MICHAEL S. A. GRAZIANO

RETHINKING CONSCIOUSNESS

A Scientific Theory of Subjective Experience

W. W. NORTON & COMPANY
Independent Publishers Since 1923

CONTENTS

CHAPTER 1: The Elephant in the Room

CHAPTER 2: Crabs and Octopuses

CHAPTER 3: The Central Intelligence of a Frog

CHAPTER 4: The Cerebral Cortex and Consciousness

CHAPTER 5: Social Consciousness

CHAPTER 6: Yoda and Darth: How Can We Find Consciousness in the Brain?

CHAPTER 7: The Hard Problem and Other Perspectives on Consciousness

CHAPTER 8: Conscious Machines

CHAPTER 9: Uploading Minds

APPENDIX: HOW TO BUILD VISUAL CONSCIOUSNESS

NOTES

INDEX

RETHINKING CONSCIOUSNESS

CHAPTER 1

The Elephant in the Room

When my son was 3 years old, I made his favorite stuffed elephant talk. At that age he couldn't tell how bad a ventriloquist I was, so the trick worked very well on him. He loved it. Over the next several years, as I improved my technique, the uncanny power of that illusion began to impress me. Ventriloquism is not just a voice that comes out of a puppet, as though out of a hidden speaker. Even in the hands of a mediocre performer like me, something special happens. The puppet comes to life with its own personality, and consciousness seems to emanate from it.

The human brain clearly must contain machinery that impels us to attribute consciousness to the puppet. But we didn't evolve that machinery to enjoy ventriloquism. Humans are social animals, and we routinely use the same trick on each other. When I talk to someone, I have an automatic impression of thoughts, emotions, and awareness emanating from that person. I'm not directly perceiving the person's mind, of course. Instead, my brain is constructing a handy model of a mind and projecting it onto the person, treating that person like my son treats the puppet.

We apply the same process to more than just people. We attribute awareness to our pet cats and dogs, and some people even swear that their houseplants are conscious. The ancients felt sure that trees and rivers were sentient; children perceive consciousness in their favorite toys; and heck, the other day I got mad at my computer. So I'm not talking about intellectually figuring out whether something has a mind or cleverly deducing what might be in that mind—although we do that, too. I'm talking about an automatic, gut intuition, which is often wrong but sometimes persuasively potent, that an essence of awareness is emanating from an object.

As I thought more about ventriloquism, I began to wonder if my own consciousness and these examples of attributing consciousness to others might stem from the same source. Maybe there is one unifying explanation: we automatically build models of minds and project them onto ourselves and other people. Our intuitions about a mysterious conscious presence, our conviction that it is present in me or you or this pet or that object, might depend on those simplified but useful models—sets of information that the brain constructs to understand its world.

This is the kind of profound insight that can only come from talking to a stuffed elephant. It also diverted my scientific work to the study of consciousness.

For 20 years, I had been studying more traditional issues in neuroscience—how the brain monitors the space immediately around the body and how it controls complex movements within that space.[1] That background in basic, nuts-and-bolts neuroscience turned out to be useful for building a theory of consciousness. In 2010, my colleagues and I began outlining what we called the attention schema theory, drawing on data from neuroscience, psychology, and evolution and adding insights from engineering.[2] The theory is part of a larger change of perspective in the scientific community.[3] The new approach does not solve the so-called "hard problem" of consciousness—how a physical brain can generate a nonphysical essence.[4] Instead, it explains why people might mistakenly think that there is a hard problem to begin with, why that mistaken intuition is built deep into us where we're unlikely to change it, and why its presence is advantageous, maybe even necessary, for the functioning of the brain.

I first understood the theory from the point of view of social interaction. At its root, however, the theory depends on a more general property of the brain: model-based knowledge.[5] The brain constructs internal models—ever-changing rich packets of information, constructed continuously and automatically, like bubbles of meaning that lie beneath the level of higher thought or of language. Those internal models represent important items that are useful to monitor, sometimes external objects and sometimes aspects of the self. The representations are simplified and distorted, like impressionistic or cubist paintings of reality, and we report the content of them as though we are reporting literal reality. We can't help it—they come built into us. Our intuitive understanding of the world around us and our understanding of ourselves, always distorted and simplified, are dependent on those internal models.

