CHAPTER 1

THINKING ON 20 WATTS

The Ultimate Scientific Challenge

Understanding how the brain works is almost certainly the most challenging scientific problem of our time. How can three pounds of tissue perform mental feats that outstrip the ability of the world's most powerful computers while consuming less energy than a dim lightbulb? Answering this question is the goal of neuroscientists, who study it at many different levels. Much of our current knowledge about how the brain works comes from studying other species, ranging from worms or fruit flies up to mammals like mice, rats, and monkeys. While this research has given us many important insights, most of us ultimately want to understand how the human brain works, and there are many aspects of the human mind that simply can't be studied in nonhuman animals: if we want to understand how humans think, we need to study humans.

This book will tell the story of how a set of new technologies has given us the ability to study how the human brain works in greater detail than ever before. These tools are known as *neuroimaging* methods, because they allow us to create images of the human brain that show us what it is made of (which we refer to as its *structure*) and what it is doing (which we refer to as its *function*). One tool in particular has revolutionized our ability to image the brain: magnetic resonance imaging (MRI). This incredibly versatile technique has provided neuroscientists with the ability to safely watch the human brain in action, which has allowed us to understand how the brain accomplishes many psychological functions. In some cases MRI can even allow us

to decode what people are experiencing or thinking about by looking at their brain activity when they are performing a task or simply resting—what some audaciously call "mind reading" but what is more accurately known as decoding. And the power of MRI is not limited to studying the brain only during a fixed point in time. MRI has also shown us how experiences change the brain, and how individual human brains change over time from childhood to old age. It has shown us that all human brains follow the same general plan, but there are also many differences between people, and these studies have given insights into the brain dysfunctions that lead to mental illness. In doing so, MRI has raised many new questions that go beyond science, ultimately addressing some of the fundamental questions about how we view ourselves as humans. If thinking is just a biological function that we can visualize with MRI, then what becomes of the mystery of human consciousness? If decisions emerge from the computations of the brain, then in what sense are "we" responsible for our choices? Is addiction a "brain disease," a failure of self-control, or both? Should we worry about the ability of marketing researchers to use brain imaging to more effectively sell us their goods? It is these kinds of questions that we will grapple with after providing an overview of both the power and the limitations of neuroimaging.

What Is Neuroimaging?

When I use the term "neuroimaging" I am referring generally to a set of techniques that allow us to look at the human brain from the outside. There are a number of different ways to do this, but I will focus mostly on MRI because it has become the most widely used tool for brain imaging owing to its safety and its flexibility. Different kinds of MRI scans can be used to measure many different aspects of the brain, and we will roughly group them into what we call *structural* and *functional* MRI. Structural MRI measures different aspects of the makeup of brain tissue, such as how much water or fat is present in the tissue. Because different parts of the brain contain different amounts of these substances, they will show up on the MRI image as brighter or

darker (see color plate 1). These aspects of the brain are very useful for detecting diseases of the brain, and for understanding differences in size and shape of different brain parts between people, but they don't tell us what the brain is *doing*—for that, we need to use functional MRI, or, as it is usually abbreviated, fMRI. fMRI came about when researchers discovered how to use MRI to detect the shadows of brain activity through its effects on the amount of oxygen in the blood. It is fMRI that provides the colorful images like the one shown on the right in color plate 1— in which parts of the brain seem to "light up." We will discuss the invention of fMRI and how it works in much more detail in chapter 2. First, we need to ask: What does "brain function" mean?

The Brain as a Computer

Each of the body's organs has evolved to serve a particular biological function: the heart pumps blood, the lungs oxygenate the blood, the digestive system extracts nutrients from food, and the kidneys filter waste products from the blood. What is the biological function of the brain? Whatever that function is, it's clearly very important—the brain accounts for only about 2% of the body's weight, but it uses about 20% of the energy consumed by the body.[1] If I had to come up with a simple label for what the function of the brain is, I would say: it processes information. Certainly not in the same way that your laptop or smartphone processes information, but nonetheless we can think of the brain as our body's central computing system, extracting information from the world and using it to choose how to act, with the goals of living long, prospering, and (most importantly for evolution) reproducing. The brain is not the body's only computer—for example, the gut has its own as well, known as the *enteric nervous system*, with about half a billion neurons—but it's certainly the most important when it comes to the things that make us uniquely human.

While it might make sense to call it a "computer," the brain is definitely not like most computers that we are familiar with in the world. Those silicon-based computers follow the recipe

attributed to John von Neumann (one of the first true "computer scientists")—they are constructed from a large number of elements, which you can think of as microscopic switches, that behave in a highly reliable and consistent way. Anyone who ever experienced the "blue screen of death" on a personal computer knows what happens when one of these elements malfunctions; digital computers just aren't very resilient. These switches also operate very quickly. The computer that I am using to write this book has a clock speed of three gigahertz, which means that it can perform three billion operations every second, and thus those little switches have to be able to turn on and off very quickly. What's even more impressive is that those operations are happening largely "in serial," meaning that they are happening one at a time (or a few at a time in the case of most current computers).

