Gerald M. Edelman

Wider Than the Sky

Gerald M. Edelman

Wider Than the Sky

a revolutionary view of consciousness

PENGUIN BOOKS

Print published in Penguin 2005
1

978-0-14-193756-4

Contents

Index

Preface

Consciousness is the guarantor of all we hold to be human and precious. Its permanent loss is considered equivalent to death, even if the body persists in its vital signs. No wonder, then, that consciousness has attracted speculation and study across the ages. Over the past twenty-five years, I have written a number of books and papers on the subject. My conviction that consciousness is susceptible to scientific study has been supported by a sharp increase in the number of publications and scientific meetings on the subject.

These developments have prompted me to present an account of consciousness to the general reader. In carrying out this project, my goals were clear: to define consciousness and to offer as simple a view of the subject as is consistent with clarity. The subject is a challenging one and it will certainly require a concentrated effort on the part of the reader. I can only promise that the reward for such effort will be a deeper insight into issues that are at the center of human concern. Accordingly, except when absolutely necessary, I have deliberately omitted many scholarly references, which may be found in abundance in my previous works. Those interested in further reading can find a number of excellent works that have informed this book listed in the Bibliographic Note. I am aware that a great barrier to understanding scientific presentations rests in the inevitable use of technical terms. The problem is compounded when one considers details related to the brain and consciousness. For this reason, I have added a glossary that I hope will provide some alleviation.

William James, whose descriptions of consciousness still stand as a high-water mark in the field, said:

> Something definite happens when to a certain brain-state a certain "sciousness" corresponds. A genuine glimpse into what it is would be the scientific achievement, before which all past achievements would pale. But at the present, psychology is in the condition of physics before Galileo and the laws of motion, of chemistry before Lavoisier and the notion that mass is preserved in all reactions. The Galileo and the Lavoisier of psychology will be famous men indeed when they come, as come they some day surely will, or past successes are no index to the future. When they do come, however, the necessities of the case will make them "metaphysical." Meanwhile the best way in which we can facilitate their advent is to understand how great is the darkness in which we grope, and never forget that the natural-science assumptions with which we started are provisional and revisable things.

I have puzzled over what James had in mind in stating that successful scientific efforts to glimpse the bases of consciousness would *necessarily* metaphysical. In any event, in this book I have tried to avoid extensive discussion of metaphysical matters. I intend to deal with explanations that rest solely on a scientific base. My hope is to disenthrall those who believe the subject is exclusively metaphysical or necessarily mysterious.

A scientific analysis of consciousness must answer the question: How can the firing of neurons give rise to subjective sensations, thoughts, and emotions? To some, the two domains are so disparate as to be irreconcilable. A scientific explanation must provide a causal account of the connection between these domains so that properties in one domain may be understood in terms of events in the other. This is the task I have set myself in this small book.

The title of the book comes from a poem by Emily Dickinson that appears as the epigraph. This poem was written in around 1862, before modern brain science began toward the end of the nineteenth century. I find it impressive that, in extolling the width and depth of the mind, Dickinson referred exclusively to the brain.

Acknowledgments

I thank Drs. Kathryn Crossin, David Edelman, Joseph Gaily, Ralph Greenspan, and George Reeke for their valuable criticism and useful suggestions. The illustrations were composed by Eric Edelman; I am grateful for his skilled and patient responses to my sometimes idiosyncratic suggestions. Darcie Plunkett provided excellent help in preparing this manuscript.

consciousness develops in certain animals.

Two subtle but important issues strongly influence our interpretation of these requirements. The first of these is the question of the causal status of consciousness. Some take the view that consciousness is a mere epiphenomenon with no material consequences. A contrary view is that consciousness is efficacious—that it causes things to happen. We will take the position, which we shall explore in detail later, that it suffices to show that the neural bases of consciousness, not consciousness itself, can cause things to happen. The second major challenge to any scientific account of consciousness is to show how a neural mechanism entails a subjective conscious state, or quale, as it is called. Before we can meet these two challenges, it is necessary to provide a sketch of the properties of consciousness and consider some matters of brain structure and function.