In the theory, our metaphysical intuitions about ourselves, about consciousness as a nonphysical inner essence—sometimes called the "ghost in the machine"[6]—are derived from a particular internal model. I call it the attention schema, for reasons that will become clear throughout this book. It is a simplified depiction of how the brain seizes on information and deeply processes it. That depiction is an efficient way for the brain to understand and monitor its own internal abilities. The same kind of internal model can also be used, to a lesser extent, to monitor and make predictions about other people.

This model-based approach can sometimes sound like a dismissal or a devaluing of consciousness—but it is decidedly not. The internal model that tells us we are conscious is deep, rich, continuous, and probably necessary. Almost nothing we do—perceiving, thinking, acting, socially interacting—would work properly without that part of the system.

In this book, I will use the terms *consciousness*, *subjective awareness*, and *subjective experience* interchangeably, although I acknowledge that those words are not always used by scholars in an equivalent way. The word *consciousness* is especially notorious for its many slippery connotations. I want to clarify, first, what I *don't* mean by it, before I get to what I do mean. Sometimes people think of consciousness as the ability to know who you are and to understand your trajectory through life. Other people think of it more as the ability to process the world around you and on that basis to make intelligent decisions. I mean neither of those things.

The most common understanding of the internal experience is probably the "stream of consciousness," the constantly changing, kaleidoscopic contents of the mind. It's that riot of stuff in your head that James Joyce famously captured in his 1922 novel *Ulysses*.[7] Joyce meticulously recorded the ever-changing sight and sound and touch of the world, the taste and smell, the memories of the recent and distant past boiling up, a running internal dialogue, the conflicting emotions and fantasies, some of them so scandalous that the book was initially banned. (The 1933 court case, *The United States v. One Book Called Ulysses*, gave us our modern legal definition of obscenity.) But again, this is not what I mean by consciousness. That stream of material is not very well defined, and its sheer volume is overwhelming to study scientifically.

Instead, imagine putting 10,000 odds and ends in a bucket. You can catalog that complicated list of items, as James Joyce did. But you can also ask a more basic question: what about the bucket? Never mind the contents for now. What is the bucket made of, and where did it come from? How does a person get to be conscious of anything at all? Consciousness can't be just the information inside us, because we're conscious of only a small amount of the huge pool of information in the brain at any one time. Something must happen to a limited amount of information to make us conscious of it. What makes that happen? That more specific question has increasingly occupied philosophers and scientists.[8] The term *consciousness* has come to mean the act of being conscious of something, rather than the material of which you are conscious.

I suspect that the gradual shift in philosophy from focusing on the many items in a stream of consciousness to the act of *being conscious* has something to do with the advance of computer technology over the past half-century. As our information technology has improved, the information content of the mind has become less mysterious, while at the same time the act of being conscious of it, of experiencing anything at all, has become more remote and seemingly unsolvable. Let's look at a few examples.

You can connect a digital camera to a computer and program the system to process the incoming visual information. The computer can extract color, shape, and size, and it can identify objects. The human brain does something similar. The difference is that people also have a subjective experience of what they see. We don't just register the information that the object is red; we have an *experience* of redness. Seeing *feels* like something. A modern computer can process a visual image, but engineers have not yet solved how to make the computer conscious of that information.

Now consider something a little more personal than visual perception: the autobiographical memories that define your trajectory through life. Nothing typifies the Joycean stream of consciousness so much as the memories that are constantly bubbling up. And yet we know how to build a machine that stores and retrieves memory. Every computer has the capacity, and scientists know the general principles, if not the details, of how memory is stored in the brain. Memory is not a fundamental mystery. It also doesn't cause consciousness. Again, the stuff *in* consciousness—in this example, a memory—is not the same as the act of *being conscious* of a memory.

I'll give one more example: decision-making. If anything defines the mystery of human consciousness, surely it must be our ability to make decisions. We take in information, process it, judge it, and make a choice about what to do next. But, again, I would say that consciousness is not an intrinsic part of decision-making. All computers make decisions. In a sense, that's the definition of a computer. It takes in information, manipulates it, and uses it to select one course of action out of many. Most of the decisions made by the human brain, possibly tens of thousands a day, occur automatically with no subjective experience. In a few select instances, we report a subjective awareness of making a decision. Sometimes we call it intention, choice, or free will. But the mere ability to make a decision does not require consciousness.

