

V.S. RAMACHANDRAN and SANDRA BLAKESLEE

Phantoms in the Brain. Human Nature and the Architecture of the Mind. HARPER PERENNIAL London, New York, Toronto and Sydney

CHAPTER 12: Do Martians See Red?

All of modern philosophy consists of unlocking, exhuming and recanting what has been said before.

V.S. RAMACHANDRAN

Why is thought, being a secretion of the brain, more wonderful than gravity, a property of matter?

CHARLES DARWIN

In the first half of the next century, science will confront its greatest challenge in trying to answer a question that has been steeped in mysticism and metaphysics for millennia: What is the nature of the self? As someone who was born in India and raised in the Hindu tradition, I was taught that the concept of the self—the "I" within me that is aloof from the universe and engages in a lofty inspection of the world around me—is an illusion, a veil called *maya*. The search for enlightenment, I was told, consists of lifting this veil and realizing that you are really "One with the cosmos." Ironically, after extensive training in Western medicine and more than fifteen years of research on neurological patients and visual illusions, I have come to realize that there is much truth to this view—that the notion of a single unified self "inhabiting" the brain may indeed be an illusion. Everything I have learned from the intensive study of both normal people and patients who have sustained damage to various parts of their brains points to an unsettling notion: that you create your own "reality" from mere fragments of information, that what you "see" is a reliable—but not always accurate—representation of what exists in the world, that you are completely unaware of the vast majority of events going on in your brain. Indeed, most of your actions are carried out by a host of unconscious zombies who exist in peaceful harmony along with you (the "person") inside your body! I hope that the stories you have heard so far have helped convince you that the problem of self—far from being a metaphysical riddle—is now ripe for scientific inquiry.

Nevertheless, many people find it disturbing that all the richness of our mental life—all our thoughts, feelings, emotions, even what we regard as our intimate selves—arises entirely from the activity of little wisps of protoplasm in the brain. How is this possible? How could something as deeply mysterious as consciousness emerge from a chunk of meat inside the skull? The problem of mind and matter, substance and spirit, illusion and reality, has been a major preoccupation of both Eastern and Western philosophy for millennia, but very little of lasting value has emerged. As the British psychologist Stuart Sutherland has said, "Consciousness is a fascinating but elusive phenomenon: it is impossible to specify what it is, what it does, or why it evolved. Nothing worth reading has been written on it."

I won't pretend to have solved these mysteries,¹ but I do think there's a new way to study consciousness by treating it not as a philosophical, logical or conceptual issue, but rather as an empirical problem.

Except for a few eccentrics (called panpsychists) who believe everything in the universe is conscious, including things like anthills, thermostats, and Formica tabletops, most people now agree that consciousness arises in brains and not in spleens, livers, pancreases or any other organ. This is already a good start. But I will narrow the scope of inquiry even further and suggest that consciousness arises not from the whole brain but rather from certain specialized brain circuits that carry out a particular style of computation. To illustrate the nature of these circuits and the special computations they perform, I'll draw from the many examples in perceptual psychology and neurology that we have already considered in this book. These examples will show that the circuitry that embodies the vivid subjective quality of consciousness resides mainly in parts of the temporal lobes (such as the amygdala, septum, hypothalamus and insular cortex) and a single projection zone in the frontal lobes—the cingulate gyrus. And the activity of these structures must fulfill three important criteria, which I call (with apologies to Isaac Newton, who described the three basic laws of physics) the "three laws of qualia" ("qualia" simply means the raw feel of sensations such as the subjective quality of "pain" or "red" or "gnocchi with truffles"). My goal in identifying these three laws and the specialized structures embodying them is to stimulate further inquiry into the biological origin of consciousness. The central mystery of the cosmos, as far as I'm concerned, is the following: Why are there always two parallel descriptions of the universe—the first-person account ("I see red") and the third-person account ("He says that he sees red when certain pathways in his brain encounter a wavelength of six hundred nanometers")? How can these two accounts be so utterly different yet complementary? Why isn't there only a third-person account, for according to the objective worldview of the physicist and neuroscientist, that's the only one that really exists? (Scientists who hold this view are called behaviorists.) Indeed, in their scheme of "objective science," the need for a first-person account doesn't even arise—implying that consciousness simply doesn't exist. But we all know perfectly well that can't be right. I'm reminded of the old quip about the behaviorist who, just having made passionate love, looks at his lover and says, "Obviously that was good for you, dear, but was it good for me?" This need to reconcile the first-person and third-person accounts of the universe (the "I" view versus the "he" or "it" view) is the single most important unsolved problem in science. Dissolve this barrier, say the Indian mystics and sages, and you will see that the separation between self and nonself is an illusion—that you are really One with the cosmos.

Philosophers call this conundrum the riddle of *qualia* or subjective sensation. How can the flux of ions and electrical currents in little specks of jelly—the neurons in my brain—generate the whole subjective world of sensations like red, warmth, cold or pain? By what magic is matter transmuted into the invisible fabric of feelings and sensations? This problem is so puzzling that not everyone agrees it is even a problem. I will illustrate this so-called qualia riddle with two simple thought experiments of the kind that philosophers love to make up. Such whimsical pretend experiments are virtually impossible to carry out in real life. My colleague Dr. Francis Crick is deeply suspicious of thought experiments, and I agree with him that they can be very misleading because they often contain hidden question-begging assumptions. But they can be used to clarify logical points, and I will use them here to introduce the problem of qualia in a colorful way.

First, imagine that you are a future superscientist with a complete knowledge of the workings of the human brain. Unfortunately you are also completely color-blind. You don't have any cone receptors (the structures in your retina that allow your eyes to discriminate the different colors), but you do have rods (for seeing black and white), and you also have the correct machinery for processing colors higher up inside your brain. If your eyes could distinguish colors, so could your brain.

Now suppose that you, the superscientist, study my brain. I am a normal color perceiver—I can see that the sky is blue, the grass is green and a banana is yellow—and you want to know what I mean by these color terms. When I look at objects and describe them as turquoise, chartreuse or vermilion, you don't have any idea what I'm talking about. To you, they all look like shades of gray.

But you are intensely curious about the phenomenon, so you point a spectrometer at the surface of a ripe red apple. It indicates that light with a wavelength of six hundred nanometers is emanating from the fruit. But you still have no idea what color this might correspond to because you can't experience it. Intrigued, you study the light-sensitive pigments of my eye and the color pathways in my brain until you eventually come up with a complete description of the laws of wavelength processing. Your theory allows you to trace the entire sequence of color perception, starting from the receptors in my eye and passing all the way into my brain, where you monitor the neural activity that generates the word "red." In short, you completely understand the laws of color vision (or more strictly, the laws of wavelength processing), and you can tell me in advance which word I will use to describe the color of an apple, orange or lemon. As a superscientist, *you have no reason to doubt the completeness of your account.* Satisfied, you approach me with your flow diagram and say, "Ramachandran, this is what's going on in your brain!"

But I must protest. "Sure, that's what's going on. But I also see red. Where is the red in this diagram?" "What is that?" you ask.

"That's part of the actual, ineffable experience of the color, which I can never seem to convey to you because you're totally color-blind."

This example leads to a definition of "qualia": they are aspects of my brain state that seem to make the scientific description incomplete—from my point of view.

As a second example, imagine a species of Amazonian electric fish that is very intelligent, in fact, as intelligent and sophisticated as you or I. But it has something we lack—namely, the ability to sense electrical fields using special organs in its skin. Like the superscientist in the previous example, 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 information is conveyed to the brain, what part of the brain analyzes this information and how the fish uses this information to dodge predators, find prey and so on. If the fish could talk, however, it would say, "Fine, but you'll never know what it feels like to sense electricity."

These examples clearly state the problem of why qualia are thought to be essentially private. They also illustrate why the problem of qualia is not necessarily a scientific problem. Recall that your scientific description is complete. It's just that the your account is incomplete epistemologically because the actual experience of electric fields or redness is something you never will know. For you, it will forever remain a "third-person" account.

