

Chapter One

Stepping into the Light

Stepping into the Light

I have always been intrigued by the specific moment when, as we sit waiting in the audience, the door to the stage opens and a performer steps into the light; or, to take the other perspective, the moment when a performer who waits in semidarkness sees the same door open, revealing the lights, the stage, and the audience.

I realized some years ago that the moving quality of this moment, whichever point of view one takes, comes from its embodiment of an instance of birth, of passage through a threshold that separates a protected but limiting shelter from the possibility and risk of a world beyond and ahead. As I prepare to introduce this book, however, and as I reflect on what I have written, I sense that stepping into the light is also a powerful metaphor for consciousness, for the birth of the knowing mind, for the simple and yet momentous coming of the sense of self into the world of the mental. How we step into the light

of consciousness is precisely the topic of this book. I write about the sense of self and about the transition from innocence and ignorance to knowingness and selfness. My specific goal is to consider the biological circumstances that permit this critical transition.

No aspect of the human mind is easy to investigate, and for those who wish to understand the biological underpinnings of the mind, consciousness is generally regarded as the towering problem, in spite of the fact that the definition of the problem may vary considerably from investigator to investigator. If elucidating mind is the last frontier of the life sciences, consciousness often seems like the last mystery in the elucidation of the mind. Some regard it as insoluble.

Yet, it is difficult to think of a more seductive challenge for reflection and investigation. The matter of mind, in general, and of consciousness, in particular, allows humans to exercise, to the vanishing point, the desire for understanding and the appetite for wonderment at their own nature that Aristotle recognized as so distinctively human. What could be more difficult to know than to know how we know? What could be more dizzying than to realize that it is our having consciousness which makes possible and even inevitable our questions about consciousness?

Although I do not see consciousness as the pinnacle of biological evolution, I see it as a turning point in the long history of life. Even when we resort to the simple and standard dictionary definition of consciousness—as an organism's awareness of its own self and surroundings—it is easy to envision how consciousness is likely to have opened the way in human evolution to a new order of creations not possible without it: conscience, religion, social and political organizations, the arts, the sciences, and technology. Perhaps even more compellingly, consciousness is the critical biological function that allows us to know sorrow or know joy, to know suffering or know pleasure, to sense embarrassment or pride, to grieve for lost love or lost life. Whether individually experienced or observed, pathos is a by-product of consciousness and so is desire. None of those personal states would ever be known to each of us without consciousness. Do not blame Eve for knowing; blame consciousness, and thank it, too.

I write this in downtown Stockholm as I look out of a window and watch a frail old man make his way toward a ferry that is about to depart. Time is short, but his gait is slow; his steps break at the ankle from arthritic pain; his hair is white; his coat is worn. It is raining persistently and the wind makes him bend slightly like a lone tree in an open field. He finally reaches the ship. He climbs with difficulty the tall step needed to get on the gangplank and starts on his way down to the deck, afraid of gaining too much momentum on the incline, head moving briskly, left and right, checking his surroundings and seeking reassurance, his whole body seemingly saying, Is this it? Am I in the right place? Where to next? And then the two men on deck help him steady his last step, ease him into the cabin with warm gestures, and he seems to be safely where he should be. My worry is over. The ship departs.

Now let your mind wander and consider that, without consciousness, the old man's discomfort, perhaps humiliation, would simply not have been known to him. Without consciousness, the two men on deck would not have responded with empathy. Without consciousness, I would not have been concerned and would never have thought that one day I might be him, walking with the same pained hesitation and feeling the same discomfort. Consciousness amplifies the impact of these feelings in the minds of the characters in this scene.

Consciousness is, in effect, the key to a life examined, for better and for worse, our beginner's permit into knowing all about the hunger, the thirst, the sex, the tears, the laughter, the kicks, the punches, the flow of images we call thought, the feelings, the words, the stories, the beliefs, the music and the poetry, the happiness and the ecstasy. At its simplest and most basic level, consciousness lets us recognize an irresistible urge to stay alive and develop a concern for the self. At its most complex and elaborate level, consciousness helps us develop a concern for other selves and improve the art of life.

Absent without Leave

Thirty-two years ago, a man sat across from me in a strange, entirely circular, gray-painted examining room. The afternoon sun was shining

on us through a skylight as we talked quietly. Suddenly the man stopped, in midsentence, and his face lost animation; his mouth froze, still open, and his eyes became vacuously fixed on some point on the wall behind me. For a few seconds he remained motionless. I spoke his name but there was no reply. Then he began to move a little, he smacked his lips, his eyes shifted to the table between us, he seemed to see a cup of coffee and a small metal vase of flowers; he must have, because he picked up the cup and drank from it. I spoke to him again and again he did not reply. He touched the vase. I asked him what was going on, and he did not reply, his face had no expression. He did not look at me. Now, he rose to his feet and I was nervous; I did not know what to expect. I called his name and he did not reply. When would this end? Now he turned around and walked slowly to the door. I got up and called him again. He stopped, he looked at me, and some expression returned to his face—he looked perplexed. I called him again, and he said, "What?"

For a brief period, which seemed like ages, this man suffered from an impairment of consciousness. Neurologically speaking, he had an absence seizure followed by an absence automatism, two among the many manifestations of epilepsy, a condition caused by brain dysfunction. This was not my first exposure to impaired consciousness but it was the most intriguing yet. From a first-person perspective, I knew what it was like to dissolve into unsolicited unknowingness and to return to consciousness—I had lost consciousness once, as a kid, in an accident, and I had general anesthesia once, as an adolescent. I also had seen patients in coma and observed, from a third-person perspective, what a state of unconsciousness looked like. In all of these instances, however, as well as in falling asleep or waking up, the loss of consciousness was radical, something like a complete power outage. But what I had just seen that afternoon in the gray circular room was far more startling. The man had not collapsed on the floor, comatose, and had not gone to sleep, either. He was both there and not there, certainly awake, attentive in part, behaving for sure, bodily present but personally unaccounted for, absent without leave.