With these and many other examples, the rise of computer technology has revealed the distinction between the content of consciousness, which is increasingly well understood at an engineering level, and the act of being conscious of it. My interest lies in that crucial, second part of the puzzle: how do we get to have a subjective experience of anything at all?

Sometimes people find that focus limiting. I've often been asked: What about memory? What about conscious choice? What about self-understanding? What about intentions and beliefs? Aren't these things the bread and butter of consciousness? I agree: all of these are important concerns and are crucial objects in the bucket of human consciousness. Still, they are not fundamental mysteries. They are matters of information processing, and we can imagine, at least in principle, how to engineer them. The fundamental mystery is the bucket itself. What is consciousness—what is it made of? How can something enter it, what is gained by entering it, and why do so few items in the brain end up there?

Traditionally, scholars assumed that something as amorphous and slippery as consciousness must be impossible to understand scientifically. But given recent insights, I am now pretty sure that it is as understandable and buildable as visual processing, memory, decision-making, or any other specific item that makes up its content.

I'VE WRITTEN ABOUT consciousness many times before. This book, however, is written entirely for the general reader. In it, I attempt to spell out, as simply and clearly as possible, a promising scientific theory of consciousness—one that can apply equally to biological brains and artificial machines.

The next few chapters start with evolution. Beginning more than half a billion years ago with the appearance of neurons, the cells that make up the brain, I'll describe the evolving complexity of the nervous system. Along the way I'll introduce the components of the attention schema theory, and by Chapter 6, the main scaffold of the theory will be in place.

I'll then discuss how the attention schema theory makes contact with other theories. The attention schema theory is one of about half a dozen main theories of consciousness that are gaining ground in the scientific literature. My impression, and the impression I try to convey in this book, is that these theories should not always be viewed as rivals, and we should not wait to see which one kills off its competitors. As different as they are—and I do disagree with a lot of what has been proposed—these many theories can also have strange, hidden connections to each other. Each one contributes important insights. I believe we are beginning to see the glimmerings of a consensus view—or maybe more like a consensus web of ideas.

The final chapters take a deep dive into the technological implications. We are close to understanding consciousness well enough to build it, and when we do, the new technology is likely to transform our civilization. Machine consciousness is just the first step. If consciousness is engineerable, then the mind is, in principle, migratable from one device to another. Though much farther down the road, it is theoretically possible to read the relevant data from a human brain and migrate that person's mind to an artificial platform.[9] The technology could allow minds to live indefinitely and to explore environments, such as interstellar space, that are hostile to biological bodies. No laws of physics stand in the way—only gadgetry that has yet to be invented.

If consciousness can be understood from a scientific and engineering perspective, then the topic is no longer just a philosophical game for scholars. It becomes an urgent practical matter. This book will follow the uses of consciousness to many possible technological futures, some good and some admittedly horrible. But good or bad, I am now pretty sure that a scientific understanding of consciousness and an ability to engineer it artificially are rapidly approaching.

CHAPTER 2

Crabs and Octopuses

SELF-REPLICATING, BACTERIAL LIFE first appeared on Earth about 4 billion years ago. For most of Earth's history, life remained at the single-celled level, and nothing like a nervous system existed until around 600 or 700 million years ago (MYA). In the attention schema theory, consciousness depends on the nervous system processing information in a specific way. The key to the theory, and I suspect the key to any advanced intelligence, is attention—the ability of the brain to focus its limited resources on a restricted piece of the world at any one time in order to process it in greater depth. In this and the next several chapters, I'll examine how attention may have evolved from early animals to humans and how the property we call consciousness may have emerged along with it.[1]

I will begin the story with sea sponges, because they help to bracket the evolution of the nervous system. They are the most primitive of all multicellular animals, with no overall body plan, no limbs, no muscles, and no need for nerves. They sit at the bottom of the ocean, filtering nutrients like a sieve. And yet sponges do share some genes with us, including at least 25 that, in people, help structure the nervous system.[2] In sponges, the same genes may be involved in simpler aspects of how cells communicate with each other. Sponges seem to be poised right at the evolutionary threshold of the nervous system. They are thought to have shared a last common ancestor with us between about 700 and 600 MYA (see the time line in Figure 2.1).[3]

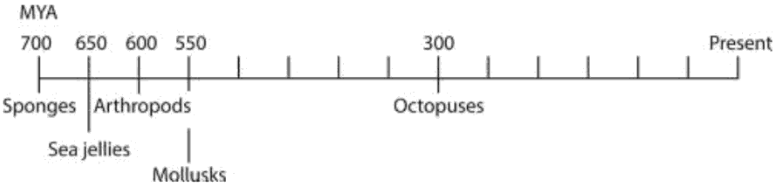

Figure 2.1 Invertebrates discussed in this chapter and their approximate time of first appearance.