For centuries philosophers have assumed that this gap between brain and mind poses a deep epistemological problem—a barrier that simply cannot be crossed. But is this really true? I agree that the barrier hasn't yet been crossed, but does it follow that it can never be crossed? I'd like to argue that there is in fact no such barrier, no great vertical divide in nature between mind and matter, substance and spirit. Indeed, I believe that this barrier is only apparent and that it arises as a result of language. This sort of obstacle emerges when there is any translation from one language to another.²

How does this idea apply to the brain and the study of consciousness? I submit that we are dealing here with two mutually unintelligible languages. One is the language of nerve impulses—the spatial and temporal patterns of neuronal activity that allow us to see red, for example. The second language, the one that allows us to communicate what we are seeing to others, is a natural spoken tongue like English or German or Japanese—rarefied, compressed waves of air traveling between you and the listener. Both are languages in the strict technical sense, that is, they are information-rich messages that are intended to convey meaning, across synapses between different brain parts in one case and across the air between two people in the other.

The problem is that I can tell you, the color-blind superscientist, about my qualia (my experience of seeing red) only by using a spoken language. But the ineffable "experience" itself is lost in the translation. The actual "redness" of red will remain forever unavailable to you.

But what if I were to skip spoken language as a medium of communication and instead hook a cable of neural pathways (taken from tissue culture or from another person) from the color-processing areas in my brain directly into the color-processing regions of your brain (remember that your brain has the machinery to see color even though your eyes cannot discriminate wavelengths because they have no color receptors)? The cable allows the color information to go straight from my brain to neurons in your brain without intermediate translation. This is a far-fetched scenario, but there is nothing logically impossible about it.

Earlier when I said "red," it didn't make any sense to you because the mere use of the word "red" already involves a translation. But if you skip the translation and use a cable, so that the nerve impulses themselves go directly to the color area, then perhaps you'll say, "Oh, my God, I see exactly what you mean. I'm having this wonderful new experience."³

This scenario demolishes the philosophers' argument that there is an insurmountable logical barrier to understanding qualia. In principle, you can experience another creature's qualia, even the electric fish's. If you could find out what the electroceptive part of the fish brain is doing and if you could somehow graft it onto the relevant parts of your brain with all the proper associated connections, then you would start experiencing the fish's electrical qualia. Now, we could get into a philosophical debate over whether you need to be a fish to experience it or whether as a human being you could experience it, but the debate is not relevant to my argument. The logical point I am making here pertains only to the electrical qualia—not to the whole experience of being a fish.

The key idea here is that the qualia problem is not unique to the mind-body problem. It is no different in kind from problems that arise from any translation, and thus there is no need to invoke a great division in nature between the world of qualia and the material world. There is only one world with lots of translation barriers. If you can overcome them, the problems vanish.

This may sound like an esoteric, theoretical debate, but let me give you a more realistic example—an experiment we are actually planning to do. In the seventeenth century the English astronomer William Molyneux posed a challenge (another thought experiment). What would happen, he asked, if a child were raised in complete darkness from birth to age twenty-one and were then suddenly allowed to see a cube? Would he recognize the cube? Indeed, what would happen if the child were suddenly allowed to see ordinary daylight? Would he experience the light, saying, "Aha! I now see what people mean by light!" or would he act utterly bewildered and continue to be blind? (For the sake of argument, the philosopher assumes that the child's visual pathways have not degenerated from the deprivation and that he has an intellectual concept of seeing, just as our superscientist had an intellectual concept of color before we used the cable.)

This turns out to be a thought experiment that can actually be answered empirically. Some unfortunate individuals are born with such serious damage to their eyes that they have never seen the world and are curious about what "seeing" really is: To them it's as puzzling as the fish's electroreception is to you. It's now possible to stimulate small parts of their brains directly with a device called a transcranial magnetic stimulator—an extremely powerful, fluctuating magnet that activates neural tissue with some degree of precision. What if one were to stimulate the visual cortex of such a person with magnetic pulses, thereby bypassing the nonfunctional optics of the eye? I can imagine two possible outcomes. He might say, "Hey, I feel something funny zapping the back of my head," but nothing else. Or he might say, "Oh, my God, this is extraordinary! I now understand what all of you folks are talking about. I am finally experiencing this abstract thing called vision. So this is light, this is color, this is seeing!"

This experiment is logically equivalent to the neuron cable experiment we did on the superscientist because we are bypassing spoken language and directly hitting the blind person's brain. Now you may ask, If he does experience totally novel sensations (what you and I call seeing), how can we be sure that it is in fact true vision? One way would be to look for evidence of topography in his brain. I could stimulate different parts of his visual cortex and ask him to point to various regions of the outside world where he experiences these strange new sensations. This is akin to the way you might see stars "out there" in the world when I hit you on the head with a hammer; you don't experience the stars as being inside your skull. This exercise would provide convincing evidence that he was indeed experiencing for the first time something very close to our experience of seeing, although it might not be as discriminating or sophisticated as normal seeing.⁴

Why did qualia—subjective sensation—emerge in evolution? Why did some brain events come to have qualia? Is there a particular *style* of information processing that produces qualia, or are there some types of neurons exclusively associated with qualia, (The Spanish neurologist Ramón y Cajal calls these neurons the "psychic neurons.") Just as we know that only a tiny part of the cell, namely, the deoxyribonucleic acid (DNA) molecule, is directly involved in heredity and other parts such as proteins are not, could it be that only some neural circuits are involved in qualia and others aren't? Francis Crick and Christof Koch have made the ingenious suggestion that qualia arise from a set of neurons in the lower layers of the primary sensory areas, because these are the ones that project to the

frontal lobes where many so-called higher functions are carried out. Their theory has galvanized the entire scientific community and served as a catalyst for those seeking biological explanations for qualia. Others have suggested that the actual patterns of nerve impulses (spikes) from widely separated brain regions become "synchronized" when you pay attention to something and become aware of it.⁵ In other words, it is the synchronization itself that leads to conscious awareness. There's no direct evidence for this yet, but it's encouraging to see that people are at least trying to explore the question experimentally.

These approaches are attractive for one main reason, namely, the fact that reductionism has been the single most successful strategy in science. As the English biologist Peter Medawar defines it, "Reductionism is the belief that a whole may be represented as a function (in the mathematical sense) of its constituent parts, the functions having to do with the spatial and temporal ordering of the parts and with the precise way in which they interact." Unfortunately, as I stated at the beginning of this book, it's not always easy to know *a priori* what the appropriate level of reductionism is for any given scientific problem. For understanding consciousness and qualia there wouldn't be much point in looking at ion channels that conduct nerve impulses, at the brain stem reflex that mediates sneezing or at the spinal cord reflex arc that controls the bladder, even though these are interesting problems in themselves (at least to some people). They would be no more useful in understanding higher brain functions like qualia than looking at silicon chips in a microscope in an attempt to understand the logic of a computer program. And yet this is precisely the strategy most neuroscientists use in trying to understand the higher functions of the brain. They argue either that the problem doesn't exist or that it will be solved some fine day as we plod along looking at the activity of individual neurons.⁶

Philosophers offer another solution to this dilemma when they say that consciousness and qualia are "epiphenomena." According to this view, consciousness is like the whistling sound that a train makes or the shadow of a horse as it runs: It plays no causal role in the real work done by the brain. After all, you can imagine a "zombie" unconsciously doing everything in exactly the same manner that a conscious being does. A sharp tap on the tendon near your knee joint sets in motion a cascade of neural and chemical events that causes a reflex knee jerk (stretch receptors in the knee connect to nerves in the spinal cord, which in turn send messages to the muscles). Consciousness doesn't enter into this picture; a paraplegic has an excellent knee jerk even though he can't feel the tap. Now imagine a much more complex cascade of events starting with long-wavelength light striking your retina and various relays, leading to your saying "red." Since you can imagine this more complex cascade happening without conscious awareness, doesn't it follow that consciousness is irrelevant to the whole scheme? After all, God (or natural selection) could have created an unconscious being that does and says all the things you do, even though "it" is not conscious.

