THE FEELING OF LIFE ITSELF

WHY CONSCIOUSNESS IS WIDESPREAD BUT CAN'T BE COMPUTED

CHRISTOF KOCH

"Koch's mind-stretching book provides a rich feast."
—*New Scientist*

The Feeling of Life Itself

Why Consciousness Is Widespread but Can't Be Computed

Christof Koch

The MIT Press
Cambridge, Massachusetts
London, England

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.
Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data is available.

ISBN: 978-0-262-04281-9

10 9 8 7 6 5 4 3 2 1

Contents

Preface: Consciousness Redux

To bring home the centrality of consciousness to life, consider a devil's bargain in which you gain unlimited wealth at the expense of your conscious experiences. You get all the money you want but must relinquish all subjective feeling, turning into a zombie. From the outside, everything appears normal—you speak, act, dispose of your vast riches, engage in a vigorous social life, and so on. Yet your inner life is gone; no more seeing, hearing, smelling, loving, hating, suffering, remembering, thinking, planning, imagining, dreaming, regretting, wanting, hoping, dreading. From your point of view, you might as well be dead, for it would feel the same—like nothing.

How experience comes into the world has been an abiding mystery since the earliest days of recorded thought. Aristotle warned his readers more than two thousand years ago that "to attain any assured knowledge about the soul is one of the most difficult things in the world." Known as the *mind–body problem,* this puzzle has occupied philosophers and scholars throughout the ages. Your subjective experience appears radically different from the physical stuff that makes up your brain. The foundational equations of physics, the periodic table of chemical elements, the endless ATGC chatter of your genes—none of these express anything about consciousness. Yet you awaken each morning to a world in which you see, hear, feel, and think. Experience is the only way you know about the world.

How does the mental relate to the physical? Most assume that the mental emerged from the physical when the physical became sufficiently complex. That is, in the eons before big brains like ours evolved on this planet, the mental did not exist. Yet are we really to believe that until that point in time (in a memorable phrase of the physicist Erwin Schrödinger), the world "remained a play before empty benches, not existing for anybody, thus

quite properly speaking not existing"? Alternatively, perhaps the mental was always present, allied with the physical, but not in a form readily recognizable? Perhaps consciousness predates the arrivals of big brains? This is the less-traveled road that I will take here.

When did *your* consciousness begin? Was your first experience one of confusion and chaos as the birthing process ejected you into a harsh world, blinding you with bright lights, overwhelming sounds, and a desperate need for oxygen? Or perhaps even earlier, of warmth and security in your mother's womb?

How will your stream of experience end? Snuffed out abruptly like a candle, or fading gradually? Can your mind, chained to a dying animal, encounter the numinous in a near-death experience? Can advanced technology come to the rescue and transition your mind to an engineered medium, a ghost in a new shell, by uploading your mind to the Cloud?

Do apes, monkeys, and other mammals hear the sounds and see the sights of life? Are dogs just machines, as René Descartes famously argued, or do they experience the world in a riot of redolent smells?

Then there is the urgent question of the day—can computers experience anything? Can digital code feel like something? Dramatic progress in machine learning has crossed a threshold, and human-level artificial intelligence may come into the world within the lifetime of many readers. Will these AIs have human-level consciousness to match their human-level intelligence?

In this book, I will show how these questions, formerly the sole province of philosophers, novelists, and moviemakers, are now being addressed by scientists. Powered by advanced instrumentation that peers deep into the brain, the science of consciousness has seen dramatic progress over the past decade. Psychologists have dissected out which cognitive operations underpin any one conscious perception. Much cognition occurs outside the limelight of consciousness. Science is bringing light to these dark passages where strange, forgotten things live in the shadows.

I dedicate two chapters to tracking the footprints of consciousness within its principle organ, the nervous system. Surprisingly, many brain regions do not contribute meaningfully to experience. This is true for the cerebellum, despite having more than four times more nerve cells than neocortex. Even in neocortical tissue, the most complex piece of highly excitable matter in the known cosmos, some sectors have a much more intimate relationship to experience than others.

In the fullness of time, the quest for the neural footprints of consciousness will pursue its prey to its lair somewhere in the thicket of the nervous system. Sooner or later, scientists will know which assemblies of nerve cells, expressing which proteins, and active in some mode, house any one experience. This discovery will be a high water mark for science. It will be enormously beneficial to neurological and psychiatric patients.

Yet knowing the neural correlates of consciousness does not answer the more fundamental question: Why these neurons and not those? Why this vibration and not that one? Identifying some form of physical activity as generating a feeling is laudable progress. But ultimately we want to know why this mechanism goes hand in hand with experience. What is it about the biophysics of the brain, but not, say, the liver—another complex biological organ—that evokes the ephemeral feelings of life?

What we need is a quantitative theory that starts with experience and proceeds from there to the brain. A theory that infers and predicts where experience can be found. To me, the most exciting development over the last decade has been the genesis of such a theory, a first in the history of thought. Integrated information theory considers the parts and their interactions that make up a whole, whether evolved or engineered, and derives, via a well-specified calculus, the quantity and the quality of the experience of this whole. The beating heart of *The Feeling of Life Itself* comprises two chapters outlining the theory and how it defines any one conscious experience in terms of intrinsic causal powers.

From the austerity of these abstract considerations, I dive into messy clinical practice. I describe how the theory has been used to build a tool to detect the presence and absence of consciousness in unresponsive patients. I next discuss some of the theory's counterintuitive predictions. If the brain is cut in the right place, its unitary mind splits into two minds that coexist within a single skull. Conversely, if the brains of two people are directly connected via a futuristic *brain-bridging* technology, their distinct minds could fuse into a single mind at the cost of their individual minds, which will be extinguished. The theory predicts that consciousness without any content, known as *pure experience* within certain meditative practices, can be achieved in a near-silent cortex.

After considering why consciousness evolved, *The Feeling of Life Itself* turns to computers. The basic tenet of today's dominant faith, its zeitgeist, is that digital, programmable computers can, in the fullness of time,

simulate anything, including human-level intelligence and consciousness. Computer experience is just a clever hack away.

According to integrated information theory, nothing could be further from the truth. Experience does not arise out of computation. Despite the near-religious belief of the digerati in Silicon Valley, there will not be a Soul 2.0 running in the Cloud. While appropriately programmed algorithms can recognize images, play Go, speak to us and drive a car, they will never be conscious. Even a perfect software model of the human brain will not experience anything, because it lacks the intrinsic causal powers of the brain. It will act and speak intelligently. It will claim to have experiences, but that will be make-believe—fake consciousness. No one is home. Intelligence without experience.