In contrast, another ancient type of animal, the sea jelly, does have a nervous system. Sea jellies don't fossilize very well, but by analyzing their genetic relationship to other animals, biologists estimate that they may have split from the rest of the animal kingdom as early as 650 MYA.[4] These numbers may change with new data, but as a plausible, rough estimate, it seems that neurons, the basic cellular components of a nervous system, first appeared in the animal kingdom somewhere between sponges and sea jellies, a little more than half a billion years ago.

A NEURON IS, in essence, a cell that transmits a signal. A wave of electrochemical energy sweeps across the membrane of the cell from one end to the other, at about 200 feet per second, and influences another neuron, a muscle, or a gland. The earliest nervous systems may have been simple nets of neurons laced throughout the body, interconnecting the muscles. Hydras work on this nerve-net principle.[5] They are tiny water creatures—transparent, flowerlike animals with sacs for bodies attached to many arms—and belong to the same ancient category as sea jellies. If you touch a hydra in one place, the nerve net spreads the signals indiscriminately, and the hydra twitches as a whole.

A nerve net doesn't process information—not in any meaningful sense. It merely transmits signals around the body. It connects the sensory stimulus (a poke on the hydra) to a muscle output (a twitch). After the emergence of the nerve net, however, nervous systems rapidly evolved a second level of complexity: the ability to enhance some signals over others. This simple but powerful trick of signal boosting is one of the basic ways that neurons manipulate information. It is a building block of almost all computations that we know about in the brain.

The eye of the crab is one of the best-studied examples.[6] The crab has a compound eye with an array of detectors, each with a neuron inside it. If light falls on one detector, it activates the neuron inside. So far so good. But in an added pinch of complexity, each neuron is connected to its nearest neighbors, and because of those connections, the neurons compete with each other. When a neuron in one detector becomes active, it tends to suppress the activity of the neurons in the neighboring detectors, like a person in a crowd who is trying to shout the loudest while shushing the people nearest to him.

The result is that if a blurry spot of light shines on the crab's eye, with the brightest part of the spot hitting one detector, the neuron in that detector becomes highly active, wins the competition, and shuts down its neighbors. The pattern of activity across the set of detectors in the eye not only signals a bright spot, but also signals a ring of darkness around it. The signal is, in this way, enhanced. The crab eye takes a fuzzy, gray-scale reality and sharpens it into a high-contrast image with exaggerated, brighter peaks and darker shadows. This signal enhancement is a direct consequence of neurons inhibiting their neighbors, a process called *lateral inhibition*.[7]

The mechanism in the eye of a crab is arguably the simplest and most fundamental example—the model A case—of attention. Signals compete with each other, the winning signals are boosted at the expense of the losing signals, and those winning signals can then go on to influence the animal's movements. *That* is the computational essence of attention. Our human attention is merely an elaborated version of it, made of the same building blocks. You can find the crab-eye method of lateral inhibition at every stage of processing in the human nervous system, from the eye to the highest levels of thought in the cerebral cortex. The origin of attention lies deep in evolutionary time, more than half a billion years ago, with a surprisingly simple innovation.

Crabs belong to an extensive group of animals, the arthropods, which includes spiders and insects and other creatures with hard, jointed exoskeletons and which branched off from other animals about 600 MYA.[8] The most famous extinct arthropod, the one with the biggest fan club today, is the trilobite—a leggy, jointed creature almost like a miniature horseshoe crab, which crawled about the bottom of Cambrian seas as early as 540 MYA. When trilobites died and sank into very fine silt on the ocean floor, their faceted eyes were sometimes fossilized in amazing detail.[9] If you look at a trilobite fossil and examine its bulging eyes through a magnifying glass, you can often still see the orderly mosaic of individual detectors. Judging from these fossilized details, the trilobite's eye must have closely resembled a modern crab's eye in its organization and is likely to have used the same trick of competition between neighboring detectors to sharpen its view of the ancient seabed.