This argument sounds reasonable but in fact it is based on the fallacy that because you can imagine something to be logically possible, therefore it is actually possible. But consider the same argument applied to a problem in physics. We can all imagine something traveling faster than the speed of light. But as Einstein tells us, this "commonsense" view is wrong. Simply being able to imagine that something is logically possible does not guarantee its possibility in the real world, even in principle. Likewise, even though you can imagine an unconscious zombie doing everything you can do, there

may be some deep natural cause that prevents the existence of such a being! Notice that this argument does not prove that consciousness must have a causal role; it simply proves that you cannot use statements that begin, "After all, I can imagine" to draw conclusions about any natural phenomenon. I would like to try a somewhat different approach to understanding qualia, which I will introduce by asking you to play some games with your eyes. First, recall the discussion in Chapter 5 concerning the so-called blind spot—the place where your optic nerve exits the back of your eyeball. Again, if you close your right eye, fix your gaze on the black spot in Figure 5.2 and slowly move the page toward or away from your eye, you will see that the hatched disk disappears. It has fallen into your natural blind spot. Now close your right eye again, hold up the index finger of your right hand and aim your left eye's blind spot at the middle of your extended finger. The middle of the finger should disappear, just as the hatched disk does, and yet it doesn't; it 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. Psychologists call this phenomenon "filling in," a useful if somewhat misleading phrase that simply means that you see something in a region of space where nothing exists.

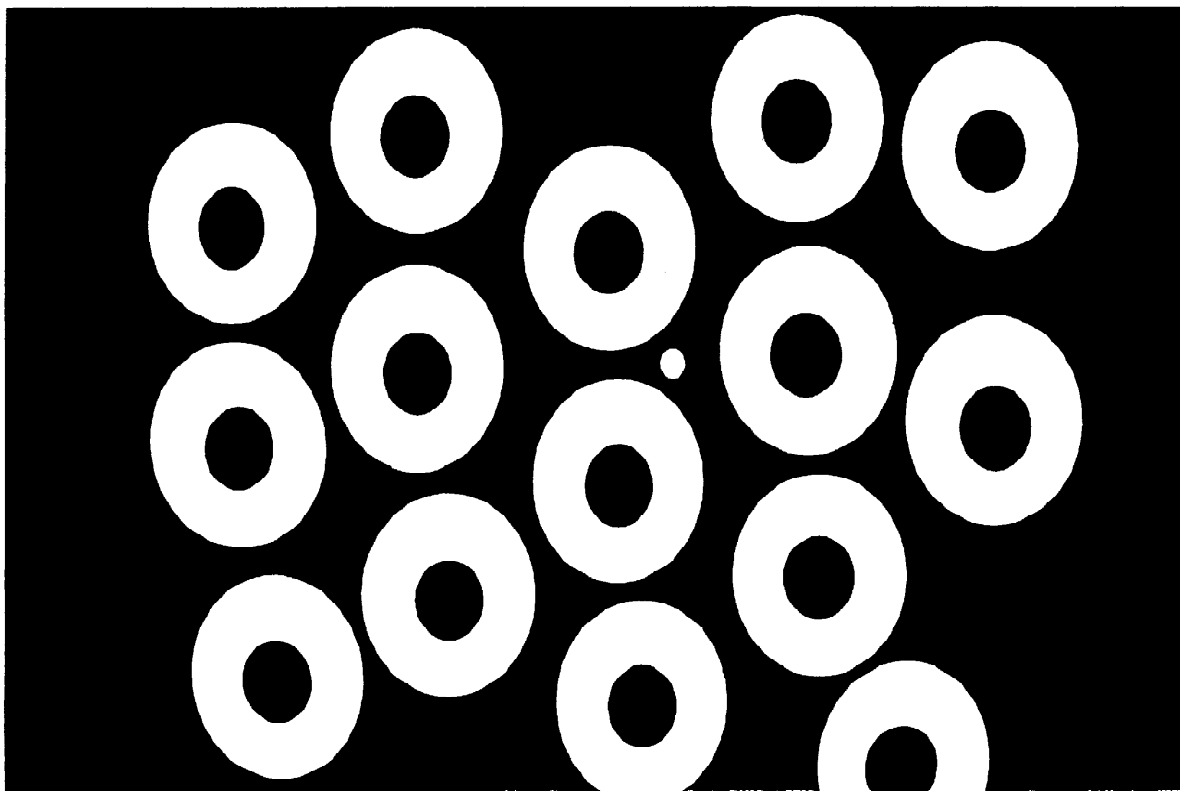


Figure 12.1 [A field of yellow doughnuts (shown in white here). Shut your right eye and look at the small white dot near the middle of the illustration with your left eye. When the page is about six to nine inches from your face, one of the doughnuts will fall exactly around your left eye's blind spot. Since the black hole in the centre of the doughnut is slightly smaller than your blind spot, it should disappear and the blind spot is then "filled in" with yellow (white) qualia from the ring so that you see a yellow disk rather than a ring. Notice that the ring "pops out" conspicuously against the background of rings. Paradoxically, you have made a target more conspicuous by virtue of your

blind spot. If the illusion doesn't work, try using an enlarged photocopy and shifting the white dot horizontally.]

This phenomenon can be demonstrated even more dramatically if you look at Figure 12.1. Again, with your right eye shut look at the small white dot on the right with your left eye and gradually move the book toward you until one of the "doughnuts" falls on your blind spot. Since the inner diameter of the doughnut—the small black disk—is slightly smaller than your blind spot, it should disappear and the white ring should encompass the blind spot. Say the doughnut (the ring) is yellow. What you will see if your vision is normal is a complete yellow homogeneous disk, which will indicate that your brain "filled in" your blind spot with yellow qualia (or white in Figure 12.1). I emphasize this because some people have argued that we all simply ignore the blind spot and don't notice what's going on, meaning that there really is no filling in. But this can't be right. If you show someone several rings, one of which is concentric with the blind spot, that concentric one will look like a homogeneous disk and will actually "pop out" perceptually against a background of rings. How can something you are ignoring pop out at you, This means that the blind spot does have qualia associated with it and, moreover, that the qualia can provide actual "sensory support." In other words, you don't merely deduce that the center of the doughnut is yellow; you literally *see* it as yellow.⁷

Now consider a related example. Suppose I put one finger crosswise in front of another finger (as in a plus sign) and look at the two fingers. Of course, I see the finger in the back as being continuous. I know it's continuous. I sort of see it as continuous. But if you asked me whether I 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 cannot be certain that I really see that missing part.

Compare these two cases, which are similar in that the brain supplies the missing information both times. What's the difference? What does it matter to you, the conscious person, that the yellow doughnut now has qualia in the middle and that the occluded part of your finger does not? The difference is that you cannot change your mind about the yellow in the middle of the doughnut. You can't think, "Maybe it's yellow, but maybe it's pink, or maybe it's blue." No, it's shouting at you, "I am yellow," with an explicit representation of yellowness in its center. 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." This scenario is highly improbable, but not inconceivable.

In other words, I can choose to assume that there might be something else behind the occluding finger, but I cannot do so with the filled-in yellow of the blind spot. Thus the crucial difference between a qualia-laden perception and one that doesn't have qualia is that the qualia-laden perception is irrevocable by higher brain centers and is therefore "tamper-resistant?" whereas the one that lacks qualia is flexible; you can choose any one of a number of different "pretend" inputs using your imagination. Once a qualia-laden perception has been created, you're stuck with it. (A good example of this is the dalmatian dog in Figure 12.2. Initially, as you look, it's all fragments. Then suddenly

everything clicks and you see the dog. Loosely speaking, you've now got the dog qualia. The next time you see it, there's no way you can avoid seeing the dog. Indeed, we have recently shown that neurons in the brain have permanently altered their connections once you have seen the dog.)⁸



Figure 12.2 [Random jumble of splotches. Gaze at this picture for a few seconds (or minutes) and you will eventually see a dalmation dog sniffing the ground mottled with shadows of leaves (hint: the dog's face is on the left towards the middle of the picture; you can see its collar and left ear). Once the dog has been seen it is impossible to get rid of it.

Using similar pictures, we showed recently that neurons in the temporal lobes become altered permanently after the initial brief exposure—once you have “seen” the dog. (Tovee, Rolls, and Ramachandran, 1996). Dalmation dog photographed by Ron James.]