This episode stayed with me and it was a good day when I felt could interpret its meaning. I did not think then, but I think now that I had witnessed the razor-sharp transition between a fully conscious mind and a mind deprived of the sense of self. During the period of impaired consciousness, the man's wakefulness, his basic ability to attend to objects, and his capacity to navigate in space had been preserved. The essence of his mental process was probably retained, as far as the objects in his surroundings were concerned, but his sense of self and knowing had been suspended. The shaping of my notion of consciousness probably began that day, without my noticing it, and the idea that a sense of self was an indispensable part of the conscious mind only gained strength as I saw comparable cases.

I maintained an interest in the issue of consciousness through the years, at once attracted to the scientific challenge posed by consciousness and repulsed by the human consequences of its impairment in neurological disease, but I remained distant. The drama of the situations in which brain damage causes coma or persistent vegetative state the conditions in which consciousness is most radically impaired, is something I would have preferred not to observe, if given a choice. Few things are as sad to watch as the sudden and forced disappearance of the conscious mind in someone who remains alive, and few things are as painful to explain to a family. How does one look a person in the eye and make clear that the quiet state of a lifetime's companion may appear like sleep but is not sleep; that there is nothing benign or restorative about this way of resting; that the once-sentient being may never return to sentience? But even if my life as a neurologist had not made me wary of consciousness, my life as a neuroscientist might have ensured I did not touch the problem. Studying consciousness was simply not the thing to do before you made tenure, and even after you did it was looked upon with suspicion. Only in recent years has consciousness become a somewhat safer topic of scientific inquiry.[1]

Still, the reason why I eventually turned to consciousness had little to do with the sociology of consciousness studies. I certainly had not planned on investigating consciousness until an impasse forced me to

do so. The impasse had to do with my work on the emotions, and that means I can blame the consequences on the passions of the soul.[2]

So here is the situation. I could understand reasonably well how different emotions were induced in the brain and played out in the theater of the body. I could also envision how both the induction of emotions and the consequent bodily changes that largely constitute an emotional state were signaled in several brain structures appropriate to map such changes, thus constituting the substrate for feeling an emotion. But, for the life of me, I could not understand how that brain substrate of feeling could become *known* to the organism having the emotion. I could not devise a satisfactory explanation for how what we conscious creatures call feeling becomes known to the feeling organism. By which additional mechanism do each of us know that a feeling is occurring within the bounds of our own organism? What else happens in the organism and, especially, what else happens in the brain, when we know that we feel an emotion or feel pain or, for that matter, when we know anything at all? I had come up against the obstacle of consciousness. Specifically, I had come up against the obstacle of self, for something like a sense of self was needed to make the signals that constitute the feeling of emotion known to the organism having the emotion.

I could see that overcoming the obstacle of self, which meant, from my standpoint, understanding its neural underpinnings, might help us understand the very different biological impact of three distinct although closely related phenomena: *an emotion, the feeling of that emotion,* and *knowing that we have a feeling of that emotion.* No less important, overcoming the obstacle of self might also help elucidate the neural underpinnings of consciousness in general.

The Problem of Consciousness

What is the problem of consciousness, then, from the perspective of neurobiology? Much as I see the matter of self as a critical issue in the elucidation of consciousness, it is important to make clear that the

problem of consciousness is not confined to the matter of self. In the simplest of summaries, I regard the problem of consciousness as a combination of two intimately related problems. The first is the problem of understanding how the brain inside the human organism engenders the mental patterns we call, for lack of a better term, the images of an object. By *object* I mean entities as diverse as a person, a place, a melody, a toothache, a state of bliss; by *image* I mean a mental pattern in any of the sensory modalities, e.g., a sound image, a tactile image, the image of a state of well-being. Such images convey aspects of the physical characteristics of the object and they may also convey the reaction of like or dislike one may have for an object, the plans one may formulate for it, or the web of relationships of that object among other objects. Quite candidly, this first problem of consciousness is the problem of how we get a "movie-in-the-brain," provided we realize that in this rough metaphor the movie has as many sensory tracks as our nervous system has sensory portals—sight, sound, taste, and olfaction, touch, inner senses, and so on. (See the glossary section of the appendix for a comment on the use of terms such as *image, representation,* and *map.*)

From the perspective of neurobiology, solving this first problem consists in discovering how the brain makes neural patterns in its nerve-cell circuits and manages to turn those neural patterns into the explicit mental patterns which constitute the highest level of biological phenomenon, which I like to call images. Solving this problem encompasses, of necessity, addressing the philosophical issue of qualia. Qualia are the simple sensory qualities to be found in the blueness of the sky or the tone of sound produced by a cello, and the fundamental components of the images in the movie metaphor are thus made of qualia. I believe these qualities will be eventually explained neurobiologically although at the moment the neurobiological account is incomplete and there is an explanatory gap.[3]

Now, for the second problem of consciousness. This is the problem of how, in parallel with engendering mental patterns for an object, the brain also engenders a sense of self in the act of knowing. To help

me clarify what I mean by *self* and *knowing*, I urge you to check their presence in your own mind right now.

You are looking at this page, reading the text and constructing the meaning of my words as you go along. But concern with text and meaning hardly describes all that goes on in your mind. In parallel with representing the printed words and displaying the conceptual knowledge required to understand what I wrote, your mind also displays something else, something sufficient to indicate, moment by moment, that *you* rather than anyone else are doing the reading and the understanding of the text. The sensory images of what you perceive externally, and the related images you recall, occupy most of the scope of your mind, but not all of it. Besides those images there is also this other presence that signifies you, as observer of the things imaged, owner of the things imaged, potential actor on the things imaged. There is a presence of you in a particular relationship with some object. If there were no such presence, how would your thoughts belong to you? Who could tell that they did? The presence is quiet and subtle, and sometimes it is little more than a "hint half guessed," a "gift half understood," to borrow words from T. S. Eliot. Later I shall propose that the simplest form of such a presence is also an image, actually the kind of image that constitutes a feeling. In that perspective, the presence of you is the feeling of what happens when your being is modified by the acts of apprehending something. The presence never quits, from the moment of awakening to the moment sleep begins. The presence must be there or there is no you.