Imagine an animal built piecemeal with "local" attention. In that animal, each part of the body would function like a separate device, filtering its own information and picking out the most salient signals. One of the eyes might say, "This particular spot is especially bright. Never mind the other spots." Meanwhile, independently, one of the legs says, "I've just been poked hard right here. Ignore the lighter touches nearby!" An animal with only this capability would act like a collection of separate agents that happen to be physically glued together, each agent shouting out its own signals, triggering its own actions. The animal's behavior would be, at best, chaotic.

For a coherent response to its environment, the animal needs a more centralized attention. Can many separate sources of input—the eyes, the body, the legs, the ears, the chemical sensors—pool their information together in one place for a global sorting and a competition among signals? That convergence would allow the animal to select the most vivid object in its environment, the one that seems most important at the moment, and then generate a single, meaningful response.

Nobody knows when that type of centralized attention first appeared, partly because nobody is certain which animals have it and which ones don't. Vertebrates have a central attention processor, which I'll describe in the next chapter. But the mechanisms of attention have not been as thoroughly studied in invertebrates. Many types of animals, such as segmented worms and slugs, do not have a central brain. Instead they have clusters of neurons, or ganglia, scattered throughout their bodies to perform local computations.[10] They probably don't have centralized attention.

Arthropods, such as crabs, insects, and spiders, are better candidates for centralized attention. They have a central brain, or at least an aggregate of neurons in the head that is larger than any of the others in their bodies.[11] That large ganglion may have evolved partly because of the requirements of vision. The eyes being in the head, and vision being the most complicated and information-intensive sense, the head gets the largest share of neurons. Some aspects of smell, taste, hearing, and touch also converge on that central ganglion. Insects are brainier than people think. When you swat at a fly and it manages to escape—as it almost always does—it isn't just darting away on a simple reflex. It probably has something that we can call central attention, or the ability to rapidly focus its processing resources on whatever part of its world is most important at the moment, in order to generate a coordinated response.[12]

Octopuses are the superstars of the invertebrates because of their astonishing intelligence. They're considered mollusks, like clams or snails. Mollusks probably first appeared about 550 MYA and remained relatively simple, at least in the organization of their nervous systems, for hundreds of millions of years.[13] One branch, the cephalopods, eventually evolved a complex brain and sophisticated behavior and may have reached something close to the modern form of an octopus around 300 MYA.[14]

Octopuses, squid, and cuttlefish are true aliens with respect to us.[15] No other intelligent animal is as far from us on the tree of life. They show us that big-brained smartness is not a one-off event, because it evolved independently at least twice—first among the vertebrates and then again among the invertebrates.

Octopuses are excellent visual predators. A good predator must be smarter and better coordinated than its prey, and using vision to locate and recognize prey is especially computationally intensive. No other sensory system has such a fire hose of varied information pouring in and such a need for an intelligent way to focus on useful subsets of that information. Attention, therefore, is the name of the game for a visual predator. Maybe that lifestyle has something to do with the expansion of octopus intelligence.

Whatever the reason, the octopus evolved an extraordinary nervous system. It can use tools, solve problems, and show unexpected creativity.[16] In a now classic demonstration, octopuses can learn to open a glass jar by unscrewing the top in order to get to a tasty morsel within. The octopus has a central brain and also an independent, smaller processor in each arm, giving it a unique mixture of centralized and distributed command.[17] The octopus also probably has self models—rich, constantly updated bundles of information to monitor its body and behavior. From an engineering perspective, it would need self models to function effectively. For example, it might have some form of a body schema that keeps track of the shape and structure of its body in order to coordinate movement. (Perhaps each arm has its own arm schema.) In that sense, you could say that an octopus knows about itself. It possesses information about itself and about the outside world, and that information results in complex behavior.

But all of these truly wonderful traits do not mean that an octopus is conscious.

Consciousness researchers sometimes use the term *objective awareness* to mean that the information has gotten in and is being processed in a manner that affects behavioral choice.[18] In that rather low-bar definition, one could say that a microwave is aware of the time setting and a self-driving car is aware of the looming obstacle. Yes, an octopus is objectively aware of itself and of the objects around it. It contains the information.

But is it *subjectively* aware? If it could talk, would it claim to have a subjective, conscious experience the same way that you or I do?