These examples demonstrate an important feature of qualia—it must be irrevocable. But although this feature is necessary, it's not sufficient to explain the presence of qualia. Why? Well, imagine that you are in a coma and I shine a light into your eye. If the coma is not too deep, your pupil will constrict,

even though you 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 as you couldn't do anything about the yellow filling in your blind spot in the doughnut example. So why does only the latter have qualia? The key difference is that in the case of the pupil's constriction, there is only one output—one final outcome—available and hence no qualia. In the case of the yellow disk, even though the representation that was created is irrevocable, you have the luxury of a choice; what you can do with the representation is open-ended. For instance, when you experienced yellow qualia, you could say yellow, or you could think of yellow bananas, yellow teeth, the yellow skin of jaundice and so on. And when you finally saw the dalmatian, your mind would be poised to conjure up any one of an infinite set of dog-related associations—the word "dog," the dog's bark, dog food or even fire engines. And there is apparently no limit to what you can choose. This is the second important feature of qualia: Sensations that 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 you to work with it. Your brain needs to hold the representation in an intermediate buffer or in so-called immediate memory. (For example, you hold the phone number you get from the information operator just long enough to dial it with your fingers.) Again this condition is not enough in itself to generate qualia. A biological system can have other reasons, besides making a choice, for holding information in a buffer. For example, Venus's-flytrap snaps shut only if its trigger hairs inside the trap are stimulated twice in succession, apparently retaining a memory of the first stimulus and comparing it with the second to "infer" that something has moved. (Darwin suggested that this evolved to help the plant avoid inadvertently shutting the trap if hit by a dust particle rather than a bug.) Typically in these sorts of cases, there is only one output possible: Venus's-flytrap invariably closes shut. There's nothing else it can do. The second important feature of qualia—choice—is missing. I think we can safely conclude, contrary to the panpsychists, that the plant does not have qualia linked to bug detection.

In Chapter 4, we saw how qualia and memory are connected in the story of Denise, the young woman living in Italy who suffered carbon monoxide poisoning and developed an unusual kind of "blindsight." Recall that she could correctly rotate an envelope to post it in a horizontal or a vertical slot, even though she could not consciously perceive the slot's orientation. But if someone asked Denise first to look at the slot and then turned off the lights before asking her to post the letter, she could no longer do so. "She" seemed to forget the orientation of the slot almost immediately and was unable to insert the letter. This suggests that the part of Denise's visual system that discerned orientation and controlled her arm movements—what we call the zombie or the how pathway in Chapter 4— not only was devoid of qualia, but also lacked short-term memory. But the part of her visual system—the what pathway—that would normally enable her to recognize the slot and perceive its orientation is not only conscious, it also has memory. (But "she" cannot use the what pathway because it is damaged; all that's available is the unconscious zombie and "it" doesn't have memory.) And I don't think this link between short-term memory and conscious awareness is coincidental.

Why does one part of the visual stream have memory and another not have it? It may be that the qualia-laden what system has memory because it is involved in making choices based on perceptual representations—and choice requires time. The how system without qualia, on the other hand, engages in continuous real-time processing running in a tightly closed loop—like the thermostat in your house. It does not need memory because it is not involved in making real choices. Thus simply posting the letter does not require memory, but choosing which letter to post and deciding where to mail it do require memory.

This idea can be tested in a patient like Denise. If you set up a situation in which she was forced to make a choice, the zombie system (still intact in her) should go haywire. For example, if you asked Denise to mail a letter and you showed her two slots (one vertical, one horizontal) simultaneously, she should fail, for how could the zombie system choose between the two? Indeed, the very idea of an unconscious zombie making choices seems oxymoronic—for doesn't the very existence of free will imply consciousness?

To summarize thus far—for qualia to exist, you need potentially infinite implications (bananas, jaundice, teeth) but a stable, finite, irrevocable representation in your short-term memory as a starting point (yellow). But if the starting point is revocable, then the representation will not have strong, vivid qualia. Good examples of the latter are a cat that you "infer" under the sofa when you only see its tail sticking out, or your ability to imagine that there is a monkey sitting on that chair. These do not have strong qualia, for good reason, because if they did you would confuse them with real objects and wouldn't be able to survive long, given the way your cognitive system is structured. I repeat what Shakespeare said: "You cannot cloy the hungry edge of appetite by bare imagination of a feast." Very fortunate, for otherwise you wouldn't eat; you would just generate the qualia associated with satiety in your head. In a similar vein, any creature that simply imagines having orgasms is unlikely to pass on its genes to the next generation.

Why don't these faint, internally generated images (the cat under the couch, the monkey in the chair) or beliefs, for that matter, have strong qualia? Imagine how confusing the world would be if they did. Actual perceptions need to have vivid, subjective qualia because they are driving decisions and you cannot afford to hesitate. Beliefs and internal images, on the other hand, should not be qualia-laden because they need to be tentative and 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 functional or computational advantage to making qualia irrevocable? One answer is stability. If you constantly changed your mind about qualia, the number of potential outcomes (or "outputs") would be infinite; nothing would constrain your behavior. At some point you need to say "this is it" and plant a flag on it, and it's the planting of the flag that we call qualia. The perceptual system follows a rationale something like this: Given the available information, it is 90 percent certain that what you are seeing is yellow (or dog or pain or whatever). Therefore, for the sake of argument, I'll assume that it is yellow and act accordingly, because if I keep saying, "Maybe it's not yellow," I won't be able to take the next step of choosing an appropriate course of action or thought. In other

words, if I treated perceptions as beliefs, I would be blind (as well as being paralyzed with indecision). Qualia are irrevocable *in order to eliminate hesitation and to confer certainty* to decisions.⁹ And this, in turn, may depend on which particular neurons are firing, how strongly they're firing and what structures they project to.

When I see the cat's tail sticking out from under the table, I "guess" or "know" there is a cat under the table, presumably attached to the tail. But I don't literally see the cat, even though I literally see the tail. And this raises another fascinating question: Are seeing and knowing—the qualitative distinction between perception and conception—completely different, mediated by different types of brain circuitry perhaps, or is there a grey area in between? Let's go back to the region corresponding to the blind spot in my eye, where I can't see anything. As we saw in the Chapter 5 discussion on Charles Bonnet syndrome, there is another kind of blind spot—the enormous region behind my head—where I also can't see anything (although people don't generally use the term "blind spot" for this region). Of course, ordinarily you don't walk around experiencing a huge 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 in the same way that you fill in the blind spot. But you don't. You can't. There is no visual neural representation in the brain corresponding to this area behind your head. You fill it in only in the trivial sense that if you are standing in a bathroom with wallpaper in front of you, you assume that the wallpaper continues behind your head. But even though you assume that there is wallpaper behind your head, you don't literally see it. In other words, this sort of "filling in" is purely metaphorical and does not fulfill our criterion of being irrevocable. In the case of the "real" blind spot, as we saw earlier, you can't change your mind about the area that has been filled in. But regarding 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."

Filling in of the blind spot is therefore fundamentally different from your failure to notice the gap behind your head. But the question remains, Is the distinction between what is going on behind your head and the blind spot qualitative or quantitative? Is the dividing line between "filling in" (of the kind seen in the blind spot) and mere guesswork (for things that might be behind your head) completely arbitrary? To answer this, consider another thought experiment. Imagine we continue evolving in such a way that our eyes migrate toward the sides of our heads, while preserving the binocular visual field. The fields of view of the two eyes encroach farther and farther behind our heads until they are almost touching. At that point let's assume you have a blind spot behind your head (between your eyes) that is identical in size to the blind spot that 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 behind our heads? I think that there will be a definite point when the images become irrevocable, and when robust perceptual representations are created, perhaps even re-created and fed back to the early visual areas. At that point the blind region behind your head becomes functionally equivalent to the normal blind spot in front of you. The brain will then suddenly switch to a completely novel mode of representing the information; it will use neurons in the sensory areas to signal the

events behind your head irrevocably (instead of neurons in the thinking areas to make educated but tentative guesses as to what might be lurking there).

Thus even though blind-spot completion and completion behind your head can be logically regarded as two ends of a continuum, evolution has seen fit to separate them. In the case of your 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 your head, however, the odds of something important being there (like a burglar holding a gun) are high enough that it would be dangerous to fill in this area irrevocably with wallpaper or whatever pattern is in front of your eyes.

