Sutherland, N.S. (1989), *The International Dictionary of Psychology* (New York: Continuum).
Tovee, M.J., Rolls, E.T. & Ramachandran, V.S. (1996), 'Rapid visual learning in neurones of the primate temporal visual cortex', *Neuroreport*, **7**, pp. 2757–60.
Tulving, E. (1983), *Elements of Episodic Memory* (Oxford: Clarendon Press).
Trimble, M.R. (1992), 'The Gastaut-Geschwind syndrome', in *The Temporal Lobes and the Limbic System*, ed. M.R. Trimble and T.G. Bolwig (Petersfield: Wrightson Biomedical Publishing Ltd.).

van Essen, D.C. (1979), 'Visual areas of the mammalian cerebral cortex', *Annual Reviews of Neuroscience*, **2**, pp. 227–63.
Waxman, S.G. & Geschwind, N. (1975), 'The interictal behavior syndrome of temporal lobe epilepsy', *Archives of General Psychiatry*, **32**, pp. 1580-6.
Zeki, S.M. (1978), 'Functional specialisation in the visual cortex of the rhesus monkey', *Nature*, **274**, pp. 423–8.
Zeki, S.M. (1993), *A Vision of the Brain* (Oxford: Oxford University Press).