Hill and Wang
A division of Farrar, Straus and Giroux
18 West 18th Street, New York 10011

Distributed in Canada by D&M Publishers, Inc.
Printed in the United States of America
Published in 2009 by Hill and Wang
First paperback edition, 2010

Grateful acknowledgment is made for permission to reprint the following material: Excerpt from Francisco Varela's translation of Antonio Machado's poem "Laying Down a Path in Walking," originally published in *Gaia, A Way of Knowing: Political Implications of the New Biology*, edited by William Irwin Thompson (Hudson, NY: Lindisfarne Press, 1987), pp. 48–64. Reprinted by permission of Amy Cohen Varela and William Irwin Thompson.

The Library of Congress has cataloged the hardcover edition as follows:
Noë, Alva.
Out of our heads : why you are not your brain, and other lessons from the biology of consciousness / Alva Noë.
p. cm.
Includes bibliographical references and index.
ISBN: 978-0-8090-7465-5 (hardcover : alk. paper)
1. Consciousness. I. Title.

QP411.N599 2009
612.8'233—dc22

2008031399

Paperback ISBN: 978-0-8090-1648-8

Designed by Jonathan D. Lippincott

www.fsgbooks.com

3 5 7 9 10 8 6 4 2

To the memory of Susan L. Hurley

PREFACE

We live in a time of growing excitement about the brain. Only the preoccupation with finding the gene for everything rivals today's widespread optimism regarding all things neuroscientific. Perception, memory, our likes and dislikes, intelligence, morality, whatever—the brain is supposed to be the organ responsible for all of it. It is widely believed that even consciousness, that Holy Grail of science and philosophy, will soon be given a neural explanation. In this era of expensive and flashy new brain-imaging technologies (such as functional magnetic resonance imaging and positron emission tomography), hardly a day goes by without the science pages of our leading newspapers and magazines publishing reports of important breakthroughs and new discoveries.

After decades of concerted effort on the part of neuroscientists, psychologists, and philosophers, only one proposition about how the brain makes us conscious—how it gives rise to sensation, feeling, subjectivity—has emerged unchallenged: we don't have a clue. Even enthusiasts for the new neuroscience of consciousness admit that at present no one has any plausible explanation as to how experience—the feeling of the redness of red!—arises from the action of the brain. Despite all the technology and the animal experimentation, we are no closer now to grasping the neural basis of experience than we were a hundred years ago. Currently, we lack even a back-of-the-envelope theory about what the behavior of individual cells contributes to con-

language. Science and philosophy are, if you like, conversations that have been going on for a long time. Of course, it will be hard for an outsider to sit down at the table and have a real sense of what is going on. And why should scientists be required to begin again anew each day so that the novice can understand what is under discussion?

The situation is different, however, if the conversation is, well, troubled. In my view, this is the case in contemporary cognitive science. The science of mind could benefit from interruption. It is time to slide our chairs back from the table and to invite intelligent latecomers to join our circle. In cognitive science, specialist jargon and technical details are too often an impediment to clear and honest thinking.

In some sense, then, this book is political. I am writing the book to change the world. Or at least to shake up the cognitive science establishment. I am aware that that's a tall order and that in some ways it may seem presumptuous even to try.

My book is political in another sense as well. American and European intellectual life is fragmented. Humanists—and I don't just mean college English professors but rather anyone whose first love is literature and art—have an awkward relation to science. For many humanists, science is a world apart. Some of them accept its findings uncritically and with indifference. Others disdain science; as far as they are concerned, science has nothing to teach us about what matters most: truth, beauty, art, meaning, experience. Scientists, for their part, have a no less problematic relation to the arts and humanities. Many of them do not appreciate the value of nonscientific research. And those who take an interest in art and literature are very often motivated to explain these phenomena away—for example, by investigating the underlying neuronal basis of aesthetic experience. (This conflict takes a curious and disturbing form where religion is concerned. On the one hand, some religious thinkers hold that religion is altogether insulated from criticism by science,

whereas others promote religious doctrine by pretending it is science. On the other hand, scientists, or at least representatives of the scientific worldview, act as though religious people are simply in error, as if they don't realize that religious doctrines lack empirical support.)

In this book I try to show, by example, that science and humanistic styles of thinking must engage each other. Physics used to be called natural philosophy (that's how Newton thought of it). In Germany today, the study of literature is known as *Literaturwissenschaft* (literary science). The idea that science and philosophy, or the humanities more generally, are separate spheres with their own standards and criteria is itself a bit of questionable ideology, a relic of the enthusiasm of an earlier modern age. Natural science is not sui generis. It is not value neutral. It is not discontinuous with broader human concerns. Nor is philosophy a free-for-all of opinion. Philosophy and science share a common aim: understanding. Science and philosophy must work together to advance toward understanding. This is especially so where the target of understanding is consciousness or, more basically, our own nature. The contemporary science of consciousness, at least as it is carried on in the mainstream, rests on shaky philosophical foundations. This makes for an alienated and distorted conception of our human life. It also makes for bad science.

In this book I argue that mind science, like biology more generally, must give pride of place to the whole, living being. I leave it to the reader to judge whether I am successful.