Consciousness belongs to the natural realm. Just like mass and charge, it has causal powers. To create human-level consciousness in a machine, the intrinsic causal powers of the human brain must be instantiated at the level of the metal, the transistors and wiring making up its hardware. I will show that the intrinsic causal powers of contemporary computers is puny compared to those of brains. Thus, artificial consciousness calls either for computer architectures radically different from those of today's machines or for a merging of neural and silicon circuits as envisioned by transhumanists.

In the concluding chapter, I survey nature's wide circuit. Because of the vast complexity of brains of so-called simple animals, IIT implies experience in parrots, ravens, octopuses, and bees. As nervous systems devolve into the primitive nerve net of jellyfish, their associated experience will lessen. But single-cell microorganisms contain untamed molecular complexity inside their cellular envelope, so they too are likely to feel an itsy-bitsy bit like something.

Integrated information theory has captured the imagination of philosophers, scientists, and clinicians as it opens myriads of doors to experimentation and because of its promise to illuminate those aspects of reality that have been, until now, beyond the pale of empirical investigations.

Any entrepreneur launching a new company against overwhelming odds must have a healthy amount of self-delusion. This is essential to remain motivated to work crazy hours year after year. Accordingly, I wrote this book assuming the theory is true and discard the scholar's cautionary habit of prefacing every statement with "under certain conditions." I do note current controversies and cite extensive and up-to-date literature in

Acknowledgments

Writing books is one of life's great pleasures, rewarding at an intellectual and emotional level over a protracted period, unlike the more fleeting pleasures of the body. Thinking about the book's content, discussing it with others, revising it, and working with editors, artists and the publisher provides a focus for one's mental energies.

I would like to thank everyone who engaged with me over the past three years of writing.

Judith Feldmann took my prose and edited it. Bénédicte Rossi was the artist who turned my cartoons into beautiful drawings. The book's title is an amalgamation of a comment by Elizabeth Koch, who exclaimed after one of my talks that "you study the feeling of life," and the title of Francis Crick's book *Life Itself: Its Origin and Nature*.

Many friends and colleagues read drafts, identified infelicities and inconsistencies and helped me sharpen the underlying concepts. In particular, I would like to acknowledge with gratitude Larissa Albantakis, Melanie Boly, Fatma Deniz, Mike Hawrylycz, Patrick House, David McCormick, Liad Mudrik, and Giulio Tononi, who took the time to carefully read the entire text and emend it. The philosophers Francis Fallon and Matthew Owen helped to clarify some conceptual troubling issues. My daughter, Gabriele Koch, edited key sections. The book is better for all of their efforts.

During the day, I'm the chief scientist and president of the Allen Institute for Brain Science in Seattle, studying the mammalian brain at the cellular level. The science we carry out at our Institute has informed many aspects of this book. I thank the late Paul G. Allen for providing the vision and the means to allow my colleagues and I to tackle hard questions under the motto of "Big Science, Team Science and Open Science." I thank the chief executive

officer of the Allen Institute, Allan Jones, for tolerating my scholarly pursuits. I gratefully acknowledge the Tiny Blue Dot Foundation for funding some of the consciousness-related research reported in the book.

Last but not least, I am grateful to my wife, Teresa Ward-Koch who, together with Ruby and Felix, reminds me about what is important in life and for letting me get away with so many late-night and early-morning bouts of solitary writing.

1 What Is Consciousness?

What is common between the delectable taste of a favorite food, the sharp sting of an infected tooth, the fullness after a heavy meal, the slow passage of time while waiting, the willing of a deliberate act, and the mixture of vitality, tinged with anxiety, just before a competitive event?

All are distinct experiences. What cuts across each is that all are subjective states. All are consciously felt. Accounting for the nature of consciousness appears elusive, with many claiming that it cannot be defined at all. Yet defining it is actually straightforward. Here goes:

Consciousness is experience.

That's it. Consciousness is any experience, from the most mundane to the most exalted. Some add *subjective* or *phenomenal* to the definition. For my purposes, these adjectives are redundant. Some distinguish *awareness* from *consciousness*. For reasons I've given elsewhere,[1] I don't find this distinction helpful and so I use these two words interchangeably. I also do not distinguish between *feeling* and *experience*, although in everyday use feeling is usually reserved for strong emotions, such as feeling angry or in love. As I use it, any feeling is an experience. Collectively taken, then, consciousness is lived reality. It is the feeling of life itself. It is the only bit of eternity to which I am entitled. Without experience, I would be a zombie, a nothing to myself.

To be sure, my mind has other aspects, as well. In particular, there is the vast domain of the non- and the unconscious that exists outside the limelight of consciousness. But the challenging part of the mind–body problem is consciousness, not nonconscious processing; it is that I can *see* something, *feel* something, that is mysterious, rather than how my visual system processes the rain of photons impinging onto my retina to identify a face. Any smartphone does the latter, but none can perform the former.

The seventeenth-century French physicist, mathematician, and philosopher René Descartes, in his *Discourse on the Method*, sought ultimate certainty as the foundation of all thought. He reasoned that, if he could assume that everything was open to doubt, including whether the outside world existed, and still know something, then that something would be certain. To this end, Descartes conceived of a "supremely powerful malicious deceiver" who could fool him about the existence of the world, his body, and everything he saw or felt. Yet what was not open to doubt was that he was experiencing *something*. Descartes concluded that because he was conscious, he existed. He expressed this, the most famous deduction in Western thought, in the memorable dictum:

> I think, therefore I am.[2]

More than a thousand years earlier, Saint Augustine of Hippo, one of the foundational Church Fathers, made a strikingly similar argument in his *City of God*, with the tag line, *si fallor sum*, or

> If I am mistaken, I exist.[3]

Less high-brow but closer to contemporary cyberpunk sensibility is Neo, the central character in the *Matrix* movie trilogy. Neo lives in a computer simulation, the Matrix, which looks and feels to him like the everyday "real" world. In reality, Neo's body, together with those of the rest of humanity, is stacked in gigantic warehouses, harvested as an energy source by sentient machines (a modern-day version of Descartes's malicious deceiver). Until Neo takes the red pill offered to him by Morpheus, he lives in complete denial of this reality; yet there is no doubt that Neo has conscious experiences, even though their content is completely delusional.

A different way of putting it is that *phenomenology*—what I experience and how my experiences are structured—is prior to what I can infer about the external world, including scientific laws. Consciousness is prior to physics.