The solution for this second problem requires the understanding of how, as I write, I have a sense of me, and how, as you now read, you have a sense of you; of how we sense that the proprietary knowledge you and I behold in our minds, this very moment, is shaped in a particular perspective, that of the individual inside of whom it is formed, rather than in some canonical, one-type-fits-all perspective. The solution also requires the understanding of how the images of an object and of the complex matrix of relations, reactions, and plans related to it are sensed as the unmistakable mental property of an automatic owner who, for all intents and purposes, is an observer, a perceiver, a

knower, a thinker, and a potential actor. This second problem is all the more intriguing since we can be certain that the solution traditionally proposed for it—a homunculus creature who is in charge of knowing—is patently incorrect. There is no homunculus, either metaphysical or in the brain, sitting in the Cartesian theater as an audience of one and waiting for objects to step into the light.[4] In other words, solving the second problem of consciousness consists in discovering the biological underpinnings for the curious ability we humans have of constructing, not just the mental patterns of an object—the images of persons, places, melodies, and of their relationships, in short, the temporally and spatially integrated mental images of something-to-be-known—but also the mental patterns which convey, automatically and naturally, the sense of a self in the act of knowing. Consciousness, as we commonly think of it, from its basic levels to its most complex, is the unified mental pattern that brings together the object and the self.

In the very least, then, the neurobiology of consciousness faces two problems: the problem of how the movie-in-the-brain is generated, and the problem of how the brain also generates the sense that there is an owner and observer for that movie. The two problems are so intimately related that the latter is nested within the former. In effect, the second problem is that of generating the *appearance* of an owner and observer for the movie *within the movie*; and the physiological mechanisms behind the second problem have an influence on the mechanisms behind the first. In spite of the intimacy of the problems, however, separating them is a way of breaking the problem of consciousness into parts and, in so doing, making the overall investigation of consciousness manageable.[5]

This book is about an attempt to deal with the obstacle of consciousness focusing squarely on the problem of self but neither neglecting nor minimizing the "other" problem of consciousness. The attempt was prompted by the impasse on emotions described earlier, but it has gone beyond addressing that particular issue. The book is about my idea of what consciousness is, in mental terms, and about how consciousness can be constructed in the human brain. I do not

claim to have solved the problem of consciousness, and at the current stage in the history of cognitive science and neuroscience, notwithstanding several new and substantial contributions, I regard the thought of solving *the* consciousness problem with some skepticism. I simply hope that the ideas presented here help with the eventual elucidation of the problem of self from a biological perspective.[6]

The background for the text is an ongoing research program that relies on varied lines of investigative activity—reflecting on facts gleaned from many years of observation of neurological patients with disorders of mind and behavior and on findings from experimental neuropsychological studies of such disorders; theorizing about the processes of consciousness as they occur in the normal human condition, using evidence from general biology, neuroanatomy, and neurophysiology; and designing testable hypotheses regarding the neuroanatomical underpinnings of consciousness informed by reflection and theory.

Approaching Consciousness

Before we go any farther it is necessary to say a few words about how to approach the problem we have defined. It would be wonderful, of course, if the contents of our minds were even more richly superposed than they already are, so that I could write this book in parallel tracks and you could read, simultaneously, about theoretical assumptions, scientific methods, and foundational facts. But we operate in a world of classical physics and I must resort to devices of the Elizabethan age: asides and digressions. I promise to be brief and stick to the essentials.

Mind, Behavior, and Brain

Consciousness is an entirely private, first-person phenomenon which occurs as part of the private, first-person process we call mind.[7] Consciousness and mind, however, are closely tied to external behaviors that can be observed by third persons. We all share these phenomena—mind, consciousness within mind, and behaviors—and we know quite well how they are intercorrelated, first because of our own self-analysis, second because of our natural propensity to analyze others. Both wisdom and the science of the human mind and behavior are based on this incontrovertible correlation between the private and the public—first-person mind, on the one hand, and third-person behavior, on the other. Fortunately, for those of us who also wish to understand the mechanisms behind mind and behavior, it so happens that mind and behavior are also closely correlated with the functions of living organisms, specifically with the functions of the brain within those organisms.[8] The power of this triangulation of mind, behavior, and brain has been apparent for over a century and a half—ever since the neurologists Paul Broca and Carl Wernicke discovered a connection between language and certain regions of the left cerebral hemisphere. The triangulation has allowed a most felicitous development: the traditional worlds of philosophy and psychology have gradually joined forces with the world of biology and created an odd but productive alliance. For example, by means of the loose federation of scientific approaches currently known as cognitive neuroscience, the alliance has permitted new advances in the understanding of vision, memory, and language. There is good reason to expect that the alliance will assist with the understanding of consciousness as well.

Over the past two decades, work in cognitive neuroscience has become especially rewarding because the development of new techniques to observe the brain in terms of its structure and function now permits us to link a certain behavior we observe, clinically or in an experiment, not only to the presumed mental counterpart of that behavior, but also to specific indices of brain structure or brain activity.