A note about the text of this book. I have made no use of footnotes or in-text references. Instead, I give references or make comments on the text in notes at the end of the book. Each chapter begins with a brief paragraph outlining the aim and topic of the chapter and ends with a brief conclusion or summary.

sciousness. This in itself is no scandal. It is a scandal if we allow the hype to obscure the fact that we are in the dark.

It is sometimes said that the neuroscience of consciousness is in its infancy. But that's not quite right, as it suggests that progress will take care of itself: it's just a matter of time and the normal process of maturation. A better image might be that of inexperienced hikers out on the trails without any clear idea where they are: they are lost and don't even know it! I am writing this book to help us figure out where we are and to show us the way forward.

In a way our problem is that we have been looking for consciousness where it isn't. We should look for it where it is. Consciousness is not something that happens inside us. It is something we do or make. Better: it is something we achieve. Consciousness is more like dancing than it is like digestion.

The aim of this book is to convince you of this. I also want to show you that this is what a genuinely biological aproach to the study of mind and human nature teaches us. The idea that the only genuinely scientific study of consciousness would be one that identifies consciousness with events in the nervous system is a bit of outdated reductionism. It is comparable to the idea that depression is a brain disease. In one sense, that is obviously true. There are neural signatures of depression. Direct action on the brain—in the form of drug therapy—can influence depression. But in another sense, it is obviously not true. It is simply impossible to understand why people get depressed—or why this individual here and now is depressed—in neural terms alone. Depression happens to living people with real life histories facing real life events, and it happens not only against the background of these individual histories but also against the background of the phylogenetic history of the species. The dogma that depression is a brain disease serves the interests of drug companies, no

doubt; it also serves to destigmatize the struggle with depression, which is a good thing. But it is false.

To move forward in our understanding of consciousness, we need to give up the internal, neural microfocus (as Susan Hurley and I once described it). The locus of consciousness is the dynamic life of the whole, environmentally plugged-in person or animal. Indeed, it is only when we take up this holistic perspective on the active life of the person or animal that we can begin to make sense of the brain's contribution to conscious experience.

This is a positive book. Human experience is a dance that unfolds in the world and with others. You are not your brain. We are not locked up in a prison of our own ideas and sensations. The phenomemon of consciousness, like that of life itself, is a world-involving dynamic process. We are already at home in the environment. We are out of our heads.

I have written this book with a particular audience in mind. I imagine that my reader is a lover of science and that he or she is fascinated by the problem of mind, by the fact of consciousness, and by how daunting it is to understand or explain these phenomena. I hope that cognitive scientists and philosophers interested in the mind will read the book and take note of its arguments. But I haven't directed my writing to them. My subject matter is basic to the conduct of normal neuroscience and psychology; it concerns what philosophers call the foundations of cognitive science. I want us to rethink what scientists have simply taken for granted: the basic, starting-point assumptions. For this reason I have tried, in writing this book, to avoid the jargon and insider-speak, the styles of language and argumentation, that already presuppose that one is a member of the cognitive science club.

I am not someone who disdains specialization and technical

1

AN ASTONISHING HYPOTHESIS

> The human body is the best picture of the human soul.
>
> —Ludwig Wittgenstein

Contemporary research on consciousness in neuroscience rests on unquestioned but highly questionable foundations. Human nature is no less mysterious now than it was a hundred years ago. If we are to understand our human nature, we need to make a fresh start. In this first chapter I lay out the basic challenge.

Consciousness Is Like Money

Stop and notice that you can believe in consciousness—appreciate the fact that we feel and think and that the world shows up for us—without believing that there is a place, or a moment in time, when and where consciousness happens or comes to be *inside* of us. As a comparison, consider that there's nothing about this piece of paper in my hand, taken in isolation, that makes it one dollar. It would be ludicrous to search for the physical or molecular correlates of its monetary value. The monetary value, after all, is not intrinsic to the piece of paper itself, but depends on the existence of practices and conventions and institutions. The marks or francs or pesos or lire in your wallet didn't change

physically when, from one day to the next, they ceased to be legal tender. The change was as real as it gets, but it wasn't a physical change in the money.

Maybe consciousness is like money. Here's a possibility: my consciousness now—with all its particular quality for me now—depends not only on what is happening in my brain but also on my history and my current position in and interaction with the wider world. It is striking that the majority of scientists working on consciousness don't even notice there is an overlooked theoretical possibility here. They tend to think that consciousness, whatever its ultimate explanation, must be something that happens somewhere and sometime in the human brain, just as digestion must take place in the stomach.

According to the now standard view, our conscious lives—the fact that we think and feel and that a world shows up for us—is achieved in us by the action of our brain. The brain produces images of the environment and manipulates those images in a process known as thought. The brain calculates and infers and eventually produces neural commands so that we act. We really are our brains, and our bodies are at most robotic tools at our brains' disposal. The brain is sole author of what is in fact a grand illusion: that we inhabit a richly detailed and meaningful world, that we are the sorts of beings we think we are. What are we, then? If the truth be told, we are brains in vats on life support. Our skulls are the vats and our bodies the life-support systems that keep us going.