Let's ask the octopus. Imagine a somewhat improbable thought experiment—and remember the experiment, because it will come in handy throughout this book. Suppose we've gotten hold of a crazy science fiction device—let's call it the Speechinator 5000—that serves as an information-to-speech translator. It has a port that can be plugged into the octopus's head, and it verbalizes the information found in the brain.

It might say things like "There is a fish" if the octopus's visual system contains information about a nearby fish. The device might say, "I am an entity with a bunch of limbs that move in this and that way." It might say, "Getting a fish out of a jar requires turning that circular part." It would say many things, reflective of the information that we know is contained inside the octopus's nervous system. But we don't know if it would say, "I have a subjective, private experience—a consciousness—of that fish. I don't just process it. I *experience* it. Seeing a fish *feels* like something." We don't know if its brain contains that type of information because we don't know what the octopus's self models tell it. It may lack the machinery to model what consciousness is or to attribute that property to itself. Consciousness could be irrelevant to the animal.

The octopus conundrum is an instructive example of how an animal can be complex and intelligent, and yet we are, so far, unable to answer the question of its subjective experience or even whether the question has any meaning for that creature.

Maybe one source of confusion here is the automatic and powerful human urge to attribute consciousness to the objects around us. As I pointed out in Chapter 1, we are prone to see consciousness in puppets and other, even less likely objects. People sometimes believe that their houseplants are conscious. An octopus, with its richly complex behavior and its large eyes filled with focused attention, is a far more compelling inkblot test, so to speak, triggering a strong social perception in us. Not only do we know, intellectually, that it gathers objective information about its world, but we can't help feeling that it must have a subjective awareness as well emanating out of those soulful eyes. But the truth is, we don't know, and the sense we get of its conscious mind says more about *us* than about the octopus. The experts who study octopuses risk becoming the least reliable observers on this point, because they are the ones most likely to be entranced by these wonderful creatures. Later, in Chapter 5, I'll return to that pervasive human aspect of consciousness—how we use it as a tool in our social tool kit and reflexively attribute it to the agents around us.

Just to be clear, I'm not saying that octopuses are *not* conscious. But the octopus nervous system is still so incompletely understood that we can't yet compare its brain organization with ours and guess how similar it might be in its algorithms and self models. To make those types of comparisons, we will need to examine animals in our own lineage, the vertebrates.

CHAPTER 3

The Central Intelligence of a Frog

I GREW UP partly on a farm in upstate New York. Every summer, all night long, we'd hear a big bullfrog croaking out his mating song in the pond behind the house. We used to call him Elvis, and the smaller answering frog voice, Priscilla. I've been fond of frogs ever since, and when I became a neuroscientist, I was interested to read about what goes on inside their heads.

A frog has a part of the brain called the tectum. The word *tectum* means "roof" in Latin, and it's the largest, most obvious hump at the top of the brain. Frogs are not alone in this feature of the brain. The tectum may have been particularly studied in amphibians, but it is also present in fish, reptiles, birds, and mammals. All vertebrates have a tectum, but no other animals do, at least as far as we know. We can make a good guess that around half a billion years ago, a species of small, jawless fish, the common ancestor of vertebrates, evolved a tectum, and all its descendants inherited that brain part from it.[1]

People have a tectum, but ours is no longer at the top of the brain. It's a relatively tiny lump—or rather, two lumps, one on each side of the midline—buried beneath piles of brain structures that expanded in our evolutionary past. When found in people and other mammals, it's usually called the superior colliculus (which is

Latin for "upper bump"). Here, for simplicity, I will call it a tectum.

For most of vertebrate evolution, the tectum was the pinnacle of intellectual achievement—the most complex, computationally sophisticated processor at the center of the brain. In a frog, the tectum takes in visual information and sorts the world into a literal map.[2] Each point on the rounded surface of the tectum corresponds to a point in space around the animal. The tectum on the right side of the frog's brain contains a neat, orderly map of the field of view of the left eye; and similarly for the left tectum and the right eye. When an erratic black dot zigzags around the frog, the eyes take in that information, the optic nerve sends the signals to the tectum, and the tectum triggers a set of muscle controllers. As a result, the tongue shoots out with impressive accuracy to snag the bug.