Chapter 2

Consciousness

THE REMEMBERED PRESENT

He all know what consciousness is: it is what you lose when you fall into a deep dreamless sleep and what you regain when you wake up. But this glib statement does not leave us in a comfortable position to examine consciousness scientifically. For that we need to explore the salient properties of consciousness in more detail, as William James did in his *Principles of Psychology.* Before doing so, it will help to clarify the subject if we first point out that consciousness is utterly dependent on the brain. The Greeks and others believed that consciousness resided in the heart, an idea that survives in many of our common metaphors. There is now a vast amount of empirical evidence to support the idea that consciousness emerges from the organization and operation of the brain. When brain function is curtailed—in deep anesthesia, after certain forms of brain trauma, after strokes, and in certain limited phases of sleep—consciousness is not present. There is no return of the functions of the body and brain after death, and postmortem experience is simply not possible. Even during life there is no scientific evidence for a free-floating spirit or consciousness outside the body: consciousness is embodied. The question then becomes: What features of the body and brain are necessary and sufficient for consciousness to appear? We can best answer that question by specifying how the properties of conscious experience can emerge from properties of the brain.

Before taking up the properties of consciousness in this chapter, we must address another consequence of embodiment. This concerns the private or personal nature of each person's conscious experience. Here is James on the subject:

> In this room—this lecture room, say—there are a multitude of thoughts, yours and mine, some of which cohere mutually, and some not. They are as little each-for-itself and reciprocally independent as they are all-belonging-together. They are neither: no one of them is separate, but each belongs with certain others and with none beside. My thought belongs with my other thoughts and your thought with your other thoughts. Whether anywhere in the room there be a mere thought, which is nobody's thought, we have no means of ascertaining, for we have no experience of the like. The only states of consciousness that we naturally deal with are found in personal consciousness, minds, selves, concrete particular I's and you's.

There is no mystery here. Since consciousness arrives as a result of each individual's brain and bodily functions, there can be no direct or collective sharing of that individual's unique and historical conscious experience. But this does not mean that it is impossible to isolate the salient features of that experience by observation, experiment, and report.

What is the most important statement one can make about consciousness from this point of view? It is that consciousness is a process, not a thing. James made this point trenchantly in his essay "Does Consciousness Exist?" To this day, many category errors have been made as a result of ignoring this point. For example, there are accounts that attribute consciousness specifically to nerve cells (or "consciousness neurons") or to particular layers of the cortical mantle of the brain. The evidence, as we shall see, reveals that the process of consciousness is a dynamic accomplishment of the distributed activities of populations of neurons in many different areas of the brain. That an area may be essential or necessary for consciousness does not mean it is sufficient. Furthermore, a given neuron may contribute to conscious activity at one moment and not at the next.

There are a number of other important aspects of consciousness as a process that may be called Jamesian properties. James pointed out that consciousness occurs only in the individual (that is, it is private or subjective), that it appears to be continuous, albeit continually changing, that it has intentionality (a term referring to the fact that, generally, it is about things), and that it does not exhaust all aspects of the things or events to which it refers. This last property has a connection to the important matter of attention. Attention, particularly focal attention, modulates conscious states and directs them to some extent, but it is not the same as consciousness. I will return to this issue in later chapters.

One outstanding property is that consciousness is unitary or integrated, at least in healthy individuals. When I consider my conscious state at the time of this writing, it appears to be all of a piece. While I am paying attention to the act of writing, I am aware of a ray of sunlight, of a humming sound across the street, of a small discomfort in my legs at the edge of the chair, and even of a "fringe," as James called it, that is of objects and events barely sensed. It is usually not entirely possible to reduce this integrated scene to just one thing, say my pencil. Yet this unitary scene will change and differentiate according to outside stimuli or inner thoughts to yet another scene. The number of such differentiated scenes seems endless, yet each is unitary. The scene is not just wider than the sky, it can contain many disparate elements—sensations, perceptions, images, memories, thoughts, emotions, aches, pains, vague feelings, and so on. Looked at from the inside, consciousness seems continually to change, yet at each moment it is all of piece—what I have called "the remembered present"—reflecting the fact that all my past experience is engaged in forming my integrated awareness of this single moment.

This integrated yet differentiated state looks entirely different to an outside observer, who possesses his or her own such states. If an outside observer tests whether I can essential to understand that differences in qualia are based on differences in the wiring and activity of parts of the nervous system. It is also valuable to understand that qualia are always experienced as parts of the unitary and integrated conscious scene. Indeed, all conscious events involve a complex of qualia. In general, it is not possible to experience only a single quale—"red," say—in isolation.