Let me offer some examples. Areas of circumscribed brain damage caused by neurological disease, which are known as lesions, have long been a mainstay of research on the neural basis of the mind. Such lesions used to be revealed only at the time of autopsy, often many years after the study of the patient had been concluded. This time lag

slowed the process of analysis and generated some uncertainty in the correlation between anatomy and behavior. Recent technical developments, however, permit us to analyze the lesions in a 3-D reconstruction of the living patient's brain at the same time behavioral or cognitive observations are being carried out. The reconstruction is displayed on a computer screen and is based on an elaborate manipulation of raw data obtained from a magnetic resonance scan. It depicts neural structures with great fidelity and allows careful dissection in virtual space rather than on a laboratory bench. The significance of this development is that a lesion analyzed in this detailed and timely manner serves as a probe to test hypotheses about how a brain system performs a certain mental function or behavior. For instance, we may postulate that a system made up of four interconnected brain regions, A, B, C, D, operates in a particular fashion. Then we may predict the kind of changes that must occur when, say, region C is destroyed. To test the validity of the prediction we study how patients with a lesion in area C behave while performing a given task. Incidentally, the same approach is used in another recently evolved area of neuroscience, molecular neurobiology. A specific gene is inactivated experimentally, in a mouse, for instance, thus causing a "lesion" (in scientific jargon this is called a "knock-out"). The investigators can then determine whether the consequences of the "knock-out" are as predicted.[9]

Another example of a new type of brain index is an area of increased or decreased brain activity revealed by a positron emission scan (PET) or a functional magnetic resonance imaging scan (fMRI). Such scans can be used not only in neurological patients but also in humans without brain diseases. Again, a specific prediction concerning the activity of a certain region during the performance of a particular mental task is used to assess the validity of the hypothesis.

Yet another index is a change in electrical conductance response measured in the skin; or a change in electrical potentials and related magnetic fields measured from the scalp; or a change in electrical potentials measured directly on the brain surface during surgery for epilepsy. Remarkably, the possibility of making intricate linkages

among private mind, public behavior, and brain function does not stop with the application of these new techniques. The cross linkages can be extended by a connection to new domains of knowledge about the anatomy and function of the nervous system, gathered by experimental neuroanatomists, neurophysiologists, neuropharmacologists, and neurobiologists who study molecular events within individual nerve cells and can, in turn, relate those events to the composition and action of specific genes. The facts gathered recently on the basis of all these developments allow us to establish progressively more detailed theories regarding the relation between certain aspects of mind and behavior and the brain. The organism's private mind, the organism's public behavior, and its hidden brain can thus be joined in the adventure of theory, and out of the adventure come hypotheses that can be tested experimentally, judged on their merits, and subsequently endorsed, rejected, or modified. (See the appendix for fundamentals of brain anatomy and organization.)

Reflecting on the Neurological and Neuropsychological Evidence

The results of neurological observations and of neuropsychological experiments reveal many facts that were the starting point for the ideas presented here. The first fact is that some aspects of the processes of consciousness can be related to the operation of specific brain regions and systems, thus opening the door to discovering the neural architecture which supports consciousness. The regions and systems in question cluster in a limited set of brain territories and no less so than with functions such as memory or language there will be an anatomy of consciousness. One of the purposes of this text is to present testable anatomical hypotheses for some aspects of the consciousness process.

The second fact is that consciousness and wakefulness, as well as consciousness and low-level attention, can be separated. This fact was based on the evidence that patients can be awake and attentive without having normal consciousness, as exemplified by the man in the circular room. In chapters 3 and 4, I discuss such patients and consider the theoretical significance of their conditions.

The third and perhaps most revealing fact is that consciousness and emotion are *not* separable. As discussed in chapters 2, 3, and 4, it is usually the case that when consciousness is impaired so is emotion. In effect, the connection between emotion and consciousness, on the one hand, and between both of these and the body, on the other, form a main theme of this book.

The fourth fact is that consciousness is not a monolith, at least not in humans: it can be separated into simple and complex kinds, and the neurological evidence makes the separation transparent. The simplest kind, which I call *core consciousness,* provides the organism with a sense of self about one moment—now—and about one place—here. The scope of core consciousness is the here and now. Core consciousness does not illuminate the future, and the only past it vaguely lets us glimpse is that which occurred in the instant just before. There is no elsewhere, there is no before, there is no after. On the other hand, the complex kind of consciousness, which I call *extended consciousness* and of which there are many levels and grades, provides the organism with an elaborate sense of self—an identity and a person, you or me, no less—and places that person at a point in individual historical time, richly aware of the lived past and of the anticipated future, and keenly cognizant of the world beside it.

In short, core consciousness is a simple, biological phenomenon; it has one single level of organization; it is stable across the lifetime of the organism; it is not exclusively human; and it is not dependent on conventional memory, working memory, reasoning, or language. On the other hand, extended consciousness is a complex biological phenomenon; it has several levels of organization; and it evolves across the lifetime of the organism. Although I believe extended consciousness is also present in some nonhumans, at simple levels, it only attains its highest reaches in humans. It depends on conventional memory and working memory. When it attains its human peak, it is also enhanced by language.

The supersense of core consciousness is the first step into the light of knowing and it does not illuminate a whole being. On the other hand, the supersense of extended consciousness eventually brings a full construction of being into the light. In extended consciousness, both the past and the anticipated future are sensed along with the here and now in a sweeping vista as far-ranging as that of an epic novel.

If it is true that core consciousness is the rite of passage into knowing, it is equally true that the levels of knowing which permit human creativity are those which only extended consciousness allows. When we think of the glory that is consciousness, and when we consider consciousness as distinctively human, we are thinking of extended consciousness at its zenith. And yet, as we shall see, extended consciousness is not an independent variety of consciousness: on the contrary, it is built on the foundation of core consciousness. The fine scalpel of neurological disease reveals that impairments of extended consciousness allow core consciousness to remain unscathed. By contrast, impairments that begin at the level of core consciousness demolish the entire edifice of consciousness: extended consciousness collapses as well. The glory that is consciousness requires the orderly enhancement of both kinds of consciousness. But if we are to elucidate the glorious combination, we are well advised to begin by understanding the simpler, foundational kind: core consciousness.[10]

Incidentally, the two kinds of consciousness correspond to two kinds of self. The sense of self which emerges in core consciousness is the *core self,* a transient entity, ceaselessly re-created for each and every object with which the brain interacts. Our traditional notion of self, however, is linked to the idea of identity and corresponds to a nontransient collection of unique facts and ways of being which characterize a person. My term for that entity is the *autobiographical self.* The autobiographical self depends on systematized memories of situations in which core consciousness was involved in the knowing of the most invariant characteristics of an organism's life—who you were born to, where, when, your likes and dislikes, the way you usually react to a problem or a conflict, your name, and so on. I use the term *autobiographical memory* to

denote the organized record of the main aspects of an organism's biography. The two kinds of self are related, and in chapter 6, I explain how the autobiographical self arises from the core self.