So far we have talked about three laws of qualia—three logical criteria for determining whether a system is conscious or not—and we have considered examples from the blind spot and from neurological patients. But you may ask, How general is this principle? Can we apply it to other specific instances when there is a debate or doubt about whether consciousness is involved? Here are some examples:

It's known that bees engage in very elaborate forms of communication including the so-called bee waggle dance. A scout bee, having located a source of pollen, will travel back to the hive and perform an elaborate dance to designate the location of the pollen to the rest of the hive. The question arises, Is the bee conscious when it's doing this?¹⁰ Since the bee's behavior, once set in motion, is irrevocable and since the bee is obviously acting on some short-term memory representation of the pollen's location, at least two of the three criteria for consciousness are met. You might then jump to the conclusion that the bee is conscious when it engages in this elaborate communication ritual. But since the bee lacks the third criterion—flexible output—I would argue that it is a zombie. In other words, even though the information is very elaborate, is irrevocable and held in short-term memory, the bee can only do one thing with that information; only one output is possible—the waggle dance. This argument is important, for it implies that mere complexity or elaborateness of information processing is no guarantee that there is consciousness involved.

One advantage my scheme has over other theories of consciousness is that it allows us unambiguously to answer such questions as, Is a bee conscious when it performs a waggle dance? Is a sleepwalker conscious? Is the spinal cord of a paraplegic conscious—does it have its own sexual qualia—when he (it) has an erection? 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 specified. For example, can a sleepwalker (while he's sleepwalking) take the "Pepsi test"—that is, choose between a Pepsi Cola and a Coca Cola? Does he have short-term memory? If you showed him the Pepsi, put it in a box, switched off the room lights for thirty seconds and then switched them on again, would he reach for the Pepsi (or utterly fail like the zombie in Denise)? Does a partially comatose patient with akinetic mutism (seemingly awake and able to follow you with his eyes but unable to move or talk) have short-term memory? We can now answer these questions and avoid endless semantic quibbles over the exact meaning of the word "consciousness."

Now you might ask, "Does any of this yield clues as to where in the brain qualia might be?" It is surprising that many people think that the seat of consciousness is the frontal lobes, because nothing dramatic happens to qualia and consciousness per se if you damage the frontal lobes— even though the patient's personality can be profoundly altered (and he may have difficulty switching attention). I would suggest instead that most of the action is in the temporal lobes because lesions and hyperactivity in these structures are what most often produce striking disturbances in consciousness. For instance, you need the amygdala and other parts of the temporal lobes for seeing the significance of things, and surely this is a vital part of conscious experience. Without this structure you are a zombie (like the fellow in the famous Chinese room thought experiment proposed by the philosopher John Searle¹¹) capable only 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.

Everyone would agree that qualia and consciousness are not associated with the early stages of perceptual processing as at the level of the retina. Nor are they associated with the final stages of planning motor acts when behavior is actually carried out. They are associated, instead, with the intermediate stages of processing¹²—a stage where stable perceptual representations are created (yellow, dog, monkey) and that have meaning (the infinite implications and possibilities for action from which you can choose the best one). This happens mainly in the temporal lobe and associated limbic structures, and, in this sense, the temporal lobes are the interface between perception and action. The evidence for this comes from neurology; brain lesions that produce the most profound disturbances in consciousness are those that generate temporal lobe seizures, whereas lesions in other parts of the brain only produce minor disturbances in consciousness. When surgeons electrically stimulate the temporal lobes of epileptics, the patients have vivid conscious experiences. Stimulating the amygdala is the surest way to "replay" a full experience, such as an autobiographical memory or a vivid hallucination. Temporal lobe seizures are often associated not only with alterations in consciousness in the sense of personal identity, personal destiny and personality, but also with vivid qualia—hallucinations such as smells and sounds. If these are mere memories, as some claim, why would the person say, "I literally feel like I'm reliving it"? These seizures are characterized by the vividness of the qualia they produce. The smells, pains, tastes and emotional feelings—all generated in the temporal lobes—suggest that this brain region is intimately involved in qualia and conscious awareness.

Another reason for choosing the temporal lobes—especially the left one—is that this is where much of language is represented. If I see an apple, temporal lobe activity allows me to apprehend all its implications almost simultaneously. Recognition of it as a fruit of a certain type occurs in the infero-temporal cortex, the amygdala gauges the apple's 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" and "awareness," each of them, you will notice, has a correlate in temporal lobe seizures, including vivid visual and auditory hallucinations, "out of body" experiences and an absolute sense of omnipotence or omniscience.¹³ Any one of this long list of

disturbances in conscious experience can occur individually when other parts of the brain are damaged (for instance, disturbances of body image and attention in parietal lobe syndrome), but it's only when the temporal lobes are involved that they occur simultaneously or in different combinations; that again suggests that these structures play a central role in human consciousness.

Until now we have discussed what philosophers call the "qualia" problem—the essential privacy and noncommunicability of mental states—and I've tried to transform it from a philosophical problem into a scientific one. But in addition to qualia (the "raw feel" of sensations), we also have to consider the self—the "I" inside you who actually experiences these qualia. Qualia and self are really two sides of the same coin; obviously there is no such thing as free-floating qualia not experienced by anyone and it's hard to imagine a self devoid of all qualia.

But what exactly is the self? Unfortunately, the word "self" is like the word "happiness" or "love"; we all know what it is and know that it's real, but it's very hard to define it or even to pinpoint its characteristics. As with quicksilver, the more you try to grasp it the more it tends to slip away. When you think of the word "self," what pops into your mind? When I think about "myself," it seems to be something that unites all my diverse sensory impressions and memories (unity), claims to be "in charge" of my life, makes choices (has free will) and seems to endure as a single entity in space and time. It also sees itself as embedded in a social context, balancing its checkbook and maybe even planning its own funeral. Actually we can make a list of all the characteristics of the "self"—just as we can for happiness—and then look for brain structures that are involved in each of these aspects. Doing this will someday enable us to develop a clearer understanding of self and consciousness—although I doubt that there will be a single, grand, climactic "solution" to the problem of the self in the way that DNA is the solution to the riddle of heredity.

What are these characteristics that define the self? William Hirstein, a postdoctoral fellow in my lab, and I came up with the following list:

The embodied self: My Self is anchored within a single body. If I close my eyes, I have a vivid sense of different body parts occupying space (some parts more felt than others)—the so-called body image. If you pinch my toe, it is "I" who experiences the pain, not "it." And yet the body image, as we have seen, is extremely malleable, despite all its appearance of stability. With a few seconds of the right type of sensory stimulation, you can make your nose three feet long or project your hand onto a table (Chapter 3)! And we know that circuits in the parietal lobes, and the regions of the frontal lobes to which they project, are very much involved in constructing this image. Partial damage to these structures can cause gross distortions in body image; the patient may say that her left arm belongs to her mother or (as in the case of the patient I saw with Dr. Riita Hari in Helsinki) claim that the left half of her body is still sitting in the chair when she gets up and walks! If these examples don't convince you that your "ownership" of your body is an illusion, then nothing will.

The passionate self: It is difficult to imagine the self without emotions—or what such a state could even mean. If you don't see the meaning or significance of something—if you cannot apprehend all its

implications—in what sense are you really aware of it consciously? Thus your emotions—mediated by the limbic system and amygdala—are an essential aspect of self, not just a "bonus." (It is a moot point whether a purebred Vulcan, like Spock's father in the original Star Trek, is really conscious or whether he is just a zombie—unless he is also tainted by a few human genes as Spock is.) Recall that the "zombie" in the "how" pathway is unconscious, whereas the "what" pathway is conscious, and I suggest that the difference arises because only the latter is linked to the amygdala and other limbic structures (Chapter 5).