I shall elaborate later on the statement that qualia reflect the ability of conscious individuals to make high-order discriminations. How does such an ability reflect the efficacy of the neural states accompanying conscious experience? Imagine an animal with primary consciousness in the jungle. It hears a low growling noise, and at the same time the wind shifts and the light begins to wane. It quickly runs away, to a safer location. A physicist might not be able to detect any necessary causal relation among these events. But to an animal with primary consciousness, just such a set of simultaneous events might have accompanied a previous experience, which included the appearance of a tiger. Consciousness allowed integration of the present scene with the animal's past history of conscious experience, and that integration has survival value whether a tiger is present or not. An animal without primary consciousness might have many of the individual responses that the conscious animal has and might even survive. But, on average, it is more likely to have lower chances of survival—in the same environment it is less able than the conscious animal to discriminate and plan in light of previous and present events.

In succeeding chapters, I will attempt to explain how conscious scenes and qualia arise as a result of brain dynamics and experience. At the outset, though, it is important to understand what a scientific explanation of conscious properties can and cannot do. The issue concerns the so-called explanatory gap that arises from the remarkable differences between brain structure in the material world and the properties of qualia-laden experience. How can the firing of neurons, however complex, give rise to feelings, qualities, thoughts, and emotions? Some observers consider the two realms so widely divergent as to be impossible to reconcile. The key task of a scientific description of consciousness is to give a causal account of the relationship between these domains so that properties in one domain may be understood in terms of events in the other.

What such an explanation cannot and need not do is offer an explanation that replicates or creates any particular quale or experiential state. Science does not do that—indeed, imagine that a gifted scientist, through an understanding of fluid dynamics and meteorology, came up with a powerful theory of a complex world event like a hurricane. Implemented by a sophisticated computer model, this theory makes it possible to understand how hurricanes arise. Furthermore, with the computer model, the scientist could even predict most of the occurrences and properties of individual hurricanes. Would a person from a temperate zone without hurricanes, on hearing and understanding this theory, then expect to experience a hurricane or even get wet? The theory allows one to understand how hurricanes arise or are entailed by certain conditions, but it cannot create the experience of hurricanes. In the same way, a brain-based theory of consciousness should give a causal explanation of its properties but, having done so, it should not be expected to generate qualia "by description."

To develop an adequate theory of consciousness, one must comprehend enough of how the brain works to understand phenomena, such as perception and memory, that contribute to consciousness. And if these phenomena can be causally linked, one would hope to test their postulated connections to consciousness by experimental means. This means that one must find the neural correlates of consciousness. Before addressing these issues, let us turn first to the brain.

Chapter 3

Elements of the Brain

The human brain is the most complicated material object in the known universe. I have already said that certain processes within the brain provide the necessary mechanisms underlying consciousness. In the past decade or so, many of these processes have been identified. Brain scientists have described an extraordinary layering of brain structures at levels ranging from molecules to neurons (the message-carrying cells of the brain), to entire regions, all affecting behavior. In describing those features of the brain necessary to our exploration I will not go into great detail. To provide a foundation for a biological theory of consciousness, however, we do need to consider certain basic information on brain structure and dynamics. This excursion will require some patience on the reader's part. It will be rewarded when we develop a picture of how the brain works.

This short survey on the brain will cover, in order, a global description of brain regions, some notion of their connectivity, the basics of the activity of neurons and their connections—the synapses—and a bit of the chemistry underlying neuronal activity. All this will be necessary to confront a number of critical questions and principles: Is the brain a computer? How is it built during development? How complex are its transactions? Are there new principles of organization unique to the brain that were selected during evolution? What parts of the brain are necessary and sufficient for consciousness to emerge? In addressing these questions, I shall use the human brain as my central reference. There are, of course, many similarities between our brains and those of other animal species, and when necessary I shall describe these similarities as well as any significant differences.

The human brain weighs about three pounds. Its most prominent feature is the overlying wrinkled and convoluted structure known as the cerebral cortex, which is plainly visible in pictures of the brain (Figure 1). If the cerebral cortex were unfolded (making the gyri, its protrusions, and the sulci, its clefts, disappear) it would have the size and thickness of a large table napkin. It would contain at least 30 billion neurons, or nerve cells, and 1 million billion connections, or synapses. If you started counting these synapses right now at a rate of one per second, you would just finish counting them 32 million years from now.

channels in the postsynaptic cell that, acting cumulatively, can cause it to fire an action potential of its own. Thus, neuronal communication occurs by a combination of controlled electrical and chemical events.