Think of it this way. I see something I've learned to call a face. Face percepts follow certain regularities: they are usually left-right symmetric; they typically consist of something conventionally called a mouth, a nose, two eyes. From closely inspecting the eyes in a face, I can infer whether the face is looking at me, whether it is angry or scared and so on. I implicitly attribute these regularities to objects, called people, existing in a world outside of me; I learn how to interact with them and I infer that I am a person like

them. As I grow up, I am so utterly habituated to this inferential process that I take it completely for granted. From these experiences, I build up a picture of the world. This inferential process is amplified and acquires immense power using the intersubjective method of science that reveal hidden aspects of reality, such as electrons and gravity, exploding stars, the genetic code, dinosaurs, and so on. But ultimately, these are all inferences; eminently reasonable ones, but inferences nonetheless. All these things could prove erroneous. But not that I experience. It is the one fact I am absolutely certain of. Everything else is conjecture, including the existence of an external world.

Denying One's Experience

The great strength of this commonsense definition—*consciousness is experience*—is that it is completely obvious. What could be simpler? Consciousness is the way the world appears and feels to me (I will talk about you in the next chapter).

A minority of researchers begs to differ. To reduce the mental discomfort of being unable to explain the central aspect of life, some philosophers, such as the wife-and-husband team of Patricia and Paul Churchland, dismissively refer to the folk belief of the reality of experience as a naive assumption, just like thinking the Earth is flat, that must and shall be overcome. They seek to eliminate the very idea of consciousness from polite discussion among the educated.[4] On this view, in some real sense, no one suffers from cruelty, torture, agony, distress, depression, or anxiety. If correct, such an eliminative stance implies that if people would only realize that they are confused about the true nature of their experiences, that consciousness doesn't really exist, suffering would vanish *tout court* from the world! Utopia achieved (of course, there wouldn't be pleasure and joy either; you can't cook an omelet without breaking some eggs). To put it mildly, I find this extremely unlikely. Such a denial of the authentic nature of experience is a metaphysical counterpart to Cotard's syndrome, a psychiatric condition in which patients deny being alive.

Others, such as Daniel Dennett, argue vociferously that, although consciousness exists, there is nothing intrinsic or special about it. As he expressed it in an interview in the *New York Times*, "The elusive subjective conscious

behavior is action—which is all fine in itself, but it is totally different from my subjective perception of the scene in front of me.

These days, it is easy for image-processing software not only to store photos but also to pick out and identify faces. The algorithm extracts information from the pixels making up the image and outputs a label, such as "Mom." Yet this straightforward transformation—image in, label out—is radically different from my experience of seeing my mother. The former is an input–output transformation; the latter is a state of being.

Explaining feelings to a zombie is a much greater challenge than explaining seeing to a person born without sight. For the blind person knows about sounds, touches, loving, hating, and so on; I just have to explain that a visual experience is like an auditory experience except visual percepts are associated with blobs that move in a certain way as the eyes swivel and the head turns and whose surfaces have peculiar properties such as color and texture. The zombie, by contrast, has no percepts of any kind to compare the feeling of seeing to.

I wake up every day to a world suffused with conscious experience. As a rational being, I seek to explain the nature of this luminous feeling, who has it and who doesn't, how it arises out of physics and my body, and whether engineered systems can have it. Just because it is more difficult to define consciousness objectively than to define an electron, a gene, or a black hole, doesn't mean that I have to abandon the quest for a science of consciousness. I just have to work harder at it.

Any Experience Is Structured

Any experience has distinctions within it. That is, any experience is structured, composed of many internal phenomenal distinctions. Consider one particular visual experience (fig. 1.1). Its central focus is my Bernese mountain dog, Ruby, sitting in a chair, onto which my legs are propped. Other objects can be seen in the background. Yet that is not all; there is more, much more. There is left and right, up and down, center and periphery, closer and farther away—an uncountable number of spatial relationships. Even when I open my eyes in the complete dark, I experience a rich notion of geometric space that extends in all directions.

The actual experience, impossible to depict in a drawing, also includes Ruby's peculiar smell and the emotional coloration that shapes my attitude

toward her. These distinct sensory and affective aspects are interwoven in a complex experiential cocktail, each with its own time-course, some swift, some more sluggish, some transient, some sustained. This is true of most experiences; each can be dissected into finer distinctions across modalities.[10]

Consider another everyday experience. Squeezed into seat 36F on a bumpy, two-hour flight after having had my morning cappuccino, I feel pressure building up in my bladder. By the time I get to a bathroom in the terminal, the urge to pee becomes almost unbearable[11]—finally, I consciously feel the urine flowing, together with a mildly pleasurable sensation as the pressure is relieved. But beyond that, I can't introspect further. I can't decompose these sensations into more primitive atomic elements. I can't get past the "veil of the Maya," to adopt Hindu parlance. My introspective spade has hit impenetrable bedrock.[12] And I certainly never experience the synapses, neurons and the other stuff inside my skull that constitutes the physical substrate of any experience. That level is completely hidden to me.

Finally, consider a rare class of conscious states: mystical experiences common to many religious traditions, whether Christian, Jewish, Buddhist, or Hindu. These are characterized as having no content: no sounds, no images, no bodily feelings, no memories, no fear, no desire, no ego, no distinction between the experiencer and the experience, the apprehender and the apprehended (nondual).

The late-medieval Dominican monastic, philosopher, and mystic Meister Eckhart encountered the Godhead in a featureless plain, the essence of his soul:

> There is the silent "middle," for no creature ever entered there and no image, nor has the soul there either activity or understanding, therefor she is not aware there of any image, whether of herself or of any other creature.[13]

Using similar language, long-term practitioners of Buddhist meditation describe naked or sheer awareness:

> Unobscured like a cloudless sky, remain in lucid and intangible openness. Unmoving like the ocean free of waves, remain in complete ease, undistracted by thought. Unchanging and brilliant like a flame undisturbed by the wind, remain utterly clean and bright.[14]

I shall return in chapter 10 to content-free or pure consciousness, as this phenomenon constitutes a striking challenge to any computational account of consciousness. Note that even pure experience is, strictly speaking, a subset (though not a proper one) of the whole and is therefore structured.

Beyond the intrinsic and structured nature of any one conscious experience, what else do I know for certain about my experience? What can I positively say that is true for any experience, no matter how mundane or how exotic?

Any Experience Is Informative, Integrated, and Definite

Three additional properties hold for any conscious experience. They cannot be doubted.

First, any experience is highly *informative*, distinct because of the way it is. Each experience is informationally rich, containing a great deal of detail, a composition of specific phenomenal distinctions, bound together in specific ways. Every frame of every movie I ever saw or will see in the future is a distinct experience, each one a wealth of phenomenology of colors, shapes, lines, and textures at locations throughout the field of view. And then there are auditory, olfactory, tactile, sexual, and other bodily experiences—each one distinct in its own way. There cannot be a generic experience. Even the experience of vaguely seeing something in a dense fog, without being clear what I am seeing, is a specific experience.