The amygdala and the rest of the limbic system (in the temporal lobes) ensures that the cortex—indeed, the entire brain—serves the organism's basic evolutionary goals. The amygdala monitors the highest level of perceptual representations and "has its fingers on the keyboard of the autonomic nervous system"; it determines whether or not to respond emotionally to something and what kinds of emotions are appropriate (fear in response to a snake or rage to your boss and affection to your child). It also receives information from the insular cortex, which in turn is driven partially by sensory input not only from the skin but also from the viscera—heart, lung, liver, stomach—so that one can also speak of a "visceral, vegetative self" or of a "gut reaction" to something. (It is this "gut reaction," of course, that one monitors with the GSR machine, as we showed in Chapter 9, so that you could argue that the visceral self isn't, strictly speaking, part of the conscious self at all. But it can nevertheless profoundly intrude on your conscious self; just think of the last time you felt nauseous and threw up.) Pathologies of the emotional self include temporal lobe epilepsy, Capgras' syndrome and Klüver-Bucy syndrome. In the first, there may be a heightened sense of self that may arise partly through a process that Paul Fedio and D. Bear call "hyperconnectivity"—a strengthening of connections between the sensory areas of the temporal cortex and the amygdala. Such hyperconnectivity may result from repeated seizures that cause a permanent enhancement (kindling) of these pathways, leading the patient to ascribe deep significance to everything around him (including himself!). Conversely, people with Capgras' syndrome have reduced emotional response to certain categories of objects (faces) and people with Klüver-Bucy or Cotard's syndrome have more pervasive problems with emotions (Chapter 8). A Cotard's patient feels so emotionally remote from the world and from himself that he will actually make the absurd claim that he is dead or that he can smell his flesh rotting.

Interestingly, what we call "personality"—a vital aspect of your self that endures for life and is notoriously impervious to "correction" by other people or even by common sense—probably also involves the very same limbic structures and their connections with the ventromedial frontal lobes. Damage to the frontal lobes produces no obvious, immediate disturbance in consciousness, but it can profoundly alter your personality. When a crowbar pierced the frontal lobes of a railway worker named Phineas Gage, his close friends and relatives remarked, "Gage wasn't Gage anymore." In this famous example of frontal lobe damage, Gage was transformed from a stable, polite, hardworking young man into a lying, cheating vagabond who could not hold down a job.¹⁴

Temporal lobe epilepsy patients like Paul in Chapter 9 also show striking personality changes, so much so that some neurologists speak of a "temporal lobe epilepsy personality." Some of them (the patients, not the neurologists) tend to be pedantic, argumentative, egocentric and garrulous. They also tend to be obsessed with "abstract thoughts." If these traits are a result of hyperfunctioning of certain

parts of the temporal lobe, what exactly is the normal function of these areas? If the limbic system is concerned mainly with emotions, why would seizures in these areas cause a tendency to generate abstract thought? Are there areas in our brains whose role is to produce and manipulate abstract thoughts? This is one of the many unsolved problems of temporal lobe epilepsy.¹⁵

The *executive self*: Classical physics and modern neuroscience tell us that you (including your mind and brain) inhabit a deterministic billiard ball universe. But you don't ordinarily experience yourself as a puppet on a string; you feel that you are in charge. Yet paradoxically, it is always obvious to you that there are some things you can do and others you cannot given the constraints of your body and of the external world. (You know you can't lift a truck; you know you can't give your boss a black eye, even if you'd like to.) Somewhere in your brain there are representations of all these possibilities, and the systems that plan commands (the cingulate and supplementary motor areas in the frontal lobes) need to be aware of this distinction between things they can and cannot command you to do. Indeed, a "self" that sees itself as completely passive, as a helpless spectator, is no self at all, and a self that is hopelessly driven to action by its impulses and urgings is equally effete. A self needs free will—what Deepak Chopra calls "the universal field of infinite possibilities"—even to exist. More technically, conscious awareness has been described as a "conditional readiness to act."

To achieve all this, I need to have in my brain not only a representation of the world and various objects in it but also a representation of myself, including my own body within that representation—and it is this peculiar recursive aspect of the self that makes it so puzzling. In addition, the representation of the external object has to interact with my self representation (including the motor command systems) in order to allow me to make a choice. (He's your boss; don't sock him. It's a cookie; it's within your reach to grab it.) Derangements in this mechanism can lead to syndromes like anosognosia or somatoparaphrenia (Chapter 7) in which a patient will with a perfectly straight face claim that her left arm belongs to her brother or to the physician.

What neural structure is involved in representing these "embodied" and "executive" aspects of the self? Damage to the anterior cingulate gyrus results in a bizarre condition called "akinetic mutism"—the patient simply lies in bed unwilling to do or incapable of doing anything even though he appears to be fully aware of his surroundings. If there's such a thing as absence of free will, this is it.

Sometimes when there is partial damage to the anterior cingulate, the very opposite happens: The patient's hand is uncoupled from her conscious thoughts and intentions and attempts to grab things or even perform relatively complex actions without her permission. For example, Dr. Peter Halligan and I saw a patient at Rivermead Hospital in Oxford whose left hand would seize the banister as she walked down the steps and she would have to use her other hand forcibly to unclench the fingers one by one, so she could continue walking. Is the alien left hand controlled by an unconscious zombie, or is it controlled by parts of her brain that have qualia and consciousness? We can now answer this by applying our three criteria. Does the system in her brain that moves her arm create an irrevocable representation? Does it have short-term memory? Can it make a choice?

Both the executive self and the embodied self are deployed while you are playing chess and assume you're the queen as you plan "her" next move. When you do this, you can almost feel momentarily

that you are inhabiting the queen. Now one could argue that you're just using a figure of speech here, that you're not literally assimilating the chess piece into your body image. But can you really be all that sure that the loyalty of your mind to your *own* body is not equally a "figure of speech"? What would happen to your GSR if I suddenly punched the queen? Would it shoot up as though I were punching your own body? If so, what is the justification for a hard-and-fast distinction between her body and yours? Could it be that your tendency normally to identify with your "own" body rather than with the chess piece is also a matter of convention, albeit an enduring one? Might such a mechanism also underlie the empathy and love you feel for a close friend, a spouse or a child who is literally made from your own body?

The mnemonic self: Your sense of personal identity—as a single person who endures through space and time—depends on a long string of highly personal recollections: your autobiography. Organizing these memories into a coherent story is obviously vital to the construction of self.

We know that the hippocampus is required for acquiring and consolidating new memory traces. If you lost your hippocampi ten years ago, then you will not have any memories of events that occurred after that date. You are still fully conscious, of course, because you have all the memories prior to that loss, but in a very real sense your existence was frozen at that time.

Profound derangement to the mnemonic self can lead to multiple personality disorder or MPD. This disorder is best regarded as a malfunction of the same coherencing principle I alluded to in the discussion of denial in Chapter 7. As we saw, if you have two sets of mutually incompatible beliefs and memories about yourself, the only way to prevent anarchy and endless strife may be to create two personalities within one body—the so-called multiple personality disorder. Given the obvious relevance of this syndrome to understanding the nature of self, it is astonishing how little attention it has received from mainstream neurology.

Even the mysterious trait called hypergraphia—the tendency of temporal lobe epilepsy patients to maintain elaborate diaries—may be an exaggeration of the same general tendency: the need to create and sustain a coherent worldview or autobiography. Perhaps kindling in the amygdala causes every external event and internal belief to acquire deep significance for the patient, so there is an enormous proliferation of spuriously self-relevant beliefs and memories in his brain. Add to this the compelling need we all have from time to time to take stock of our lives, see where we stand; to review the significant episodes of our lives periodically—and you have hypergraphia, an exaggeration of this natural tendency. We all have random thoughts during our day-to-day musings, but if these were sometimes accompanied by mini-seizures—producing euphoria—then the musings themselves might evolve into obsessions and entrenched beliefs that the patient would keep returning to whether in his speech or in his writing. Could similar phenomena provide a neural basis for zealotry and fanaticism?

The unified self—imposing coherence on consciousness, filling in and confabulation: Another important attribute of self is its unity—the internal coherence of its different attributes. One way to approach the question of how our account of qualia relates to the question of the self is to ask why something like filling in of the blind spot with qualia occurs. The original motive many philosophers

had for arguing that the blind spot is not filled in was that there is no person in the brain to fill it in for—that no little homunculus is watching.