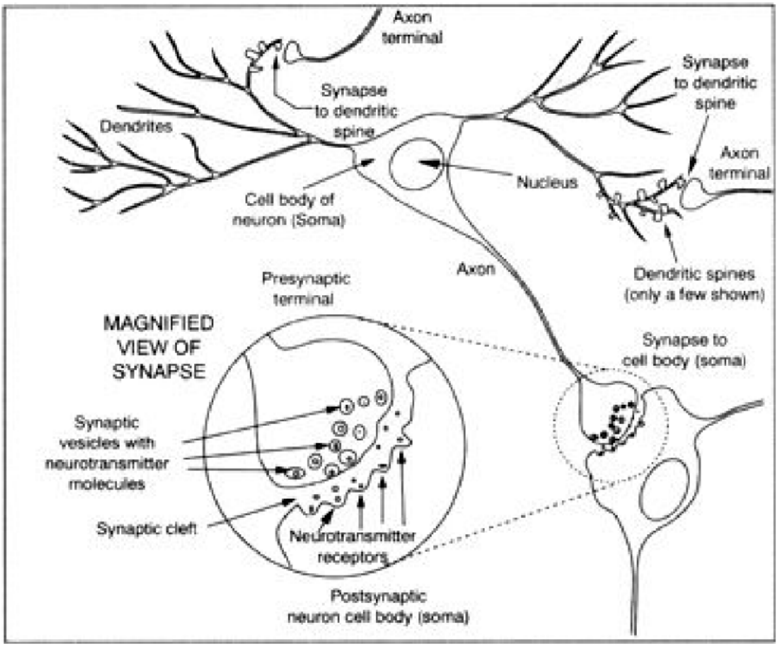

Figure 2. A diagram illustrating synaptic connections between two neurons. An action potential traveling down the axon of the presynaptic neuron causes the release of a neurotransmitter into the synaptic cleft. The transmitter molecules bind to receptors in the postsynaptic membrane, changing the probability that the postsynaptic neuron will fire. (Because of the number of different shapes and kinds of neurons, this drawing is of necessity a greatly simplified

cartoon.)

Now try to imagine the enormous numbers of neurons firing in various areas of the brain. Some firings are coherent (that is, they are simultaneous), others are not. Different brain regions have different neurotransmitters and chemicals whose properties change the timing, amplitude, and sequences of neuronal firing. To achieve and maintain the complex patterns of dynamic activity in healthy brains, some neurons are inhibitory, suppressing the firing of others, which are excitatory. Most excitatory neurons use the substance glutamate as their neurotransmitter, while the inhibitory neurons use GABA (gamma-aminobutyric acid). We can ignore the chemical details for now and simply accept that the effects of different chemical structures are different and that their distribution and occurrence together can have significant effects on neural activity.

I started by describing the cortex. With the picture of a polar neuron in mind, we can turn briefly to other key regions of the brain. One of the most important anatomical structures for understanding the origin of consciousness is the thalamus. This structure, which is located at the center of the brain, is essential for conscious function, even though it is only somewhat larger than the last bone in your own thumb. When nerves from different sensory receptors serving different modalities (located in your eyes, ears, skin, and so on) travel to your brain, they each connect in the thalamus with specific clusters of neurons called nuclei. Postsynaptic neurons in each specific thalamic nucleus then project axons that travel and map to particular areas of the cortex. A well-studied example is the projection from the neurons of the retina through the optic nerve to the part of the thalamus called the lateral geniculate nucleus and then to the primary visual cortical area, called VI (for "visual area 1").

There is one striking feature of the many connections between the thalamus and the cortex: not only does the cortex receive many axons from thalamic neurons but there are also reciprocal axonal fibers going from the cortex back to the thalamus. We speak therefore of thalamocortical projections and corticothalamic projections. Reciprocal connections of this type abound within the cortex itself; such reciprocal connections are called corticocortical tracts. A striking example of these is the fiber bundle called the corpus callosum, which connects the two cortical hemispheres and consists of more than 200 million reciprocal axons. Cutting the corpus callosum leads to a split-brain syndrome, which in some cases can lead to the remarkable appearance of two separate and very different consciousnesses.

Each specific thalamic nucleus (and there are many) does not connect directly to any of the others. Surrounding the periphery of the thalamus, however, there is a layered structure called the reticular nucleus, which connects to the specific nuclei and which can inhibit their activity. The reticular nucleus, it is suspected, acts to switch or "gate" the activities of the specific thalamic nuclei, yielding different patterns of expression of such sensory modalities as sight, hearing, and touch. Another set of thalamic nuclei called intralaminar nuclei receive connections from certain lower structures in the brainstem that are concerned with activation of multiple neurons; these then project to many different areas of the cortex. The activity of these intralaminar nuclei is suspected to be essential for consciousness in that it sets appropriate thresholds or levels of cortical response—with too high a threshold, consciousness would be lost.