I recently attended a Blind Café during which I underwent a sort of a reverse birth. I shuffled from a lit vestibule through a long, black, narrow birth-canal into a completely dark chamber—so dark that I was unable to see my wife's hand that she waved in front of me. We groped for chairs, sat down, introduced ourselves to the other guests and started to eat in Stygian darkness—very, very carefully. It was an utterly unique experience designed to introduce sighted folks to the world of the blind. Yet even in this pitch-black room I had a distinct visual experience, specific, and, combined with its echoes and its feels, distinct from waking up in a pitch-dark hotel room.

Second, any experience is *integrated*, irreducible to its independent components. Each experience is unitary, holistic, including all phenomenal distinctions and relations within that experience. I experience the entire drawing, including my body on the couch and the room, not just the legs and, independently, the hand. I don't experience the left side independently of the right side or the dog divorced from the lounge chair on which she is squatting. I experience the whole thing. When somebody tells me about their honeymoon, I have a distinct image of the couple going off on

a romantic get-away rather than imagining the sweet substance produced by bees in addition to the large object in the sky.[15]

Third, any experience is *definite* in content and spatiotemporal grain. It is unmistakable. Looking again at the domestic scene in figure 1.1, I perceive my dog and the world in chiaroscuro, in perspective, from the sofa, with my right eye shut. There is a distinct content of consciousness that is "in" while everything else is out, not experienced. The world I see isn't bordered by a line beyond which things are gray or dark, such as behind my head. It simply doesn't exist. The strokes of the brush are painted onto the canvas; everything else is not.

My experience is what it is with a definite content. If it were anything more (seeing while experiencing a pounding headache, say) or anything less (like the drawing but without dog), it would be a different experience.

In summary, every conscious experience has five distinct and undeniable properties: each one exists for itself, is structured, informative, integrated and definite. These are the five essential hallmarks of any and all conscious experiences, from the commonplace to the exalted, from the painful to the orgiastic.

Any Experience Has a Point of View and Occurs in Time

Some researchers argue that experiences may have other properties in addition to these five; for example, that each experience comes with a unique point of view—a first-person account, the subject's perspective. I am looking at the drawing; I am at the center of this world.[16] I suspect that centeredness emerges from the representation of space as it is given to me by my visual, auditory, and tactile senses. Each one of these three associated sensory spaces has one particular location singled out, which is where the eyes, ears, and my body respectively are located. As it is obviously important that what I see, what I hear, and what I feel all refer to a common space (so that, for example, the sounds I hear emanating from moving lips are assigned to the colocalized face), "I" am located at this singular point, the origin of my own space. Furthermore, this center is also the focus of any behaviors, such as moving the eyes with its attendant shift in perspective. Thus, having a perspective, a view from somewhere rather than from nowhere, emerges in a natural way from the structure of sensorimotor contingencies, without having to postulate any additional fundamental property.

A more compelling case could be made that any one experience takes place at a particular moment, the present *now*. Defining this now in an objective manner has defied philosophers, physicists, and psychologists since time immemorial. There is no doubt that lived life has three distinct temporal dominions: the past, the present, and the future, with the experienced present being the intermediate between the past and the future.[17] The past encompasses everything that has already happened. It is immutable, even though how I recall events within my memory palace is susceptible to reinterpretations and to subsequent occurrences that seemingly violate causality. The future is the sum total of everything that hasn't happened yet; it is open-ended and contingent. The bleeding edge of the future forever turns into the specious present that irrevocably recedes into the past as soon as it is experienced.

Yet there are uncommon experiences in which the perception of time ceases. For those taking hallucinogens, for example, the flow of the river of time, the duration of the present now, can slow down and even stop altogether. Time crawls when one's attention is utterly and fully engaged, such as during a dangerous climb up a sheer wall of granite. Movies like *The Matrix* visualize this slowing of perceived time through the well-known bullet-time effect. In other words, the flow of time is not a universal property of all experiences, but only of most.[18]

Thus, what remains is the quintet of essential properties that any and all conscious experiences have:

> Every conscious experience exists for itself, is structured, is the specific way it is, is one, and is definite.[19]

So that's how it is for me. How is it for you? What can I confidently state about the experiences of others? How can their experiences be studied in the laboratory? I cover these questions in the next chapter.

ecosystems. Abduction is a form of reasoning that deals with probabilities and likelihoods. The conclusion of a solid abductive argument is a hypothesis that best explains all known facts. We daily abduce the best explanation of a dizzying variety of phenomena—diagnosing the most likely cause of a skin rash, a malfunctioning car, a leaking pipe, a financial or political crisis.

The search for the most likely explanation of all relevant facts is the very opposite mindset of that of conspiracy-minded folks who perceive the malevolent action of their particular Boogieman behind every event (the CIA, Jews, communists). This leads to a contrived, byzantine chain of reasoning, involving the collusion of thousands of individuals, extremely unlikely to have taken place. Sightings of the Virgin Mary in a cheese sandwich, gigantic alien face artifacts on Mars, and the moon-landing conspiracy are all sorry examples of breakdowns of inference to the best explanation.[2]

Sherlock Holmes is a master of abductive reasoning, with the BBC series *Sherlock* visualizing his inferences with vivid graphic overlays. Despite Holmes's claim that he's practicing a "science of deduction," he rarely deduces anything—for that would imply logical necessity. From the two propositions "All men are mortal" and "Socrates is a man," we can deduce by necessity that Socrates will die. In real life, the situation is never that clear. Typically, Holmes abduces the most likely explanation of the facts, as in his celebrated exchange with the police in the short story "Silver Blaze":

> Inspector Gregory: "Is there any point to which you would wish to draw my attention?" Holmes: "To the curious incident of the dog in the night-time." Inspector: "The dog did nothing in the night-time." Holmes: "That was the curious incident."

Holmes abduced that the dog didn't bark because it knew the perpetrator. Abductive reasoning is all the rage in computer science and artificial intelligence, giving software powerful reasoning abilities. An example is IBM's question-and-answering computer system using natural language, Watson, employed in medical diagnostics.[3]

Probing the Conscious Minds of Others

Unlike my own mind, which I am directly acquainted with, I can only abduce the existence of other conscious minds. I can never directly experience them. In particular, I abduce that you and other people have experiences like I do unless I have strong reasons to believe otherwise (say, because they have a brain injury or are severely intoxicated). With this assumption in place,

I can look for systematic connections between consciousness and the physical world.