Since there's no little man, they argued, the antecedent is also false: Qualia are not filled in, and thinking so is a logical fallacy. Since I argue that qualia are in fact filled in, does this mean that I believe they are filled in for a homunculus? Of course not. The philosopher's argument is really a straw man. The line of reasoning should run, If qualia are filled in, they are filled in for *something* and what is that "something"? There exists in certain branches of psychology the notion of an executive, or a control process, which is generally thought to be located in the prefrontal and frontal parts of the brain. I would like to suggest that the "something" that qualia are filled in for is not a "thing" but simply another brain process, namely, executive processes associated with the limbic system including parts of the anterior cingulate gyrus. This process connects your perceptual qualia with specific emotions and goals, enabling you to make choices—very much the sort of thing that the self was traditionally supposed to do. (For example, after having lots of tea, I have the sensation or urge—the qualia—to urinate but I'm giving a lecture so I choose to delay action until the talk is finished but also choose to excuse myself at the end instead of taking questions.) An executive process is not something that has all the properties of a full human being, of course. It is not a homunculus. Rather, it is a process whereby some brain areas such as those concerned with perception and motivation influence the activities of other brain areas such as ones dealing with the planning of motor output. Seen this way, filling in is a kind of treating and "preparing" of qualia to enable them to interact properly with limbic executive structures. Qualia may need to be filled in because gaps interfere with the proper working of these executive structures, reducing their efficiency and their ability to select an appropriate response. Like our general who ignores gaps in data given to him by scouts to avoid making a wrong decision, the control structure also finds a way to avoid gaps—by filling them in.¹⁵ Where in the limbic system are these control processes? It might be a system involving the amygdala and the anterior cingulate gyrus, given the amygdala's central role in emotion and the anterior cingulate's apparent executive role. We know that when these structures are disconnected, disorders of "free will" occur, such as akinetic mutism¹⁶ 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."

The *vigilant self*: A vital clue to the neural circuitry underlying qualia and consciousness comes from two other neurological disorders—penduncular hallucinosis and "vigilant coma" or akinetic mutism. The anterior cingulate and other limbic structures also receive projections from the intralaminar thalamic nuclei (cells in the thalamus), which in turn are driven by clusters of cells in the brain stem (including the cholinergic lateral segmental cells and the pendunculopontine cells). Hyperactivity of these cells can lead to visual hallucinations (penduncular hallucinosis), and we also know that schizophrenics have a doubling of cell number in these very same brain stem nuclei—which may contribute to their hallucinations.

Conversely, damage to the intralaminar nucleus or to the anterior cingulate results in coma vigilance or akinetic mutism. Patients with this curious disorder are immobile and mute and react sluggishly, if at

all, to painful stimuli. Yet they are apparently awake and alert, moving their eyes around and tracking objects. When the patient comes out of this state, he may say, "No words or thoughts would come to my mind. I just didn't want to do or think or say anything." (This raises a fascinating question: Can a brain stripped of all motivation record any memories at all? If so, how much detail does the patient remember? Does he recall the neurologist's pinprick? Or the cassette tape that his girlfriend played for him?) Clearly these brain stem and thalamic circuits play an important role in consciousness and qualia. 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 an integral part of the circuitry that embodies qualia and consciousness. Are they analogous to the power supply of a VCR or TV set or to the actual magnetic recording head and the electron gun in the cathode-ray tube?

The conceptual self and the social self: In a sense, our concept of self is not fundamentally different from any other abstract concept we have— such as "happiness" or "love." Therefore, a careful examination of the different ways in which we use the word "I" in ordinary social discourse can provide some clues as to what the self is and what its function might be.

For instance, it is clear that the abstract self-concept also needs to have access to the "lower" parts of the system, so that the person can acknowledge or claim responsibility for different self-related facts: states of the body, body movements and so on (just as you claim to "control" your thumb when hitching a ride but not your knee when I tap the tendon with my rubber hammer). Information in autobiographical memory and information about one's body image need to be accessible to the self-concept, so that thought and talk about self are possible. In the normal brain there are specialized pathways that allow such access to occur, but when one or more of these pathways is damaged, the system tries to do it anyway, and confabulation results. For instance, in the denial syndrome discussed in Chapter 7, there is no access channel between information about the left side of the body and the patient's self concept. But the self-concept is set up to try automatically to include that information. The net result of this is anosognosia or denial syndrome; the self "assumes" that the arm is okay and "fills in" the movements of that arm.

One of the attributes of the self-representation system is that the person will confabulate to try to cover up deficits in it. The main purposes of doing this, as we saw in Chapter 7, are to prevent constant indecisiveness and to confer stability on behavior. But another important function may be to support the sort of created or narrative self that the philosopher Dan Dennett talks about—that we present ourselves as unified in order to achieve social goals and to be understandable to others. We also present ourselves as acknowledging our past and future identity, enabling us to be seen as part of society. Acknowledging and taking credit or blame for things we did in the past help society (usually kin who share our genes) incorporate us effectively in its plans, thereby enhancing the survival and perpetuation of our genes.¹⁷

If you doubt the reality of the social self, ask yourself the following question: Imagine that there is some act you've committed about which you are extremely embarrassed (love letters and Polaroid photographs from an illicit affair). Assume further that you now have a fatal illness and will be dead in two months. If you know that people rummaging through your belongings will discover your secrets,

will you do your utmost to cover your tracks? If the answer is yes, the question arises, Why bother?

After all, you know you won't be around, so what does it matter what people think of you after you're gone? This simple thought experiment suggests that the idea of the social self and its reputation is not just an abstract yarn. On the contrary, it is so deeply ingrained in us that we want to protect it even after death. Many a scientist has spent his entire life yearning obsessively for posthumous fame—sacrificing everything else just to leave a tiny scratchmark on the edifice.

So here is the greatest irony of all: that the self that almost by definition is entirely private is to a significant extent a social construct—a story you make up for others. In our discussion on denial, I suggested that confabulation and self-deception evolved mainly as by-products of the need to impose stability, internal consistency and coherence on behavior. But an added important function might stem from the need to conceal the truth from other people.

The evolutionary biologist Robert Trivers¹⁸ has proposed the ingenious argument that self-deception evolved mainly to allow you to lie with complete conviction, as a car salesman can. After all, in many social situations it might be useful to lie—in a job interview or during courtship ("I'm not married"). But the problem is that your limbic system often gives the game away and your facial muscles leak traces of guilt. One way to prevent this, Trivers suggests, may be to deceive yourself first. If you actually believe your lies, there's no danger your face will give you away. And this need to lie efficiently provided the selection pressure for the emergence of self-deception.

I don't find Trivers's idea convincing as a general theory of self deception, but there is one particular class of lies for which the argument carries special force: lying about your abilities or boasting. Through boasting about your assets you may enhance the likelihood of getting more dates, thereby disseminating your genes more effectively. The penalty you pay for self-deception, of course, is that you may become delusional. For example, telling your girlfriend that you're a millionaire is one thing; actually believing it is a different thing altogether, for you may start spending money you don't have! On the other hand, the advantages of boasting successfully (reciprocation of courtship gestures) may outweigh the disadvantage of delusion—at least up to a point. Evolutionary strategies are always a matter of compromise.

So can we do experiments to prove that self-deception evolved in a social context? Unfortunately, these are not easy ideas to test (as with all evolutionary arguments), but again our patients with denial syndrome whose defences are grossly amplified may come to our rescue. When questioned by the physician, the patient denies that he is paralyzed, but would he deny his paralysis to himself as well? Would he do it when nobody was watching? My experiments suggest that he probably would, but I wonder whether the delusion is amplified when others are present. Would his skin register a galvanic response as he confidently asserted that he could arm wrestle? What if we showed him the word "paralysis"? Even though he denies the paralysis, would he be disturbed by the word and register a strong GSR? Would a normal child show a skin change when confabulating (children are notoriously prone to such behavior)? What if a neurologist were to develop anosognosia (the denial syndrome) as the result of a stroke? Would he continue to lecture on this topic to his students—blissfully unaware that he himself was suffering from denial? Indeed, how do I know that I am not such a person? It's

only through raising questions such as these that we can begin to approach the greatest scientific and philosophical riddle of all—the nature of the self.

Our revels now are ended. These our actors,
As I foretold you, were all spirits and
Are melted into air, into thin air....
We are such stuff
As dreams are made on,
And our little life
Is rounded with a sleep.