A fifth fact: not infrequently, consciousness is simply explained in terms of other cognitive functions, such as language, memory, reason, attention, and working memory. While such functions are indeed necessary for the top tiers of extended consciousness to operate normally, the study of neurological patients suggests that they are not required for core consciousness. Accordingly, a theory of consciousness should *not* be just a theory of how memory, reason, and language help construct, from the top down, an interpretation of what goes on in the brain and mind. To be sure, memory, intelligent inferences, and language are critical to the generation of what I call the autobiographical self and the process of extended consciousness. Some interpretation of the events that take place in an organism can surely arise after the process of autobiographical self and extended consciousness are in place. But I do not believe consciousness began that way, at that high a level in the hierarchy of cognitive processes and that late in the history of life and of each of us. I propose that the earliest forms of consciousness precede inferences and interpretations—they are part of the biological transition that eventually enables inferences and interpretations. Accordingly, a theory of consciousness should account for the simpler, foundational kind of the phenomenon which occurs close to the nonconscious representation of the organism for whose sake the entire show is put together and which can support the later developments of identity and person.

Moreover, a theory of consciousness should *not* be just a theory of how the brain attends to the image of an object. As I see it, natural low-level attention precedes consciousness, while focused attention follows the unfolding of consciousness. Attention is as necessary to consciousness as having images. But attention is not sufficient for consciousness and is not the same as consciousness.

Finally, a theory of consciousness should *not* be just a theory of how the brain creates integrated and unified mental scenes, although the production of integrated and unified mental scenes is an important aspect of consciousness, especially at its highest levels. Those scenes do not exist in a vacuum. I believe they are integrated and unified *because* of the singularity of the organism and *for* the benefit of that single organism. The mechanisms that prompt the integration and unification of the scene require an explanation.

By focusing the explanatory efforts on how the sense of self in the act of knowing an object appears in the mind, I am open to the criticism that I am *just* addressing the problem of so-called self-consciousness and neglecting the remainder of the problem, namely the qualia problem. I would answer the criticism as follows. If "self-consciousness" is taken to mean "consciousness with a sense of self," then all human consciousness is necessarily covered by the term—there is just no other kind of consciousness as far as I can see. I would add that the biological state we describe as sense of self and the biological machinery responsible for engendering it may well have a hand in optimizing the processing of the objects to be known—having a sense of self is not only required for knowing, in the proper sense, but may influence the processing of whatever gets to be known. In other words, the biological processes that pose the second problem of consciousness probably plays a role in the biological processes that pose the first. When I address the self problem, I address the qualia issue with respect to the representation of the organism having consciousness.[11]

A Search for Self

How do we ever know that we are seeing a given object? How do we become conscious in the full sense of the word? How is the sense of self in the act of knowing implanted in the mind? The way into a possible answer for the questions on self came only after I began seeing the problem of consciousness in terms of two key players, the *organism* and the *object,* and in terms of the *relationships* those players hold

in the course of their natural interactions. The organism in question is that within which consciousness occurs; the object in question is any object that gets to be known in the consciousness process; and the relationships between organism and object are the contents of the knowledge we call consciousness. Seen in this perspective, consciousness consists of constructing knowledge about two facts: that the organism is involved in relating to some object, and that the object in the relation causes a change in the organism.

The new perspective also makes the biological realization of consciousness a treatable problem. The process of knowledge construction requires a brain, and it requires the signaling properties with which brains can assemble neural patterns and form images. The neural patterns and images necessary for consciousness to occur are those which constitute proxies for the organism, for the object, and for the relationship between the two. Placed in this framework, understanding the biology of consciousness becomes a matter of discovering how the brain can map *both* the two players *and* the relationships they hold.

The general problem of representing the object is not especially enigmatic. Extensive studies of perception, learning and memory, and language have given us a workable idea of how the brain processes an object, in sensory and motor terms, and an idea of how knowledge about an object can be stored in memory, categorized in conceptual or linguistic terms, and retrieved in recall or recognition modes. The neurophysiologic details of these processes have not been worked out, but the contours of these problems are understandable. From my perspective, neuroscience has been dedicating most of its efforts to understanding the neural basis of what I see as the "object proxy." In the relationship play of consciousness, the object is exhibited in the form of neural patterns in the sensory cortices appropriate to map its characteristics. For example, in the case of the visual aspects of an object, the neural patterns are constructed in a variety of regions of the visual cortices, not just one or two, but many, working in concerted fashion to map the varied aspects of the object in visual terms.[12] On the side of

the organism, however, matters are quite different. To indicate how different matters are, let me suggest an exercise.

Look up from the page, at whatever is directly in front of you, observe intently, and then return to the page. As you did so, the many stations of your visual system, from the retinas to several regions of the brain's cerebral cortex, shifted rapidly from mapping the book's page, to mapping the room in front of you, to mapping the page again. Now turn around 180 degrees and look at what is behind you. Again, mapping of the page vanished swiftly so that the visual system could map the new scene you were contemplating. The moral of the story: in quick succession, precisely the *same* brain regions constructed several entirely *different* maps by virtue of the different motor settings the organism assumed and of the different sensory inputs the organism gathered. The image constructed in the brain's multiplex screens changed remarkably.