Psychophysics (literally the "physics of the soul") is the science that seeks to elucidate quantitative relationships between stimuli—a tone, a spoken word, a color field, a picture flashed onto a screen, a heated probe to the skin—and their elicited experiences. A branch of psychology, psychophysics has uncovered reliable, consistent, reproducible, and lawful regularities between objective stimuli and subjective reports.[4]

Although this chapter focuses on seeing, perception is a broad term that includes the traditional five sensory faculties—sight, sound, smell, touch and taste—as well as pain, balance, heartbeat, nausea, and other epigastric sensations.

To quantify phenomenology under laboratory conditions, psychologists don't rely on flowery descriptions. Instead, they ask simple questions. Lots of them. In a typical experiment, volunteers, paid for their time and effort, will stare at a screen while a picture, for instance, a barely visible face or a butterfly superimposed onto a grainy background of checkered light and shadow, is flashed onto a screen. Immediately afterward, they are confronted by the question, "Did you see a face or a butterfly?" (fig. 2.1). Only two answers are permitted, "face" or "butterfly." Not "I'm not sure I saw much" or "Sorry, I don't know." When in doubt, guess.

In practice, subjects don't speak but push buttons on a keyboard, allowing for consistent and rapid action. In this manner, researchers can quickly collect responses from several hundred trials. Pressing a button also tracks the subject's reaction time, which can be mined for further insights.

The experience is reduced to a series of button presses. Averaged over a block of individual trials, such responses are an *objective measure of perception,* as there is a correct answer available to the researchers (since they have access to the computer program that generated the face or butterfly image and so know the right answer). That is, a proverbial third person knows if what the subject reported corresponds to what was on the screen.

While the timing of the button press is easy enough to measure, the swiftness of vision is more difficult to determine. Comparing the timing of EEG signals from your brain when you see the face to the timing when you didn't (as in figure 2.1) indicates that visual experience arises as early as 150 msec or as late as 350 msec after the stimulus enters your eyes.[5]

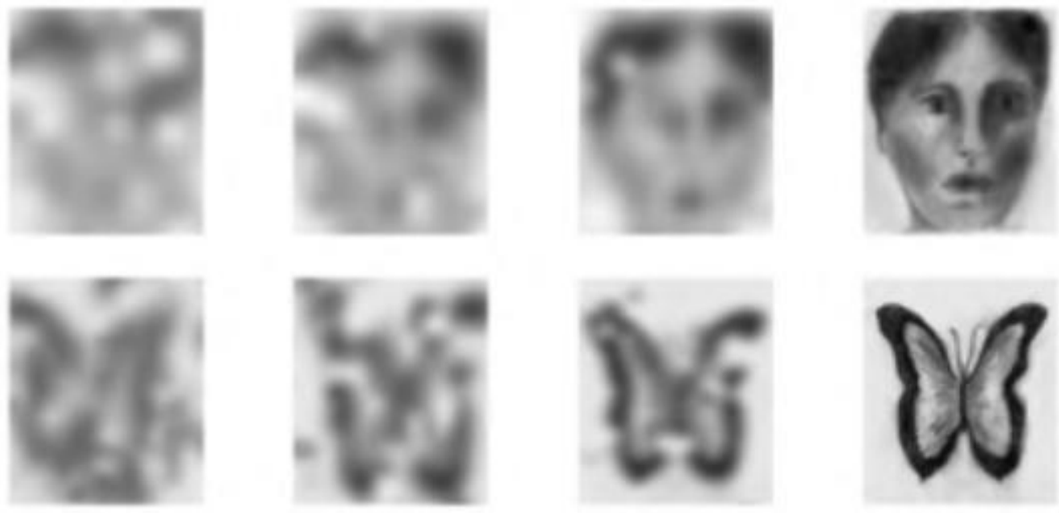

Figure 2.1
Probing experience: Images of faces or butterflies are made easy or more difficult to recognize by overlaying visual noise while you push a button, indicating whether you saw a face or a butterfly. For any one level of such noise, trials when the stimulus is correctly perceived are compared to those when you failed to do so, even though the same picture was present on your retina. (Adapted from Genetti et al., 2011.)

The visibility of the figure is manipulated to make it easier or more difficult to discern. When images are flashed for a mere 1/60th of a second, the perceptual judgment of subjects can vary considerably from trial to trial. Consider the 50 percent face image of figure 2.1 flashed briefly onto the screen. You might respond "face" three times but then on the fourth trial push the butterfly button. As the object emerges from its noisy background to become more visible (75 percent or 100 percent images), you are more and more likely to respond correctly, until you do so on almost every trial. There is a steady progression from not being able to make the distinction, to doing better than chance, to getting it right each time.[6]

Repeating this experiment with many subjects results in similar though not identical response rates as a function of visibility. Responses don't depend much on which pictures are used, butterflies or, say, pictures of animals

or houses. This is reassuring and reinforces my prior "We're all conscious" assumption.

Perceptual research of this sort reveals that sensory perception is not a passive reflection or a simple mapping of the outer world onto an inner mental screen. Perception is an active process, "a construction of a description of the world," as the influential theoretician David Marr argued.[7] You are intimately familiar with this world, because it is the one you see, hear, and otherwise experience. You infer this world from the data impinging upon your eyes, ears, and other receptors by sophisticated but unconscious processes. That is, you don't look at the world and say to yourself, "Hmm, that surface reflects light in this way and is occluding another surface, and there is a shadow falling on the first surface cast by another surface far away, with a bright light source on the upper right." No, you look and see a bunch of people under a bright Harvest moon, partially obscuring each other. All of this is inferred based on the available retinal information, as well as your previous visual experiences and those of your ancestors (encoded in your genes).

Perception is a construction of precisely those features that are useful in our struggle to survive a world of eat-or-be-eaten.

How perception happens is hidden from your conscious mind. You simply look and see. Indeed, I still remember when, many decades ago, I tried to explain to my parents—a doctor and a diplomat—why I study vision. They couldn't see the point, as it was so obviously trivial. Similarly, the myriad of software operations underlying even basic tasks on your computer are completely hidden from you behind the simplicity of the user interface.

Visual illusions reveal the sometimes striking dissonance between appearances and reality. Consider the "Lilac chaser," which has its own Wikipedia page (https://en.wikipedia.org/wiki/Lilac_chaser). When you keep your eyes steady on the central fixation cross, you'll see a single greenish disk traveling on a circular trajectory, round and round. Yet the actual stimuli are eleven pink disks on a circle, with the twelfth disk missing in one location; the location of this missing disk, essentially a hole, travels around the circle. What you see is not there, and what is out there on the screen is not what you see!