William Shakespeare

During the last three decades, neuroscientists throughout the world have probed the nervous system in fascinating detail and have learned a great deal about the laws of mental life and about how these laws emerge from the brain. The pace of progress has been exhilarating, but—at the same time—the findings make many people uncomfortable. It seems somehow disconcerting to be told that your life, all your hopes, triumphs and aspirations simply arise from the activity of neurons in your brain. But far from being humiliating, this idea is ennobling, I think. Science—cosmology, evolution and especially the brain sciences—is telling us that we have no privileged position in the universe and that our sense of having a private non-material soul "watching the world" is really an illusion (as has long been emphasized by Eastern mystical traditions like Hinduism and Zen Buddhism). Once you realize that far from being a spectator, you are in fact part of the eternal ebb and flow of events in the cosmos, this realization is very liberating. Ultimately this idea also allows you to cultivate a certain humility—the essence of all authentic religious experience. It is not an idea that's easy to translate into words but comes very close to that of the cosmologist Paul Davies, who said:

Through science, we human beings are able to grasp at least some of nature's secrets. We have cracked part of the cosmic code. Why this should be, just why *Homo sapiens* should carry the spark of rationality that provides the key to the universe, is a deep enigma. We, who are children of the universe—animated stardust—can nevertheless reflect on the nature of that same universe, even to the extent of glimpsing the rules on which it runs. How we have become linked into this cosmic dimension is a mystery. Yet the linkage cannot be denied.

What does it mean? What is Man that we might be party to such privilege? I cannot believe that our existence in this universe is a mere quirk of fate, an accident of history, an incidental blip in the great cosmic drama. Our involvement is too intimate. The physical species *Homo* may count for nothing, but the existence of mind in some organism on some planet in the universe is surely a fact of fundamental significance. Through conscious beings the universe has generated self-awareness. This can be no trivial detail, no minor by-product of mindless, purposeless forces. We are truly meant to be here.

Are we? I don't think brain science alone, despite all its triumphs, will ever answer that question. But that we can ask the question at all is, to me, the most puzzling aspect of our existence.

Ramachandran's Footnotes (pp. 296-298)

1. For clear introductions to the problem of consciousness, see Humphrey, 1992; Searle, 1992; Dennett, 1991; P. Churchland, 1986; P.M. Churchland, 1993; Galin, 1992; Baars 1997; Block, Ramachandran and Hirstein, 1997; Penrose, 1989.

The idea that consciousness—especially introspection—may have evolved mainly to allow you to simulate other minds (which inspired the currently popular notion of a "theory of other minds" module) was first proposed by Nick Humphrey at a conference that I had organized in Cambridge over twenty years ago.

2. Another very different type of translation problem also arises between the code or language of the left hemisphere and that of the right (see note 16, Chapter 7).

3. Some philosophers are utterly baffled by this possibility, but it's no more mysterious than striking your ulnar nerve at the elbow with a hammer to generate a totally novel electrical "tingling" qualia even though you may have never experienced anything quite like it before (or even the very first time a boy or girl experiences an orgasm).

4. Thus an ancient philosophical riddle going back to David Hume and William Molyneux can now be answered scientifically. Researchers at NIH have used magnets to stimulate the visual cortex of blind people to see whether visual pathways have degenerated or become reorganized, and we have also begun some experiments here at UCSD. But to my knowledge, the specific question of whether a person can experience a qualia or subjective sensation totally novel to him or her has never been explored empirically.

5. The pioneering experiments in this field were performed by Singer, 1993, and Gray and Singer, 1989.

6. It is sometimes asserted—on grounds of parsimony—that one does not need qualia for a complete description of the way the brain works, but I disagree with this view. Occam's razor—the idea that the simplest of competing theories is preferable to more complex explanations of unknown phenomena—is a useful rule of thumb, but it can sometimes be an actual impediment to scientific discovery. Most 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 resulted from rejecting Occam's razor and asking what if some deeper generalization were true which was not required by the available data, but which made unexpected predictions (which later turned out to be parsimonious, after all). It's ironic that most scientific discoveries result 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 that are not called for by the current data.

7. Please note that I am using the phrase "filling in" in a strictly metaphorical sense—simply for lack of a better one. I don't want to leave you with the impression that there is a pixel-by-pixel rendering of the visual image on some internal neural screen. But I disagree 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. What we mean by "filling in" is simply this: that one quite literally sees visual stimuli (such as patterns and colors) 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 to benefit a homunculus but to make some aspects of the information explicit for the next level of processing.

8. Tovee, Rolls and Ramachandran, 1996. Kathleen Armel, Chris Foster and I have recently shown that if two completely different "views" of this dog are presented in rapid succession, naive subjects can see only chaotic, incoherent motion of the splotches, but once they see the dog, it is seen to jump or turn in the appropriate manner—emphasizing the role of the "top-down" object knowledge in motion perception (see Chapter 5).

9. Sometimes qualia become deranged, leading to a fascinating condition called synesthesia, in which a person quite literally tastes a shape or sees color in a sound. For example, one patient, a synesthete, claimed that chicken has a distinctly "pointy" flavor and told his physician, Dr. Richard Cytowic, "I wanted the taste of this chicken to be pointed, but it came out all round . . . well, I mean it's nearly spherical; I can't serve this if it doesn't have points." Another patient claimed to see the letter "U" as being yellow to light brown in color, whereas the letter "N" was a shiny varnished ebony hue. Some synesthetes see this union of the senses as a gift to inspire their art, not as brain pathology.

Some cases of synesthesia tend to be a bit dubious. A person claims to see a sound or taste a color, but it turns out that she is merely being metaphorical—much the same way that you might speak of a sharp taste, a bitter memory or a dull sound (bear in mind, though, that the distinction between the metaphorical and the literal is extremely blurred in this curious condition). However, many other cases are quite genuine. A graduate student, Kathleen Armel, and I recently examined a patient named John Hamilton who had relatively normal vision up until the age of five, then suffered progressive deterioration in his sight as a result of retinitis pigmentosa, until finally at the age of forty he was completely blind. After about two or three years, John began to notice that whenever he touched objects or simply read Braille, his mind would conjure up vivid visual images, including flashes of light, pulsating hallucinations or sometimes the actual shape of the object he was touching. These images were highly intrusive and actually interfered with his Braille reading and ability to recognize objects through touch. Of course, if you or I close our eyes and touch a ruler, we don't hallucinate one, even though we may visualize it in our mind's eye. The difference, again, is that your visualization of the ruler is usually helpful to your brain since it is tentative and revocable—you have control over it—whereas John's hallucinations are often irrelevant and always irrevocable and intrusive. He can't do anything about them, and to him they are a spurious and distracting nuisance. It seems that the tactile signals evoked in John's somatosensory areas—his Penfield map—are being sent all the way back to his deprived visual areas, which are hungry for input. This is a radical idea, but it can be tested by using modern imaging techniques.

Interestingly, synesthesia is sometimes seen in temporal lobe epilepsy, suggesting that the merging of sense modalities occurs not only in the angular gyrus (as is often asserted) but also in certain limbic structures.

10. This question arose in a conversation I had with Mark Hauser.

11. Searle, 1992.

12. Jackendorf, 1987.

13. The patient may also say, "This is it; I finally see the truth. I have no doubts anymore." It seems ironic that our convictions about the absolute truth or falsehood of a thought should depend not so much on the propositional language system, which takes great pride in being logical and infallible, but on much more primitive limbic structures, which add a form of emotional qualia to thoughts, giving them a "ring of truth." (This might explain why the more dogmatic assertions of priests as well as scientists are so notoriously resistant to correction through intellectual reasoning!)

14. Damasio, 1994.

15. I am, of course, simply being metaphorical here. At some stage in science, one has to abandon or refine metaphors and get to the actual mechanism—the nitty-gritty of it. But in a science that is still in its infancy, metaphors are often useful pointers (For example, seventeenth-century scientists often spoke of light as being made of waves or particles, and both metaphors were useful up to a point, until they became assimilated into the more mature physics of quantum theory. Even the gene—the independent particle of beanbag genetics—continues to be a useful word, although its actual meaning has changed radically over the years.)

16. For an insightful discussion of akinetic mutism, see Bogen, 1995, and Plum, 1982.

17. Dennett, 1991.

18. Trivers, 1985.