Now consider this: while your visual system changed dutifully at the mercy of the objects it mapped, a number of regions in your brain whose job it is to regulate the life process and which contain preset maps that represent varied aspects of your body did not change at all in terms of the *kind* of object they represented. The body remained the "object" all along and will remain so until death ensues. But not only was the *kind* of object precisely the same; the degree of change occurring in the object—the body—was quite small. Why was that so? Because only a narrow range of body states is compatible with life, and the organism is genetically designed to maintain that narrow range and equipped to seek it, through thick and through thin.

What we have in this situation, then, is an intriguing asymmetry that may be phrased in the following terms: some parts of the brain are free to roam over the world and in so doing are free to map whatever object the organism's design permits them to map. On the other hand, some other parts of the brain, those that represent the organism's own state, are not free to roam at all. They are stuck. They can map nothing but the body and do so within largely preset maps. They

are the body's captive audience, and they are at the mercy of the body's dynamic sameness.

There are several reasons behind this asymmetry. First, the composition and general functions of the living body remain the same, in terms of their quality, across a lifetime. Second, the body changes that continuously do occur are small, in terms of their quantity. They have a narrow dynamic range because the body must operate with a limited range of parameters if it is to survive; the body's internal state must be relatively stable by comparison to the environment surrounding it. Third, that stable state is governed from the brain by means of an elaborate neural machinery designed to detect minimal variations in the parameters of the body's internal chemical profile and to command actions aimed at correcting the detected variations, directly or indirectly. (I will address the neuroanatomy of this system in chapter 5. The system is made of not one but many units, the most important of which are located in the brain stem, hypothalamus, and basal forebrain sections of the brain.) In short, the organism in the relationship play of consciousness is the entire unit of our living being, our body as it were; and yet, as it turns out, the part of the organism called the brain holds within it a sort of model of the whole thing. This is a strange, overlooked, and noteworthy fact, and is perhaps the single most important clue as to the possible underpinning of consciousness.

I have come to conclude that the organism, as represented inside its own brain, is a likely biological forerunner for what eventually becomes the elusive sense of self. The deep roots for the self, including the elaborate self which encompasses identity and personhood, are to be found in the ensemble of brain devices which continuously and *nonconsciously* maintain the body state within the narrow range and relative stability required for survival. These devices continually represent, *nonconsciously,* the state of the living body, along its many dimensions. I call the state of activity within the ensemble of such devices the *proto-self,* the nonconscious forerunner for the levels of self which appear in our minds as the conscious protagonists of consciousness: core self and autobiographical self.

Should some readers get worried, at this point, that I am falling into the abyss of the homunculus trap, let me say immediately and vehemently that that is not the case. The "model of the body-in-the-brain" to which I am referring is nothing at all like the rigid homunculus creature of old-fashioned neurology textbooks. Nothing in it looks like a little person inside a big person; the model "perceives" nothing and "knows" nothing; it does not talk and it does not make consciousness. The model is, instead, a collection of brain devices whose main job is the automated management of the organism's life. As we shall discuss, the management of life is achieved by a variety of innately set regulatory actions—secretion of chemical substances such as hormones as well as actual movements in viscera and in limbs. The deployment of these actions depends on the information provided by nearby neural maps which signal, moment by moment, the state of the entire organism. Most importantly, neither the life-regulating devices nor their body maps are the generators of consciousness, although their presence is indispensable for the mechanisms that do achieve core consciousness.

This is the key issue, as argued in chapter 5: in the relationship play of consciousness, the organism is represented in the brain, abundantly and multifariously, and that representation is tied to the maintenance of the life process. If this idea is correct, life and consciousness, specifically the self aspect of consciousness, are indelibly interwoven.

Why We Need Consciousness

If you find the connection between life and consciousness surprising, consider the following. Survival depends on finding and incorporating sources of energy and on preventing all sorts of situations which threaten the integrity of living tissues. It is certainly true that without actions organisms such as ours would not survive since the sources of energy required for renewing the organism's structure and maintaining life would not be found and harnessed to the service of the organism, never mind staving off environmental dangers. But on their own,

without the guidance of images, actions would not take us far. Good actions need the company of good images. Images allow us to choose among repertoires of previously available patterns of action and optimize the delivery of the chosen action—we can, more or less deliberately, more or less automatically, review mentally the images which represent different options of action, different scenarios, different outcomes of action. We can pick and choose the most appropriate and reject the bad ones. Images also allow us to invent new actions to be applied to novel situations and to construct plans for future actions—the ability to transform and combine images of actions and scenarios is the wellspring of creativity.

If actions are at the root of survival and if their power is tied to the availability of guiding images, it follows that a device capable of maximizing the effective manipulation of images in the service of the interests of a particular organism would have given enormous advantages to the organisms that possessed the device and would probably have prevailed in evolution. Consciousness is precisely such a device.

The pathbreaking novelty provided by consciousness was the possibility of connecting the inner sanctum of life regulation with the processing of images. Put in other words, it was the possibility of bringing the system of life regulation—which is housed in the depths of the brain in regions such as the brain stem and hypothalamus—to bear on the processing of the images which represent the things and events which exist inside and outside the organism. Why was this really an advantage? Because survival in a complex environment, that is, efficient management of life regulation, depends on taking the right action, and that, in turn, can be greatly improved by purposeful preview and manipulation of images in mind and optimal planning. Consciousness allowed the connection of the two disparate aspects of the process—inner life regulation and image making.

Consciousness generates the knowledge that images exist within the individual who forms them, it places images in the organism's perspective by referring those images to an integrated representation

of the organism, and, in so doing, allows the manipulation of the images to the organism's advantage. Consciousness, when it appears in evolution, announces the dawn of individual forethought.

Consciousness opens the possibility of constructing in the mind some counterpart to the regulatory specifications hidden in the brain core, a new way for the life urge to press its claims and for the organism to act on them. Consciousness is the rite of passage which allows an organism armed with the ability to regulate its metabolism, with innate reflexes, and with the form of learning known as conditioning, to become a minded organism, the kind of organism in which responses are shaped by a mental *concern* over the organism's own life. Spinoza said that the effort to preserve oneself is the first and unique foundation of virtue.[13] Consciousness enables that effort.