Or so mainstream neuroscience, and writers of science fiction, would have it. Is my body a robot that my brain inhabits? Is the world a grand illusion? Is this really an intelligible conception of ourselves?

Are You Your Brain?

The fundamental assumption of much work on the neuroscience of consciousness is that consciousness is, well, a neuroscientific phenomenon. It happens inside us, in the brain.

All scientific theories rest on assumptions. It is important that these assumptions be true. In this book I will try to convince you that this starting assumption of consciousness research is badly mistaken. Consciousness does not happen in the brain. That's why we have been unable to come up with a good explanation of its neural basis.

Francis Crick, the Nobel Prize–winning codiscoverer of the structure of the DNA molecule, has proposed (in a book titled *The Astonishing Hypothesis*) that "you, your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules." With a flourish, he adds, "This hypothesis is so alien to the ideas of most people alive today that it can truly be called astonishing."

What is striking about Crick's hypothesis is how astonishing it isn't. It isn't surprising to be told that there is a thing inside each of us that thinks and feels and wants and decides. This was the view of the seventeenth-century philosopher René Descartes, who held that each of us is identical to an interior something whose essence is consciousness; each of us, really, is an internal *res cogitans*, or thinking thing. And this is the doctrine promulgated by many religious traditions. Of course, the religions, and Descartes himself, didn't teach that that thing inside us that thinks and feels is a part of our body, a bit of flesh, such as the brain. They supposed that it was something immaterial or spiritual, and so, in that sense, that it was something unnatural. How could mere matter—mere meat—achieve the powers of thought and feeling? Such a possibility boggles the mind. It is precisely

on this point, and only on this point, really, that today's neuroscientist breaks with tradition. As Patricia Churchland, a prominent philosopher of neuroscience, has written: "The weight of evidence now implies that it is the *brain*, rather than some nonphysical stuff, that feels, thinks, decides."

But what needs to be kept clearly in focus is that the neuroscientists, in updating the traditional conception of ourselves in this way, have really only succeeded in replacing one mystery with another. At present, we have no better understanding of how "a vast assembly of nerve cells and their associated molecules" might give rise to consciousness than we understand how supernatural soul stuff might do the trick. Which is just to say that the you-are-your-brain idea is not so much a working hypothesis as it is the placeholder for one.

Consciousness researchers in neuroscience like to think that they have broken with philosophy. They have left it behind and set off on the path of science. As Crick has written: "No longer need one spend time attempting . . . to endure the tedium of philosophers perpetually disagreeing with each other. Consciousness is now largely a scientific problem."

Crick is right that the problem of consciousness is now a problem for science. But this doesn't mean that it is no longer a problem for philosophy. For one thing, the aims of philosophy and of science are not different: to achieve understanding of the problems that matter to us. But that's just the beginning: it is a mistake to think that the new neuroscience of consciousness has broken with philosophy or moved beyond it. In fact, as we have been discovering, Crick and other neuroscientists have simply taken a specific family of philosophical assumptions for granted, so much so that their own reliance on them has become all but invisible to themselves. But the fact of the reliance is everywhere in evidence. Its perturbing influence is felt in the seeming mandatoriness of what we can think of as a kind of "gastric

juices" conception of consciousness—that is, the idea that consciousness happens in the head the way digestion happens in the stomach. I mentioned before that it is overoptimistic to think of the new neuroscience of consciousness as in its infancy. Developmentally, it would be more apt to characterize it as like a teenager. Like teenagers, neuroscience is in the grip of technology; it has a grandiose sense of its own abilities; and it is entirely lacking a sense of the history of what, for it, seems so new and exciting.

A Really Astonishing Hypothesis

It *would* be astonishing to learn that you are *not* your brain. All the more so to be told that the brain is not the thing inside of you that makes you conscious because, in fact, there is no thing inside of you that makes you conscious. It would then turn out that contemporary neuroscience has been in the thrall of a false dichotomy, as if the only alternative to the idea that the thing inside you that thinks and feels is immaterial and supernatural is the idea that the thing inside you that thinks and feels is a bit of your body. It would be astonishing to be told that we've been thinking about consciousness the wrong way—as something that happens in us, like digestion—when we should be thinking about it as something we do, as a kind of living activity.

In this book I advance this truly astonishing hypothesis: to understand consciousness in humans and animals, we must look not inward, into the recesses of our insides; rather, we need to look to the ways in which each of us, as a whole animal, carries on the processes of living in and with and in response to the world around us. The subject of experience is not a bit of your body. You are not your brain. The brain, rather, is part of what you are.