The Lilac chaser is fool-proof. Even though you know it is an illusion, you can't break it. It is an extreme example of the difference between your perception of the external world and its actual metric properties (size,

distance, and so on). Most of the time the conflict between appearance and reality is minor and inconsequential. In that sense, perception is by and large reliable. But at times, the discrepancy can be striking, demonstrating the limits of perception. Even mind-enhancing drugs will not let you escape the cage that is your brain—the world-in-itself, Kant's storied *Ding an sich*, is never directly accessible.

I am an avid rock climber, in search of that peculiar combination of intense fear and exhilaration encountered on the high crag, when time melts into a taut presence. Recently, I was on a narrow rock ledge, high up on a mountain side; it was snowing, with a high wind blowing. I had to cross a chasm on a wooden rope bridge, one side of which was badly frayed. With both feet in full contact with the plank, I slowly and deliberately shuffled across, in control but for the slight shaking of my calf muscles, the "Elvis" or "sewing machine leg" familiar to climbers. Midway between the two walls, over the abyss, I forced myself to look down, at the riverbed far below, before moving on to the relative safety of the narrow ledge on the other side.

Yet in reality, and embarrassingly, I was walking across a wooden stud on the carpeted floor of an office, wearing immersive virtual reality goggles! My visual experience of the cavernous spaces around and below me, my sense of being there, the sound of the wind in my ears—all of it induced a palpable sense of arousal and tension. The abstract knowledge that I was safe did not eliminate the sense of danger I experienced. A visceral demonstration of the limits of perception.

Plumbing the Depths of Consciousness

Psychophysics explores the relation between first-person experiences and third-person objective measures, such as response rates. For some, though, this isn't good enough. They argue that objective measures don't truly capture the subjective nature of an experience. To get closer to the actual phenomenology, psychologists invented *subjective measures*, probing what people know about their experience, a simple form of self-consciousness.

Recall the experiment in which a degraded picture is flashed onto the screen. After you pressed the "face" or "butterfly" button, you are asked to reflect on this button press and indicate how confident you are about your answer. This could take the form of a four-point confidence scale, with a 1 indicating "I'm guessing," a 2 "I may have seen a face," a 3 "I think I saw

your eyes move; you just can't see your own. Your brain suppresses these short segments of what would look blurry and replaces them by splicing in a stationary scene, like in a movie studio. The same is true of the eye *blinks* you make every couple of seconds (this editing-out doesn't occur for voluntary *winks*). All this furious editing goes unnoticed; you see a steady world as you look around.

Given that you make more than 100,000 daily saccades, each one lasting between 20 and 100 milliseconds, saccadic and blink suppression adds up to more than an hour a day during which you are effectively blind! Yet until scientists started studying eye movements, no one was aware of this remarkable fact.

Eye movements are but one instance of a sophisticated set of processes, implemented by specialized brain circuits, that make up a lived life. Neurological and psychological sleuthing has uncovered a menagerie of such specialized processes. Hitched to the eyes, ears, the equilibrium organ, and other sensors, these servomechanisms control our eyes, neck, trunk, arms, hands, fingers, legs, and feet. They are responsible for everyday actions—shaving, washing, tying shoelaces, biking to work, typing on a computer keyboard, texting on a phone, playing soccer, and on and on. Francis Crick and I called these specialized sensory-cognitive-motor routines *zombie agents*.[13] They manage the fluid and rapid interplay of muscles and nerves that is at the heart of all skills. They resemble *reflexes*—blinking, coughing, jerking your hand away from a hot stove, or being startled by a sudden loud noise. Classical reflexes are automatic, fast, and depend on circuits in the spinal cord or in the brainstem. Zombie behaviors can be thought of as more flexible and adaptive reflexes that involve the forebrain.

Saccadic eye movements are controlled by such a zombie agent while bypassing consciousness. You can become conscious of the routine action of a zombie agent, but only after the fact. While I was trail running in the mountains in Southern California, something made me look down. My right leg instantly lengthened its stride, for my brain had detected a rattlesnake sunning itself on the stony path where I was about to put my foot. Before I had consciously seen the reptile, before I had experienced the attendant adrenaline rush, and before the snake gave its ominous warning rattle—I had avoided stepping on it and sped past. If I had depended on the conscious feeling of fear to control my legs, I would have trod on the snake. Experiments confirm that motor action can, indeed, be faster than thought,

with the onset of corrective motor action preceding conscious perception by about a quarter of a second. Likewise, consider a world-class sprinter running one hundred meters in ten seconds. By the time the runner consciously hears the pistol, he is already several strides out of the starting block.

Learning a new sport—playing tennis, sailing, sculling, or mountaineering—takes a great deal of both physical and mental discipline. In climbing, the novice learns where to place her hands, feet, and body to smear, stem, lie back, lock off the wrist or the fingers in a crack. The climber pays attention to the flakes and grooves that turn a vertical granite cliff into a climbable wall with holds and learns to ignore the void beneath her. A sequence of distinct sensory-cognitive-motor actions is stitched into a smoothly executing motor program. After hundreds of hours of intense and dedicated training these labors result in thoughtless, flawless flow, a divine experience. Constant repetition recruits specialized brain circuits, colloquially referred to as *muscle memory*, that render the skill effortless, the motion of the body fluid, without wasted effort. The expert climber never gives a thought to the minutiae of her actions requiring a marvelous merging of muscle and nerve.

Indeed, much of what goes on in unexamined life is not accessible to consciousness or bypasses it altogether. A widespread phenomenon is *mind blanking*: the mind is seemingly nowhere while the body carries on with the routine aspects of everyday living.[14] Virginia Woolf, an astute observer of the inner self, had this to say:

> Often when I have been writing one of my so-called novels I have been baffled by this same problem; that is, how to describe what I call in my private shorthand "non-being." Every day includes much more non-being than being. ... Although it was a good day the goodness was embedded in a kind of nondescript cotton wool. This is always so. A great part of every day is not lived consciously. One walks, eats, sees things, deals with what has to be done; the broken vacuum cleaner. ... When it is a bad day the proportion of non-being is much larger.[15]

Just as you can never catch the light inside your refrigerator turned off, because every time you open the door of the fridge the light is on, you can't experience not experiencing. By randomly pinging subjects on their smartphones to ask what they were aware of at that precise point in time, psychologists discovered that the blank mind, defined by no experience at all (quite the opposite of pure experience mentioned in the previous chapter), is common over the course of a day at the office, doing chores at home

or while at the gym, driving, or watching TV. Mindfulness, "being in the moment," counteracts the blank mind.