The Beginning of Consciousness

Once I could envision how the brain might put together the patterns that stand for an object and those that stand for the organism, I began considering the mechanisms that the brain may use to represent the relationship between object and organism. I was looking specifically for how the brain might represent the fact that when an organism is engaged in the processing of an object, the object *causes* the organism to react and, in so doing, change its state. A possible solution is presented in chapters 6, 7, and 8. I propose that we become conscious when the organism's representation devices exhibit a specific kind of wordless knowledge—the knowledge that the organism's own state has been changed by an object—and when such knowledge occurs along with the salient representation of an object. The sense of self in the act of knowing an object is an infusion of *new* knowledge, continuously created within the brain as long as "objects," actually present or recalled, interact with the organism and cause it to change.

The sense of self is the first answer to a question the organism never posed: To whom do the ongoing mental patterns now unfolding

belong? The answer is that they belong to the organism, as represented by the proto-self. Later I indicate how the brain assembles the wordless knowledge necessary to produce this unrequested answer. At this point, however, I can say that the simplest form in which the wordless knowledge emerges mentally is the feeling of knowing—the feeling of what happens when an organism is engaged with the processing of an object—and that only thereafter can inferences and interpretations begin to occur regarding the feeling of knowing.

In a curious way, consciousness begins as the feeling of what happens when we see or hear or touch. Phrased in slightly more precise words, it is a feeling that accompanies the making of any kind of image—visual, auditory, tactile, visceral—within our living organisms. Placed in the appropriate context, the feeling marks those images as ours and allows us to say, in the proper sense of the terms, that we see or hear or touch. Organisms unequipped to generate core consciousness are condemned to making images of sight or sound or touch, there and then, but cannot come to know that they did. From its most humble beginnings, consciousness is knowledge, knowledge consciousness, no less interconnected than truth and beauty were for Keats.

Coping with Mystery

There has been a lack of agreement among those studying the problem of consciousness, not only about what consciousness is but also about the prospects of understanding its biological underpinnings. There has also been some degree of puzzlement, and even worry, among those who are not scholars of consciousness but simple day-to-day users, about the human consequences of elucidating the biology of consciousness. For some nonspecialists, consciousness and mind are virtually indistinguishable, and so are consciousness and conscience, and consciousness and soul, or consciousness and spirit. For them, maybe for you, mind, consciousness, conscience, soul, and spirit form one big region of strangeness that sets humans apart, that

separates the mysterious from the explainable and the sacred from the profane. It should not be surprising to discover that the manner in which this sublime conflation of human properties is approached matters greatly to any sensible human being, and even that offense can be taken at seemingly dismissive accounts of its nature. Anyone who has faced death will know precisely what I am referring to, perhaps because the irreversibility of death focuses our thoughts sharply on the monumental scale of the human minded life. It should not take death, however, to make anyone sensitive to this issue. Life should be enough to make us approach the human mind with respect for its dignity and stature, and, almost paradoxically, with tenderness for its fragility.

Let me make something clear, however. Science helps us make distinctions among phenomena and science can now successfully distinguish among several components of the human mind. Consciousness and conscience are in fact distinguishable: consciousness pertains to the knowing of any object or action attributed to a self, while conscience pertains to the good or evil to be found in actions or objects. Consciousness and mind are also distinguishable: consciousness is the part of mind concerned with the apparent sense of self and knowing. There is more to mind than just consciousness and there can be mind without consciousness, as we discover in patients who have one but not the other.

In its progress, science proposes explanations for the phenomena it manages to distinguish. In the case of mind, it manages to explain parts of the big region of strangeness. It gleans *some* mechanisms behind *some* phenomena which contribute to the creation of the admirable human mind we so respect. Yet the admirable creation does not vanish just because we manage to explain some of the component mechanisms necessary for it to occur. The appearance *is* the reality—the human mind as we directly sense it. When we explain the mind, we get to keep that reality while we satisfy part of our curiosity regarding the sleight of hand behind the appearance.

Another issue I must make clear: solving the mystery of consciousness is not the same as solving all the mysteries of the mind. Consciousness is an indispensable ingredient of the creative human mind, but it is not all of human mind, and, as I see it, it is not the summit of mental complexity, either. The biological tricks that cause consciousness have powerful consequences, but I see consciousness as an intermediary rather than as the culmination of biological development. Ethics and the law, science and technology, the work of the muses and the milk of human kindness, those are my chosen summits for biology. Surely, we would have none of that without the wonders of consciousness at the source of each new achievement. Still, consciousness is a sunrise, not the midday sun, and a sunset even less. Understanding consciousness says little or nothing about the origins of the universe, the meaning of life, or the likely destiny of both. After solving the mystery of consciousness and making a dent on a few related mysteries of mind, assuming science achieves either, there is enough mystery left to last many a scientific lifetime, enough awe at nature to keep us modest for the foreseeable future. After considering how consciousness may be produced within the three pounds of flesh we call brain, we may revere life and respect human beings more, rather than less.

Hide and Seek

Sometimes we use our minds not to discover facts but to hide them. We use part of the mind as a screen to prevent another part of it from sensing what goes on elsewhere. The screening is not necessarily intentional—we are not deliberate obfuscators all of the time—but deliberate or not, the screen does hide.

One of the things the screen hides most effectively is the body, our own body, by which I mean the ins of it, its interiors. Like a veil thrown over the skin to secure its modesty, but not too well, the screen partially removes from the mind the inner states of the body, those that constitute the flow of life as it wanders in the journey of each day.

The alleged vagueness, elusiveness, and intangibility of emotions and feelings are probably symptoms of this fact, an indication of how we cover the representation of our bodies, of how much mental imagery based on nonbody objects and events masks the reality of the body. Otherwise we would easily know that emotions and feelings are tangibly about the body. Sometimes we use our minds to hide a part of our beings from another part of our beings.