A Note on Terminology, and the Thesis Restated

In this book I use the term "consciousness" to mean, roughly, experience. And I think of experience, broadly, as encompassing thinking, feeling, and the fact that a world "shows up" for us in perception. Many writers have sought to define terms more narrowly than this. No doubt there are important distinctions that can and, for certain purposes at least, should be drawn. For example, a contrast is often made between thought or cognition, on the one hand, and sensation and feeling, or phenomenal experience, on the other. The contrast is between planning and carrying out an action, for example, and, say, experiencing the taste of licorice. When people draw this distinction it is usually because they think it is much easier to explain thought, say, than it is to explain the quality of our conscious episodes. For example, many theorists hold that thinking is a matter of computation and that we shed light on how brains think by comparing them with computers. As I discuss in Chapter 7, it is far from true that computers can think; moreover, I argue there, computers can't think largely for the same reason that brains can't. Meaningful thought arises only for the whole animal dynamically engaged with its environment, or so I contend. And indeed the same is true for the quality of our conscious episodes. The taste of licorice is not something that happens in our brains (although it is true that when we eat licorice, we do so by putting it in our mouths).

Conscious states are typically states that I can talk about, that influence what I do, and so they are states that I can make use of in planning. For example, my dislike of the taste of licorice is something that informs my larger cognitive and behavioral life. Among other things, it will influence my shopping behavior. Such a state is available or accessible to thought and talk; it is sometimes said that this marks a distinctive sort of consciousness

that the philosopher Ned Block has named access consciousness. The access consciousness of my feelings about licorice is one thing, however, and the experience of the licorice itself is another. The latter is an episode in what Block has called phenomenal consciousness, and the question of whether an episode is phenomenally conscious is, or so it seems, altogether different from the question of whether it is access conscious. To ask whether an episode is phenomenally conscious is to ask, in the philosopher Thomas Nagel's phrase, whether "there is something it is like to be" in that state. To ask whether it is access conscious is to ask whether the occurrence of the state influences what we say and do and want and plan and so on.

Other distinctions abound. To be conscious, as opposed to being unconscious, is to be awake, aroused, alert, as opposed to being asleep or knocked out. In ordinary language, self-consciousness means a kind of interfering attentiveness to how others view oneself. In philosophy and cognitive science, self-consciousness means something different. Self-consciousness is that feature of experience by virtue of which our experiences are *ours*. Experiences have a kind of "mine"-ness that makes them, distinctively, our own, or so some thinkers have maintained. Freud famously hypothesized the importance of unconscious desires and wishes in explaining human psychology.

Distinctions are useful, depending on your purposes. For my purposes, these distinctions don't matter in particular. When they do, I'll try to be careful to be clear about what I am referring to. The problem of consciousness, as I am thinking of it here, is that of understanding our nature as beings who think, who feel, and for whom a world shows up.

Another terminological issue arises in connection with the words "mind" and "brain." The latter refers to a part of the body found in the head and connected up to a larger system known as the nervous system. It is widely believed that the brain and the

larger nervous system of which it is a part play a special role in explaining our powers of mind (e.g., thought, memory, perception, emotion, and the like). Indeed, some scientists and philosophers think that the mind *is* the brain. Be that as it may, it is important to realize that no one holds that the concept of brain and the concept of mind are the same. To have a mind is, roughly, in my sense, to be conscious—that is, to have experience and to be capable of thought, feeling, planning, etc. To have a brain, on the other hand, is to have a certain kind of bodily organ or part. Ordinary language is sometimes a bit confused about this, so we need to be careful. Being intelligent, for example, is said to be a matter of having brains.

My central claim in this book is that to understand consciousness—the fact that we think and feel and that a world shows up for us—we need to look at a larger system of which the brain is only one element. Consciousness is not something the brain achieves on its own. Consciousness requires the joint operation of brain, body, and world. Indeed, consciousness is an achievement of the whole animal in its environmental context. I deny, in short, that you are your brain. But I don't deny that you have a brain. And I certainly don't deny that you have a mind. To have a mind, though, requires more than a brain. Brains don't have minds; people (and other animals) do.

The Man with Two Brains

I have always been a fan of Carl Reiner's hilarious movie *The Man with Two Brains*. Steve Martin plays the lead role, a brain surgeon named Dr. Hfuhruhurr who falls in love with a disembodied brain awaiting a transplant. This is the brain of the woman of his dreams. Now all he needs—all *she* needs—is a body! He sets about a villainous scheme to get his hands on the

body of the beautiful and svelte Dolores Benedict (played by Kathleen Turner). The joke is that, unbeknownst to him, the self whose brain he loves has an eating disorder. By the time she recovers from the brain transplant, she's morbidly and unattractively obese. (He loves her anyway!)

This is the stuff of science fiction. Pretty far-fetched, to be sure. The fact that we find it at all comprehensible, let alone compelling, shows that the "astonishing hypothesis" of the establishment neuroscientist now belongs to the conventional wisdom of the culture at large. We think of ourselves—or find it easy to take seriously the idea of ourselves—as dependent on our brains in a special sort of way, very different from the way we depend on our hearts, say. You gotta have heart, yes. But it is the brain, with its distinctive neuronal snap, crackle, and pop, that is our ground. We inhere in our brains. What makes us the kind of thing we are—beings who can feel and reason and think and see—is accomplished in our bodies by our brains.