Somewhere in the brain, the body is monitored; love, joy, and fear are born; thoughts arise, are mulled over, and discarded; plans are made; memories are laid down. Yet the conscious self may be turned off or is oblivious to this furious activity. You are a stranger to your mind.

That most of the operations of the mind are inaccessible to consciousness should not be surprising. After all, you don't feel your liver metabolizing the alcohol in the pinot noir from last night, you don't experience the trillions of bacteria happily colonizing your intestine, and you are deaf to your immune system fighting off some bug.

The nonconscious is a bound of the mind that lay undiscovered until philosophers and psychologists—in particular, Friedrich Nietzsche, Sigmund Freud, and Pierre Janet—inferred its existence in the late 1800s. Its remoteness reflects our deeply felt intuition that the conscious mind is all there is. It also explains why so much of philosophy of mind has been barren. You can't introspect your way to the unconscious layers of your mind. Yogi Berra might have quipped "You don't know what you don't experience."

The existence of the nonconscious throws the question of the physical basis of consciousness into stark relief. What is the difference between unconscious and conscious actions of the mind?

On the Limits of Behavioral Methods

You might think that scientists wouldn't touch any measure characterized as subjective with a ten feet pole. Yet subjective doesn't mean arbitrary. Subjective measures follow well-established regularities that can be checked. As a rule, as stimulus duration or the strength of the central object relative to its background decreases (fig. 2.1), both objective response rate and subjective confidence decrease and reaction time lengthens—the less certain you are about what you experienced, the slower you respond. In other words, the first-person perspective can be validated by third-person measures.

It is good scientific practice to assume that volunteers may not always faithfully carry out the instructions of the experimentalist—either because they can't follow instructions (as is the case with babies and infants, who require special techniques), misunderstand them, or don't want to follow

them (because subjects are bored and press buttons at random or want to cheat). It is critical to design appropriate controls; adding catch trials where the answer is known, repeating experiments to check for consistency, and cross-validating with other data to keep such inappropriate responses to a minimum.

There are subjects, however, who are as isolated as any early twentieth-century polar explorer wintering in the arctic—patients with severe disorders of consciousness following traumatic brain injury, encephalitis, meningitis, stroke, drug or alcohol intoxication, or cardiac arrest. Disabled and bedridden, they are unable to talk about or otherwise signal their mental state. Unlike comatose patients who have few reflexes and lie immobile, in a profound state of unconsciousness, vegetative-state patients cycle through periods of eyes opening and closing, resembling sleep (without necessarily having sleep-associated brain-wave activity).[16] They may move their limbs reflexively, grimace, turn their head, groan, spasmodically move their hands. To the naive bedside observer, these movements and sounds suggest that the patient is awake, desperately trying to communicate with his or her loved ones.

Consider Terri Schiavo, the woman in Florida who lingered for fifteen years in a vegetative state until her medically induced death in 2005. Given the public fight between her husband, who advocated discontinuing her life support, and her devout parents, who believed that their daughter had some measure of awareness, the case caused a huge uproar. It was litigated up and down the judicial chain and eventually landed on the desk of then-president George W. Bush. Her husband ultimately prevailed in his wish to have his wife taken off life support.[17]

Properly diagnosing vegetative-state patients is challenging. Who can say with certainty whether these patients experience pain and distress, living in the gray zone between fleeting consciousness and nothingness? Fortunately, neurotechnology is coming to the rescue of such patients, as I shall detail in chapter 9.

Thus, the absence of reproducible, willful behavior is not always a sure sign of unconsciousness. Conversely, the presence of some behaviors is likewise not always a sure sign of consciousness. A variety of reflex-like behaviors—eye movements, posture adjustments, or mumbling in one's sleep—bypass consciousness. Sleepwalkers are capable of complex, stereotyped behaviors—moving

about, dressing and undressing, and so on, without subsequent recall or other evidence of awareness.[18]

So yes—behavioral methods of inferring experience in others do have limitations; but even these limitations can be studied objectively. And as science's understanding of consciousness grows, the frontier between the known and the unknown is constantly being pushed back.

So far, I have only spoken about people and their experiences. What about animals? Do they too see, hear, smell, love, fear, and grieve?