I could describe the hiding of the body as a distraction, but I would have to add that it is a very adaptive distraction. In most circumstances, rather than concentrating resources on our inner states, it is perhaps more advantageous to concentrate one's resources on the images that describe problems out in the world or on the premises of those problems or on the options for their solution and their possible outcomes. Yet this skewing of perspective relative to what is available in our minds has a cost. It tends to prevent us from sensing the possible origin and nature of what we call self. When the veil is lifted, however, at the scale of understanding permitted to the human mind, I believe we can sense the origin of the construct we call self in the representation of individual life.

Perhaps it was easier to get a more balanced perspective in earlier times when there was no veil, when the environments were relatively simple, long before electronic media and jet travel, long before the printed word, before the empire, and ahead of the city-state. It must have been easier to sense the life within, when the brain provided a lopsided view in the opposite direction, tilted toward the dominant representation of the internal states of the organism. If it ever was like that, perhaps at some magic brief time between Homer and Athens, lucky humans would have perceived in an instant that all of their amusing antics were about life and that underneath every image of the outside world, there stood the ongoing image of their living bodies. Or perhaps they would not have perceived as much because they lacked the frame of reference that current knowledge about biology provides us. Be that as it may, I suspect they were able to sense more about themselves than many of us, the unforewarned, are able to

sense these days. I marvel at the ancient wisdom of referring to what we now call mind by the word *psyche* which was also used to denote breath and blood.

I suggest that the highly constrained ebb and flow of internal organism states, which is innately controlled by the brain and continuously signaled in the brain, constitutes the backdrop for the mind, and, more specifically, the foundation for the elusive entity we designate as self. I also suggest that those internal states—which occur naturally along a range whose poles are pain and pleasure, and are caused by either internal or external objects and events—become unwitting nonverbal signifiers of the goodness or badness of situations relative to the organism's inherent set of values. I suspect that in earlier stages of evolution these states—including all of those we classify as emotions—were entirely unknown to the organisms producing them. The states were regulatory and that was enough; they produced some advantageous actions, internally or externally, or they assisted indirectly the production of such actions by making them more propitious. But the organisms carrying out these complicated operations knew nothing of the existence of those operations and actions since they did not even know, in the proper sense of the word, of their own existence as individuals. True enough, organisms had a body and a brain, and brains had some representation of the body. Life was there, and the representation of life was there, too, but the potential and rightful owner of each individual life had no knowledge that life existed because nature had not invented an owner yet. There was being but not knowing. Consciousness had not begun.

Consciousness begins when brains acquire the power, the simple power I must add, of telling a story without words, the story that there is life ticking away in an organism, and that the states of the living organism, within body bounds, are continuously being altered by encounters with objects or events in its environment, or, for that matter, by thoughts and by internal adjustments of the life process. Consciousness emerges when this primordial story—the story of an object causally changing the state of the body—can be told using the

universal nonverbal vocabulary of body signals. The apparent self emerges as the feeling of a feeling. When the story is first told, spontaneously, without it ever having been requested, and forevermore after that when the story is repeated, knowledge about what the organism is living through automatically emerges as the answer to a question never asked. From that moment on, we begin to know.

I suspect consciousness prevailed in evolution because knowing the feelings caused by emotions was so indispensable for the art of life, and because the art of life has been such a success in the history of nature. But I will not mind if you prefer to give my words a twist and just say that consciousness was invented so that we could know life. The wording is not scientifically correct, of course, but I like it.

The Feeling of What Happens



BODY AND EMOTION
IN THE MAKING
OF CONSCIOUSNESS

Antonio Damasio

A HARVEST BOOK
HARCOURT, INC.
San Diego New York London

For Hanna

Copyright © 1999 by Antonio Damasio

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Requests for permission to make copies of any part of the work should be mailed to the following address: Permissions Department, Harcourt, Inc., 6277 Sea Harbor Drive, Orlando, Florida 32887-6777.

"Notes on the Reality of the Self" from *Materialism* by Jorie Graham. Copyright © 1995 by Jorie Graham. Reprinted by permission of The Ecco Press.

All figures courtesy of Dr. Hanna Damasio.

Excerpt from "Dry Salvages" in *Four Quartets*, copyright 1941 by T. S. Eliot and renewed 1969 by Esme Valerie Eliot, reprinted by permission of Harcourt, Inc.

Library of Congress Cataloging-in-Publication Data
Damasio, Antonio R.
The feeling of what happens: body and emotion in the making of consciousness/Antonio R. Damasio.—1st ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-15-100369-6
ISBN 0-15-601075-5 (pbk.)
1. Consciousness. 2. Consciousness—Physiological aspects.
3. Emotions—Physiological aspects. I. Title.
BF311.D33 1999
153—dc21 99-26357

Text set in Spectrum MT
Designed by Lori McThomas Buley
Printed in the United States of America
First Harvest edition 2000
CEFDB

Or the waterfall, or music heard so deeply
That it is not heard at all, but you are the music
While the music lasts. These are only hints and guesses,
Hints followed by guesses; and the rest
Is prayer, observance, discipline, thought and action.
The hint half guessed, the gift half understood, is Incarnation.

T. S. ELIOT
"Dry Salvages" from *Four Quartets*

The question of who I was consumed me.

I became convinced I should not find the image
of the person that I
was: Seconds passed. What rose to the surface in me
plunged out of sight again. And yet I felt
the moment of my first investiture
was the moment I began to represent myself—
the moment I began to live—by degrees—second by
second—unrelentingly—Oh mind what you're doing!—

do you want to be *covered* or do you want to be *seen*?—

And the garment—how it becomes you!—starry
with the eyes of
others,
weeping—

JORIE GRAHAM
"Notes on the Reality of the Self" from *Materialism*