I ask again: Is this a plausible conception of ourselves? Reiner's movie casts an interesting light on this question. The film itself needs to present us with communication between the Martin character and his beloved brain-in-a-cookie-jar. But how can it do this? How, for example, to capture the fact that the lovely female voice Martin hears—what we in the audience experience as a voice-over—is actually the voice of the person in that brain-in-a-cookie-jar? Film normally trades on the ventriloquist effect. We hear the voice coming from the mouth because we see the mouth move in synchrony with just those words. Vision captures and directs what we hear. In fact, this is an important part of normal speech perception. The problem with a speaking brain is that it has no mouth. What ties the sounds to the brain? What makes them *its* words? The movie strikes on a silly but funny solution. The brain glows and pulsates in synchrony with its spoken words. What makes this solution interest-

ing, as well as silly and funny, is that, in a way, it's cheating. Brains don't pulsate or change colors, and by introducing this feature you are, in effect, giving the brain a body or, more important, a face (what the brain is supposed to lack). And maybe that's not just a somewhat confused filmic conceit but something of a conceptual necessity. It's hard even to conceive of a consciousness that lacks a face. That's why, tragically, even friends and family find it difficult to empathize with Parkinson's patients whose faces have grown masklike. And that's why, in a love scene in *The Man with Two Brains*, the Steve Martin character puts a scarf around the base of his love's brain-in-a-cookie-jar, a hat on top, and bright red candy-wax lips on the front. Wittgenstein wrote that it is only of what looks and behaves like a person that we say it sees, thinks, feels. The problem with a brain is that it doesn't look and behave like a person.

Consciousness in a Petri Dish?

If the new neuroscience establishment is right, then it ought to be possible, at least in principle, to have consciousness in a petri dish. All that would be required for consciousness in a petri dish is that the cells be wired up to each other and stimulated in a suitable matter.

My own view is that the suggestion that cells in a dish could be conscious—or that you could have a conscious brain in a vat—is absurd; it's time to overhaul our starting assumptions about what consciousness is if they lead us to such a conclusion.

Consider, first of all, that the vat, or petri dish, couldn't be a mere dish or bucket, as Evan Thompson and Diego Cosmelli have discussed in an essay. It would have to supply energy to nourish the cells' metabolic activity and it would have to be capable of flushing away waste products. The vat would have to be

very complicated and specialized in order to control the administration of stimulation to the brain comparable to that normally provided to a brain by its environmentally situated body. If you actually try to think through the details of this thought experiment—this is something scientists and philosophers struck by the brain-in-a-vat idea almost never do—it's clear that the vat would have to be, in effect, something like a living body. But then, it would seem, the thought experiment teaches us what we knew already: not that we are our brains but rather that living animals like us can be, well, conscious.

Presumably it is an empirical question just how many cells would be necessary for conscious activity. It is consistent with what we now know to think it could very well turn out that in order to get consciousness in a vat, you'd have to have a whole, suitably activated, healthy brain in the vat. Recent work on the neural basis of visual consciousness has, as a matter of fact, tended to suggest that large-scale, ongoing interactions between widely separate areas of the brain are necessary for visual consciousness.

But now ask yourself: Do we have any reason, in advance of our Frankensteinian researches, to think that the whole brain is the outer limit of what might be needed for consciousness in a petri dish? If we can't draw the boundary in advance at this or that brain region, then how can we be confident that we can draw the boundary at the limits of the brain itself? Maybe consciousness depends on reliable interactions between what is going on in the brain and what is going on in nonbrain parts of the body. It could even turn out that consciousness depends on interactions between the brain and the body and bits of the world nearby. So maybe, to get consciousness in the dish, we'd need not only brain and body but also a reasonable facsimile of the environment in the dish too.

The point of this line of questioning will by now be clear. Our

philosopher-neuroscientists with their brain-in-a-vat fantasies fail to notice that their fantasies are taking a stand on what surely is an open empirical question: How much is minimally necessary for consciousness in a petri dish?

Taking the Problem Seriously

These are hard questions. And they aren't merely academic. Consider the case of a thirty-nine-year-old Belgian stroke victim who fell into a coma. Laura Spinney in the *Guardian* (April 15, 2004) reported:

> Doctors concluded that she was unlikely to regain consciousness and, after a time, diagnosed her condition as persistent vegetative state (PVS). One of the criteria on which they based their decision was her inability to blink or track a moving object with her eyes. It was only when they discovered that the stroke had damaged a cranial nerve, preventing her from opening her eyes, that they realized their error. If they opened her eyes for her, she followed their instructions. Having regained full consciousness soon after her stroke, she revealed she had overheard all the bedside discussions as to whether it was worth keeping her alive. At no point had she wanted to die.

The misdiagnosis of persistent vegetative state is horrifying but all too understandable. In normal circumstances it isn't difficult to know whether someone is uncomfortable, or in pain. In normal circumstances, how we feel finds expression in our faces and in our movements. These movements of the face, voice, and body are not mere signals to others, devices for effectively communicating with them. We don't first feel glad and then choose

to express our gladness to others in a smile, just as we don't first feel pain and then produce a grimace for the information of others. As William James first noticed, the grimace and the smile belong to our state of consciousness. They are not so much evidence of what is going on within us as they are, in fact, enactments of our condition. They are its natural expression. And there are probably good evolutionary reasons for this. The fright my fellow human being (or monkey or chimpanzee or whatever) feels at the arrival of a predator is not of much less significance to me than it is to him, and group cohesion surely depends on our ability to read each other's mind.