The logic of this input-output device was demonstrated particularly vividly by the neuroscientist Roger Sperry. In the 1960s, he performed surgery on a frog, removing the eyes, turning them upside down, and putting them back in.[3] The eyes took. Frogs have an amazing ability for regeneration. The optic nerve regrew from the eye to the tectum, reestablishing the internal visual map. When the frog was healed and could see again and a fly buzzed up above its head, its tongue shot down to the floor. If the fly buzzed to the frog's right, the tongue shot out to the left. The frog's central intelligence was a simple, beautifully efficient machine that collected specific nerve inputs and matched them with corresponding outputs. It was, unfortunately, tricked by scientific manipulation. The altered frog had to be fed by hand or it would have starved.

The frog's tectum is not limited to vision. It also collects information from the ears and from touch receptors across the skin.[4] A map of the frog's body surface, of the auditory space around the animal, and of visual space converge and are partly integrated in the tectum. It's the highest level of integration in the amphibian brain—the central processor that pulls together scattered signals pouring in from the environment, focuses on the most important event occurring at each moment in time, and then triggers a response.[5] The tectum is the frog's central attention mechanism.

Scientists can probe the brain with astonishing precision, like a computer engineer probing a circuit board. One standard method involves an electrode—a hair-thin, stiff wire coated in plastic insulation except at the tip. Only about a tenth of a millimeter of bare wire is exposed. Like a miniature detection wand, it can pick up electrical activity within a microscopically short distance of the bare metal. A long, flexible wire extends from the back end of the electrode, connecting it to a rack of equipment. The electrode is usually clamped in place by precision machinery and moved into position, 1 micrometer at a time, to study a targeted brain area.

The setup is sensitive enough to measure the activity of individual neurons in the brain. When a neuron near the electrode tip fires off a signal to its neighbors, the device picks up that tiny electrical pulse. The signal is amplified and piped to a loudspeaker, and the experimenter hears a click. Normally, a neuron might fire one or two clicks a second in a random pattern, but if an event occurs that recruits the neuron, the cell may fire off a sudden burst of clicks at a rate of 100 or so a second. One of the most thrilling pastimes of a neuroscientist is listening in on the clicking of an individual neuron and wondering what role it plays in the brain.

Each neuron in the frog's tectum acts like a detector.[6] The neuron monitors a particular region of space—for example, an area directly above the head—and its rate of clicking increases when an object enters that space. The neurons vary—some prefer a visual stimulus moving in a particular way, some prefer a touch or surrounding water.

Pit vipers, such as rattlesnakes or moccasins, have their own version of infrared vision: a pair of specialized, heat-sensing organs located about halfway between their eyes and nostrils. Those organs send information to the tectum, which contains a map of heat signals overlaid on the more usual visual map of space.[11] The ability of the snake to orient its head toward prey, and the accuracy of its strike, is thought to depend on that multisensory map.

An owl's tectum has a visual map aligned with an auditory map.[12] When the bird hunts, it can aim its strike at the correct location either by the sight of its prey or, when hunting at night, by the sound of the animal rustling in the grass.

Stimulate a monkey's superior colliculus, and a swift, coordinated head-and-eye movement unfolds.[13] The monkey orients toward the mapped location in space. I don't know of any studies applying electrical stimulation to the human superior colliculus, but we are a species of primate and presumably have the same mechanism as in a monkey. When you turn to look at something, especially when an unexpected event causes you to orient in a fast, reflexive manner, your tectum is probably triggering that seemingly effortless, well-coordinated behavior.

All vertebrates use the tectum in more or less the same way, though with some extra bells and whistles, depending on the species. The brain area collects sensory information, picks out the most vivid event happening nearby, and orients the animal, physically pointing the sense organs toward that event.

That kind of orienting is sometimes called "overt attention."[14] It's a simple way to solve a fundamental problem: too much goes on in the environment for the brain to process it all. An animal needs to pick an item of interest and filter out the rest. If you can point your eyes and ears toward one object, then you will automatically filter out other, more peripheral events. The tectum

INDEX

Page numbers listed correspond to the print edition of this book. You can use your device's search function to locate particular terms in the text.