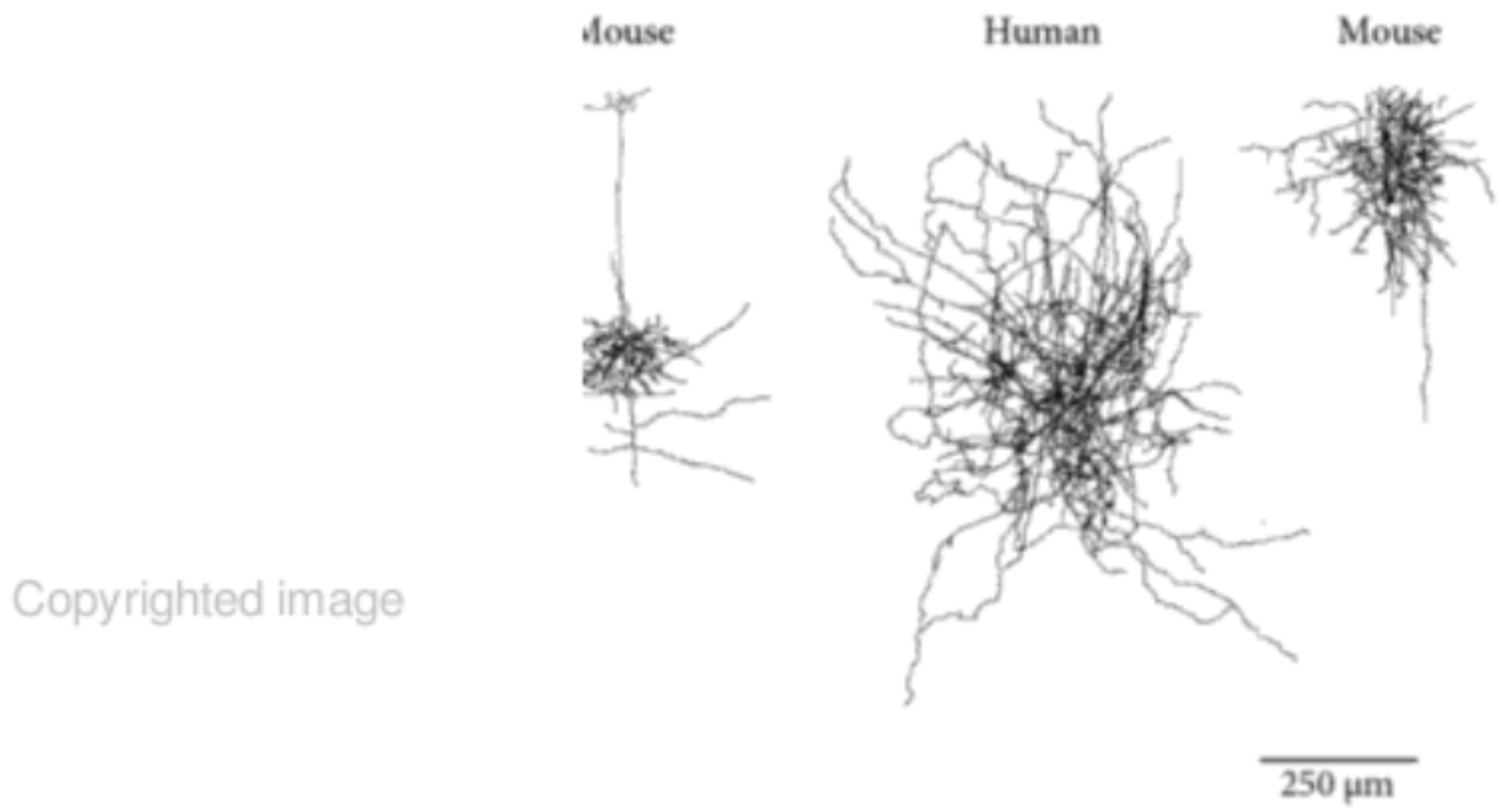

Figure 3.1

Mouse and human neurons: Two human and two mouse neocortical neurons from the Allen Institute of Brain Science. Their morphologies are similar, except that human cells are taller. (Data provided by Staci Sorensen at the Allen Institute.)

Second, the architecture of the nervous system is remarkably conserved across all mammals. Most of the close to nine hundred distinct annotated macroscopic structures found in the human brain are present in the mouse brain, the animal of choice for experimentalists, even though it is a thousand times smaller.[5]

It is not easy, even for a neuroanatomist armed with a microscope, to distinguish human nerve cells from their murine counterpart, once the scale bar has been removed (fig. 3.1).[6] That's not to say human neurons are the same as mouse neurons—they are not; the former are more complex, have more dendritic spines, and look to be more diverse than the latter. The same story holds at the genomic, synaptic, cellular, connectional, and architectural levels—we see a myriad of quantitative but no qualitative differences between the brains of mice, dogs, monkeys, and people. The receptors and pathways that mediate pain are analogous across species.

The human brain is big, but other creatures, such as elephants, dolphins, and whales, have bigger ones. Embarrassingly, some not only have a larger neocortex but also one with twice as many cortical neurons as humans.[7]

Third, the behavior of mammals is kindred to that of people. Take Ruby—she loves to lick the remaining cream off the whisk I use to whip heavy cream by hand—no matter where she is in the house or garden she comes running in as soon as she hears the sounds of the metal wire loops striking the glass. Her behavior tells me that she enjoys the sweet and fatty whipping cream as much as I do; I infer that she has a pleasurable experience. Or when she yelps, whines, gnaws at her paw, limps, and then comes to me, seeking aid: I infer that she's in pain because under similar conditions I act similarly (sans gnawing). Physiologic measures confirm this inference: dogs, just like people, have an elevated heart rate and blood pressure and release stress hormones into their bloodstream when in pain. Dogs not only experience pain from physical injuries but can also suffer, for example if they are beaten or otherwise abused or when an older pet is separated from its litter mate or its human companion. This is not to argue that dog-pain is identical to people-pain; it is not. But all the evidence is compatible with the supposition that dogs, and other mammals, not only react to noxious stimuli but also experience the awfulness of pain and suffering.

As I write this chapter, the world is witnessing a killer whale carrying her baby calf, born dead, for more than two weeks and a thousand miles across the waters of the Pacific Northwest. As the corpse of the baby orca keeps on falling off and sinking, the mother has to expend considerable energy to dive after it and retrieve it, an astonishing display of maternal grief.[8]

Monkeys, dogs, cats, horses, donkeys, rats, mice, and other mammals can all be taught to respond to forced-choice experiments of the sort outlined earlier—modified from those used by people to accommodate paws and snouts, and using food or social rewards in lieu of money. Their responses are remarkably similar to the way people behave, once differences in their sensory organs are accounted for.[9]

Experience without Voice

The most obvious trait that distinguishes humans from other animals is language. Everyday speech represents and communicates abstract symbols and concepts. It is the bedrock of mathematics, science, civilization, and all of our cultural accomplishments.

Many classical scholars assign to language the role of kingmaker when it comes to consciousness. That is, language use is thought to either directly

enable consciousness or to be one of the signature behaviors associated with consciousness. This draws a bright line between animals and people. On the far shore of this Rubicon live all creatures, small and large—bees, squids, dogs, and apes; while they have many of the same behavioral and neuronal manifestations of seeing, hearing, smelling, and experiencing pain and pleasure that people have, they have no feelings. They are mere biological machines, devoid of any inner light. On the near shore of this Rubicon lives a sole extant species, *Homo sapiens*.[10] Somehow, the same sort of biological stuff that makes up the brains of creatures across the river is superadded with sentience (Descartes's *res cogitans* or the Christian soul) on this side of the Rubicon.

One of the few remaining contemporary psychologists who denies the evolutionary continuity of consciousness is Euan Macphail. He avers that language and a sense of self are necessary for consciousness. According to him, neither animals nor young children experience anything, as they are unable to speak and have no sense of self—a remarkable conclusion that must endear him to parents and pediatric anesthesiologists everywhere.[11]

What does the evidence suggest? What happens if somebody loses their ability to speak? How does this affect their thinking, sense of self, and their conscious experience of the world? *Aphasia* is the name given to language disorders caused by limited brain damage, usually but not always to the left cortical hemisphere. There are different forms of aphasia—depending on the location of the damage, it can affect the comprehension of speech or of written text, the ability to properly name objects, the production of speech, its grammar, the severity of the deficit, and so on.[12]

The neuroanatomist Jill Bolte Taylor rocketed to fame on the strength of her TED Talk and a subsequent bestselling book about her experience while suffering a stroke.[13] At age thirty-seven, she suffered a massive bleeding in her left hemisphere. For the next several hours, she became effectively mute. She also lost her inner speech, the unvoiced monologue that accompanies us everywhere, and her right hand became paralyzed. Taylor realized that her verbal utterances did not make any sense and that she couldn't understand the gibberish of others. She vividly recalls how she perceived the world in images while experiencing the direct effect of her stroke, wondering how to communicate with people. Hardly the actions of an unconscious zombie.

Two objections to Taylor's compelling personal story is that her narrative can't be directly verified—she suffered the stroke at home, alone—and that