The point is that circumstances are not normal in the clinic. Obviously, the mere absence of the normal behavioral markers of consciousness does not entail the absence of consciousness. But what is the alternative to looking and listening to what someone says and does, to how they look? The Belgian stroke patient was lucky twice over. First, her mental presence was in fact detected. Second, she quickly recovered. Other patients with severe paralysis and loss of speech have not been so lucky. For example, thirty-two-year-old Julia Tavalaro spent six years in a New York chronic care hospital where she was known as "the vegetable" before a loved one noticed indicators of consciousness. In fact, she was entirely conscious the whole time: she was simply unable to give any signs to others. She'd spent six years trapped inside an inert body, unable to communicate with the outside world in any way. She eventually returned home and died at the age of sixty-eight. This condition, now known as locked-in syndrome, is brought about by brain-stem injury typically caused by a stroke. Because of the anatomy of the brain stem, patients with "classical" locked-in syndrome are typically able to move their eyes and deploy elaborate blinking and looking codes to communicate. Several such patients have written books. I have seen a videotape of a man with locked-in syn-

drome. In the initial shots you see the impassive and inert face of a man who appears to be blinking reflexively. The camera slowly pans back and you realize that the man is in fact staring at a computer screen. With his blinking he is actively controlling a cursor on the screen and managing an online database of locked-in syndrome sufferers in France!

But there are also known cases of total locked-in syndrome. The correct diagnosis of total locked-in syndrome—or, indeed, the more typical eye-movement locked-in syndrome—is extremely difficult. Tellingly, family members or caregivers are more likely to make the diagnosis than physicians. Sadly, it is almost certain that until recently all patients with locked-in syndrome have been mistakenly supposed to be mere vegetables, lacking all sentience, and have probably been allowed to endure slow and painful deaths by starvation. There are very few documented cases of total locked-in syndrome. This in itself is a frightening fact.

One does not need to turn to extreme forms of brain injury such as locked-in syndrome to appreciate the practical importance of the problem we are isolating. When my four-year-old son August was in the hospital for a hernia operation, before they wheeled him into surgery I asked the anesthesiologist whether he could assure me that August would not suffer any pain or discomfort during the operation. He replied that there was no cause for worry: he would personally monitor August's heart rate and would watch his face closely for signs of discomfort. I was reassured that the doctor would be paying attention. But I certainly wondered whether the absence of these very primitive behavioral and physiological indicators was reliable evidence that my son was free of awareness of what he was undergoing.

Locked-in syndrome, and the medical practice of anesthesiology, are forceful reminders that doctors can't afford to rely

alone on behavioral expressions of mental state. Persistent vegetative state, in contrast, serves to remind us of the converse. The persistent vegetative state is thought to be a condition of wakefulness without consciousness. But it is not uncommon for patients in this condition to respond to sounds, to sit up and move their eyes, to shout out, to grimace, to laugh, smile, or cry. Suppose it is your beloved who lies there jumping at the sound of the door slamming, her eyes darting around furiously. She cries out in seeming rage or purrs with apparent contentment. What would convince you that your loved one is unfeeling, absent, that she has become a vegetable? Whereas with locked-in syndrome we are challenged to believe that behind the masklike wall of a face there is a lively intellect at work, with persistent vegetative state we struggle to take seriously the thought that there is an absence of feeling and subjectivity behind what moves us as an expressive face.

Looking into the Head

We might turn to the technologies of brain scanning in the hope that these will enable us to look into the living brain itself to find out what is going on in there. The fact that brain-imaging studies of patients with locked-in syndrome—positron emission tomography (PET), functional magnetic resonance imaging (fMRI), as well as electroencephalography (EEG)—tend to show normal levels of cortical activity can be taken to be a confirmation of the judgment that patients with locked-in syndrome have normal mental lives. It is much harder, though, when we turn to patients in the persistent vegetative state. Here what confronts us is not so much direct evidence of the lack of consciousness as the absence of normal brain-imaging findings. Does the absence of normal brain profiles in patients in the persistent vege-

tative state help us decide whether they are sentient or not? Would the mere absence of normal patterns of neural activity as modeled by functional imaging technologies such as fMRI or PET satisfy you that your loved one was now little more than a vegetable?

Actually, things are more complicated. Although patients in the persistent vegetative state show markedly reduced global brain metabolism, so do people in slow-wave sleep and patients under general anesthesia. But sleepers and surgery patients wake up and resume normal consciousness, whereas patients in the persistent vegetative state rarely do. Remarkably, in the small number of cases in which brain imaging has been attempted in patients who have recovered from the persistent vegetative state, regaining full consciousness, it would appear that global metabolic levels remain low even after full recovery. Moreover, external stimuli such as sounds or pinpricks produced significant increases in neuronal activity in primary perceptual cortices. Interesting new work by Steven Laureys and his colleagues in Belgium indicates that vegetative patients show strikingly impaired functional connections between distant cortical areas and between cortical and subcortical structures. In addition, they show that in cases where consciousness is recovered, even if overall metabolic activity stays low, these functional connections between brain regions are restored. These findings are important and point in the direction of a deeper understanding of what is happening in the brain in the persistent vegetative state.