Note: Page numbers in italics indicate figures.

abstract thinking, 34
affordances, 51–52
agendas, 46
alchemy, 118–19
Alexa, 132
al-Haytham, Hasan Ibn, 62
Allen Institute, 147
amoebas, 37
amphibians, 25, 26, *45*
amygdala, 130
ancestor worship, 156, 157
animals, 1, 44–45, *45*, 50–51. *See also specific animals*
archosaurs, 28
arthropods, *9*, 10–12, *45*
artificial attention, 125–27
artificial consciousness, 118–37
artificial face recognition, 127
artificial intelligence, 7
  conscious machines, 118–37
  impact of, 152–61
  mind uploading, 138–66
Asimov, Isaac, "The Bicentennial Man," 134

attention, 8, 20–21, 34, 54–56
  artificial, 125–27
  awareness and, 109–13
  bottom-up, 33
  centralized, 11–12
  clutter and, 55, 56
  colloquial vs. scientific use of the word, 173–74
  consciousness and, 109–11
  control over, 21, 23
  covert, 21, 25, 35–36, 41–42, 43, *45*, 47
  evolution of, 8–15
  facial expressions and, 56, 57
  gaze direction and, 56, 57
  how it works, 55–56
  James's definition of, 109–10
  "local," 11
  neuroscientific definition of, 110
  overt, 20, 25, 35, 41–42, 44, *45*, 47
  saliency and, 55, 56, 57
  shifting focus of, 36
  tectum and, 34–35
  temporal dynamics of, 56, 57
  top-down, 33
attention-related networks, 81
attention schema, 21, 23–24, 39–43, 44, 48, 58–64, 81, 86–87, 174–76
  conscious machines and, 126, 127–28
  evolution of, 25
  hemispatial neglect and, 89
  as universal connector, 115–17
  vertebrates and, 23, 25
  attention schema theory, 2–3, 7, 8, 38, 43, 64, 80–81, 82
  awareness and, 109–13
  as engineering answer to alchemical mystery of consciousness, 119
  global workspace theory and, 105–7
  as guide for building conscious machine, 125–37
  higher-order thought theory and, 107–9
  illusionism and, 96–101
  integrated information and, 113–17

  machine consciousness and, 124–25
  other perspectives on consciousness and, 91–117
  phantom limbs and, 101–5
  practical benefits of, 119–20
attention schema zombies, 89–90
auditory cortex, 83
awareness
  attention and, 109–13
  attention schema theory and, 109–13
  objective, 14, 24–25, 169
  subjective, 3, 14, 15, 25, 169

Baars, Bernard, 105
bacterial life, self-replicating, 8
bees, 36–37
behavioral prediction, 51–60, 64
being conscious, 4–5, 6
beliefs, 46
biased competition, 33, 34
binocular rivalry, 67–69, 71, 72–73
birds, 20, 25, 28, *45*
  consciousness and, 44
  evolution of, 29
  intelligence of, 29–30
  as therapod dinosaurs, 29
  wulst of, 29–30
Blade Runner, 134
Blanke, Olaf, 104–5
blindness, 86
blindsight, 111–12
blobology, neuroscience and, 66–67
Blue Brain Project, 147
bodily self-knowledge. *See* body schema
body schema, 13, 22, 101–5, 174–76
body sensation, 131–32
bottom-up attention, 33
Botvinick, Matthew, 102
the brain, 1. *See also specific anatomy; specific animals*

First Edition

For information about special discounts for bulk purchases, please contact W. W. Norton Special Sales at specialsales@wwnorton.com or 800-233-4830

Jacket design by Kelly Winton
Jacket Image by Liu zishan / Shutterstock
Book design by Ellen Cipriano
Production manager: Lauren Abbate

The Library of Congress has cataloged the printed edition as follows:

Names: Graziano, Michael S. A., 1967– author.
Title: Rethinking consciousness : a scientific theory of
subjective experience / Michael S.A. Graziano.
Description: First Edition. | New York : W.W. Norton & Company, 2019. |
Includes bibliographical references and index.
Identifiers: LCCN 2019014794 | ISBN 9780393652611 (hardcover)
Subjects: LCSH: Consciousness. | Experience—Psychological aspects.
Classification: LCC B808.9 .G73 2019 | DDC 153—dc23
LC record available at https://lccn.loc.gov/2019014794

ISBN 9780393652628 (ebook)

W. W. Norton & Company, Inc., 500 Fifth Avenue, New York, N.Y. 10110
www.wwnorton.com

W. W. Norton & Company Ltd., 15 Carlisle Street, London W1D 3BS