We may now turn to some other brain structures that are important to our efforts to track down the neural bases of consciousness. These are large subcortical regions that include the hippocampus, the basal ganglia, and the cerebellum. The hippocampus is an evolutionarily ancient cortical structure lined up like a pair of curled sausages along the inner skirt of the temporal cortex, one on the right side and another on the left. In cross section, each sausage looks like a sea horse, hence the name "hippocampus." Studies of the neural properties of the hippocampus provide important examples of some of the synaptic mechanisms underlying memory. One such mechanism, which should *not* be equated with memory itself, is the change in the strength, or efficacy, of hippocampal synapses that occurs with certain patterns of neural stimulation. As a result of this change, which can be either positive for long-term potentiation or negative for long-term depression, certain neural pathways are dynamically favored over others.

The point to be stressed is that, while synaptic change is essential for the function of memory, memory is a system property that also depends on specific neuroanatomical connections.

Increased synaptic strength or efficacy within a pathway leads to a higher likelihood of conduction across that pathway, whereas decreases in synaptic strength diminish that likelihood. Various patterns have been found for the so-called synaptic rules governing these changes, following the initial proposals of Donald Hebb, a psychologist, and Friedrich von Hayek, an economist who, as a young man, thought quite a bit about how the brain works. These scholars suggested that an increase in synaptic efficacy would occur when pre- and postsynaptic neurons fired in close temporal order. Various modifications of this fundamental rule have been seen in different parts of the nervous system. What is particularly striking about the hippocampus, where these rules have been studied in detail, is the fact that bilateral removal of this structure leads to a loss of episodic memory, the memory of specific episodes or experiences in life. A very famous patient, H. M., whose hippocampi were removed to cure epileptic seizures, could not, for example, convert his short-term memory of events into a permanent narrative record, a condition that was depicted dramatically in the movie *Memento.* It is believed that such a long-term record results when particular synaptic connections between the hippocampus and the cortex are strengthened. When these connections are severed, the corresponding cortical synaptic changes cannot take place and the ability to remember episodes over the long term is lost. Such a patient can remember episodes up to the time of the operation, but loses long-term memory thereafter. It is intriguing that in some acetylcholine; the dopaminergic nuclei, which release dopamine; and the histaminergic system, which resides in a subcortical region called the hypothalamus, a region that affects many critical body functions.

The striking feature of such value systems is that, by projecting diffusely, each affects large populations of neurons simultaneously by releasing its neurotransmitter in the fashion of a leaky garden hose. By doing so, these systems affect the probability that neurons in the neighborhood of value-system axons will fire after receiving glutamatergic input. These systems bias neuronal responses affecting both learning and memory and controlling bodily responses necessary for survival. It is for this reason that they are termed value systems. In addition, there are other loci in the brain with modulatory functions mediated by substances called neuropeptides. An example is enkephalin, an endogenous opioid that regulates responses to pain. In addition, there are other brain areas, such as the amygdala, which are involved in emotional responses, such as fear. For our purposes, these areas need not be described in detail.

To summarize our account so far, we may say that, in a gross sense, there are three main neuroanatomical motifs in our brains (Figure 3). The first is the thalamocortical motif, with tightly connected groups of neurons connected both locally and across distances by rich reciprocal connections. The second is the polysynaptic loop structure of the inhibitory circuits of the basal ganglia. The third consists of the diffuse ascending projections of the different value systems. Of course, this generalization is a gross oversimplification, given the exquisite detail and individuation of neural circuitry. But as we shall see, it provides a useful simplification; we can dispose of it once we have seen its uses.

Figure 3. Fundamental arrangements of three kinds of neuroanatomical systems in the brain. The top diagram shows the gross topology of the thalamocortical system, which is a dense meshwork of reentrant connectivity between the cortex and the thalamus and among different cortical areas. The middle diagram shows the long polysynaptic loops connecting the cortex with subcortical structures such as the basal ganglia. In this case, these loops go from the basal ganglia to the thalamus, thence to the cortex and back from the target areas of cortex to the ganglia. These loops are, in general, not reentrant. The bottom diagram shows one of the diffusely projecting value systems, in which the locus coeruleus distributes a "hairnet" of fibers all over the brain. These fibers release the neuromodulator noradrenaline when the locus coeruleus is activated.