But this doesn't change the fact that at present we are not even close to being able to use brain imaging to get a look inside the head to find out whether there is consciousness or not. Consider these simple questions: Does a patient in the persistent vegetative state feel physical pain—for example, the pain of thirst or hunger, or the prick of a pin? Does she hear the sound

of the door slamming? We know she turns her head in response to the sound, and we know she withdraws her hand from the pinprick. We also know that there is some significant neural activity produced in primary perceptual cortices by these stimuli. Is the patient in the persistent vegetative state a robot, responding reflexively to stimulation, but without actually feeling anything? And, more important, is this something that brain imaging could ever help us decide?

We don't know how to answer these questions. It is disturbing to learn that so far there are no theoretically satisfying or practically reliable criteria for deciding when a person with brain injury is conscious or not. At present, doctors and relatives have to deal with these questions without guidance from science or medicine. For example, the press tended to treat the widely discussed case of Terri Schiavo as one in which science, armed with cold hard facts about the nature of Schiavo's brain damage, did battle with family members who were blinded at once by their love for their daughter and their religious fundamentalism. Sadly, science doesn't have the hard facts.

The New Phrenology?

It would be hard to overstate the extent to which the fervor about the brain-based view of consciousness is driven by the development in the last few years of new technologies of brain imaging. Until very recently, postmortem autopsy has been just about the only way to examine the brain of a person with known neurological deficits. Ethical considerations prevent scientists from deploying the sorts of invasive techniques that are used on animals. The brain has remained, for science, a black box. At best we have been able to draw conclusions about its design and functionality by looking at what possessors of brains can say and

do. Things are different now, or so it is widely believed. The development of PET and more recently fMRI—technologies of functional imaging—now enable us to penetrate the black box. Brain imaging provides colorful pictures of the brain, enabling us to see how it lights up in action as it performs its functions.

Given the huge personal and institutional investment in brain-scanning methods and technologies, it is understandable that there is so much hype about the power of functional imaging. It is hard to doubt that these technologies will add to our ability to move forward in our quest to understand the conscious mind. But this is all the more reason to pause and step back from the hype. In fact, functional imaging raises important and still unresolved methodological problems.

PET and fMRI yield multicolored images. The colors are meant to correspond to levels of neural activity: the pattern of the colors indicates the brain areas where activity is believed to occur; brighter colors indicate higher levels of activity. It is easy to overlook the fact that images of this sort made by fMRI and PET are not actually pictures of the brain in action. The scanner and the scientist perform a task that is less like gathering a photographic or X-ray image than it is like the process whereby a police sketch artist produces a drawing of a suspect based on interviews with a number of different witnesses. Such drawings carry valuable information about the criminal, to be sure, but they are not direct records of the criminal's face; they are, rather, graphical renderings based on perhaps conflicting reports of what different individuals claim to have seen. Such a composite sketch reflects a conjecture or hypothesis about, rather than a recording of, the perpetrator. Indeed, there is nothing in the process that even guarantees that there is a single perpetrator, let alone that the sketch is a good likeness.

In a similar way, images produced by PET and fMRI are not in any straightforward way traces of the psychological or mental

phenomena. Rather, they represent a conjecture or hypothesis about what we think is going on in the brains of subjects. To appreciate this, consider that we face a problem from the very beginning about how to decide what neural activity is relevant to a mental phenomenon we want to understand. Scientists start from the assumption that to every mental task—say, the judgment that two given words rhyme—there corresponds a neural process. But how do we decide which neural activity going on inside you when you make a rhyming judgment is the neural activity associated with the mental act? To do that, we need to have an idea about how things would have been in the brain if you hadn't performed the rhyming judgment; that is, we need a baseline against which to judge whether or not the deviation from the baseline corresponds to the mental act. One way to do this is by comparing the image of the brain at rest with the image of the brain making a rhyming judgment. The rhyming judgment presumably depends on the neural activity by virtue of which these two images differ. But how do we decide what the brain at rest looks like? After all, the brain is never at rest. For example, there are stages of sleep when your brain is working harder than it does at most times during the day!

Comparison provides the best method available for uncovering the areas of the brain that are critically involved in the performance of a cognitive function. For example, suppose you were to produce a bunch of PET images of people listening to recordings of spoken words and then making judgments about whether given pairs of words rhyme. To isolate the activation responsible for the rhyming judgment, as distinct from that responsible for the auditory perception of the spoken words, a standard procedure would be to compare these images with a second set of images of people listening to recordings of spoken words but not making rhyming judgments. Whatever areas are active in the first set of images, but not the second, would be

plausible candidates for the place in the brain where the rhyming judgment takes place.

This method of comparison is cogent and it holds promise. But it is worth stressing that its reliability depends on a number of background assumptions, not all of which are unproblematic, as Guy C. Van Orden and Kenneth R. Paap have convincingly argued. For one thing, sticking to our example, the comparison method assumes that there is no feedback between what the brain is doing when we make a rhyming judgment and what the brain is doing when we perceive the words. If there is indeed feedback, then it would follow that overlapping regions in the images do not necessarily correspond to a common neural factor.

As a matter of fact, it is highly likely that there is feedback. Neural activity in the brain during perception, for example, is not a one-way thing. Neural activity is characterized by loops and two-directionality. There are neural pathways heading back into the brain from the senses, but there are even more neural pathways heading back out again. This should not be surprising. Consider how much easier it is to hear a sound that you are expecting than one that you are not expecting. The assumption that there is no feedback in the neural circuitry is the flip side of a different assumption that we can factor the cognitive act itself into distinct, modular acts of perceiving the words (on the one hand) and judgments about whether they rhyme (on the other). That's a substantive empirical claim about the character and composition of cognitive acts themselves and certainly not something that can be simply taken for granted.

I am using the rhyming case as an illustrative example. My aim is not to show that, in fact, the method of comparison is misguided. What I do want to bring out is that brain scanners don't simply show us what is going on when we listen and judge. In a way, these considerations about feedback in the brain and cognitive models are only the tip of the iceberg. PET and fMRI have

very low spatial and temporal resolution. When you localize events in the brain using these techniques, you localize them to cubic regions of between 2 and 5 mm—that is, to regions in which there are hundreds of thousands of cells. If there is specialization or differentiation among these cells, that won't show up in the picture. Nor, for that matter, can we be sure exactly when neural events are happening. Cellular events unfold at the scale of thousandths of a second, but it can require much longer time scales (large portions of a minute) to detect and process signals for making images. For these reasons, scientists have developed techniques of normalizing data. Typically, data from different subjects is averaged. The averaging process involves the loss of considerable information. After all, brains differ from one another no less than faces and fingerprints do. Just as the average American taxpayer has no set height and weight, so averaged neural activity has no set location in any particular brain. For this reason, scientists project their findings onto an idealized, stock brain. The pictures we see in the science magazines are not snapshots of a particular person's brain in action.

Finally, putting all this to one side, it is important to be clear that there is no sense in which PET or fMRI pictures deliver direct information about consciousness or cognition. They do not even deliver direct representations of neural activity. Functional brain-imaging systems such as PET and fMRI build images based on the detection of physical magnitudes (such as radio or light waves) that are believed to be reliably correlated with metabolic activity. For example, in PET, one injects a positron-emitting isotope into the bloodstream; PET detects the emission of gamma rays caused by the collision of positrons and electrons. In this way, the PET image carries indirect information about metabolic activity based on the direct measurement of a physical magnitude, which is in turn supposed to carry information about neural activity. The latter supposition is not unreasonable. Neural events require oxygen, and so they require blood. The neural

activity, in its turn, is supposed to correlate to significant mental activity. Brain scans thus represent the mind at three steps of removal: they represent physical magnitudes correlated to blood flow; the blood flow in turn is correlated to neural activity; the neural activity in turn is supposed to correlate to mental activity. If all the assumptions are accurate, a brain-scan image may contain important information about neural activity related to a cognitive process. But we need to take care not to be misled by the visual, pictorial character of these images. Brain scans are not pictures of cognitive processes in the brain in action.

Conclusion: You Are Not Your Brain

Empirical research on consciousness and human nature takes for granted that the problem for science is to understand how consciousness arises in the brain. That consciousness arises in the brain goes unquestioned. In the meantime, guns blazing, engines roaring, we are going nowhere in our quest to understand what we are. In this chapter I ask whether our inability to explain consciousness and the workings of our minds stems precisely from our unquestioned assumptions. In the remainder of this book I seek to demonstrate that the brain is not the locus of consciousness inside us because consciousness has no locus inside us. Consciousness isn't something that happens inside us: it is something that we do, actively, in our dynamic interaction with the world around us. The brain—that particular bodily organ—is certainly critical to understanding how we work. I would not wish to deny that. But if we want to understand how the brain contributes to consciousness, we need to look at the brain's job in relation to the larger nonbrain body and the environment in which we find ourselves. I urge that it is a body- and world-involving conception of ourselves that the best new science as well as philosophy should lead us to endorse.

2

CONSCIOUS LIFE

> My attitude towards him is an attitude towards a soul. I am not of the *opinion* that he has a soul.
>
> —Ludwig Wittgenstein

I begin with what can seem to be the most challenging of problems about consciousness, what philosophers call "the problem of other minds." Can we know the minds of others? How do we decide whether other people are conscious? And what about the consciousness of other species? The problem of other minds can seem insurmountable. This is because we think that the problem we face is a theoretical one: how to acquire knowledge of another's mind on the basis of what he or she says and does, or on the basis of a neural signature. But we don't face this problem. The basis of our confidence in the minds of others is practical. We cannot take seriously the possibility that others lack minds because doing so requires that we take up a theoretical, detached stance on others that is incompatible with the kind of life that we already share with them. All this points to something paradoxical about the science of the mind: science requires detachment, but mind can only come into focus if we take up an altogether different, more engaged attitude. Does this mean a science of the mind must be impossible? No. There is a way forward for science. The solution comes when we recognize that there is a rigorously empirical alternative to mechanistic detach-