How does the brain differ from a digital computer? To answer this, we need to understand how the cells in the brain process information, so it's time for a whirlwind tour of the physiology of the brain. The brain is made largely of two types of cells. Neurons are the cells that we have traditionally thought are essential to the brain's computing power. There is a second class of cells, known as glia, and these give the brain its structural scaffold and provide biological support to neurons. Until recently it was thought that glia were just supporting actors, but it's becoming more clear that they also play an important role in information processing.[2] However, throughout this book we will focus on the activity of neurons, since they are still the main type of cell studied by neuroscientists.

To understand how neurons work, let's trace the path of a signal from the world to the brain. When I brew my morning espresso and take a sniff, the smell that I experience starts with molecules from the coffee contacting my olfactory bulb, one of the only parts of the brain that is exposed directly to the outside world (right inside our noses). Those molecules hit a special type of neuron called an olfactory receptor, and cause changes in its cell membrane that increase the electrical charge within the cell. When this increase reaches a certain level, the cell suddenly changes its electrical properties, releasing what is

called an *action potential*—a very sudden and large increase in its electrical charge. In common parlance, we say that the neuron "fired" or "spiked," because the change is so sudden. When this happens, the action potential travels down the length of the neuron and is ultimately transmitted to the other neurons that it connects to, and the cycle starts over—if the next neuron gets a strong enough input then it too will fire, and so on. If it's strong enough, the signal from the olfactory receptors will cause a cascade of activity between connected neurons that will ultimately reach my cerebral cortex, possibly triggering memories of the daily trips to the espresso bar on my first visit to Italy, or the desire for a piece of chocolate or pastry to go with it.

When we compare neurons in the brain to the central processing unit (CPU) in a digital computer, there is one important similarity. Most neurons behave in an "all-or-none" manner—just like a digital transistor, a neuron either fires or it doesn't, and all of its action potentials are basically the same in terms of their size and timing. This means that my olfactory neurons don't signal that the coffee smell is stronger by firing larger action potentials, but rather by firing more of them in succession or firing them more rapidly (figure 1.1). However, in almost every other way the brain computes very differently from a digital computer. First, brains are very slow compared to a digital computer—we measure the speed of computer operations in nanoseconds (billionths of a second), whereas the speed of firing of neurons is on the order of milliseconds (thousandths of a second). Second, individual neurons are noisy and unreliable. Of the many millions of neurons that are sensitive to the molecules in coffee, a different subset of them is going to fire each time I smell coffee. Third, brains process information in a highly parallel manner—rather than doing a few things at once very quickly like a CPU, the brain does lots of things at the same time, but does each one relatively slowly.

All of these features of the brain add up to a very different kind of computer, but that's a good thing. Most importantly, brains are resilient. If you drop your laptop and damage the motherboard, it's very unlikely that it will ever work again, and it certainly is not going to fix itself. The brain, on the other hand,

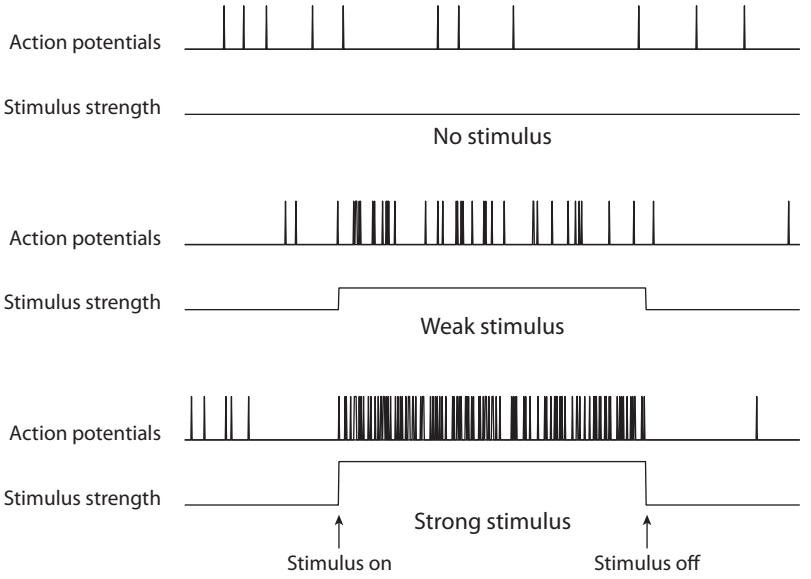

Figure 1.1. A simulated example of how an individual neuron responds to stimulation. The *top panel* shows action potentials as small spikes in the line, occurring even in the presence of no stimulation; most neurons fire randomly on occasion even when they are not stimulated. The *middle panel* shows the neuron's response to a weak stimulus, and the *bottom panel* shows the response to a strong stimulus. Note that the size of the action potential does not grow as the stimulus gets stronger; it simply fires more frequently.

is remarkably robust. Take the case of Lisa, whom I met when I was a postdoctoral fellow many years ago.[3] Lisa grew up as a relatively normal child, but around age 12 started suffering from severe epileptic seizures. Ultimately the seizures were so life-threatening and uncontrollable that at age 16 her doctors turned to a last-resort treatment known as hemispherectomy, which involves removing one entire hemisphere of the brain—fully half of her cerebral cortex (see figure 1.2). The seizures arose from her left hemisphere, which in most people is the side of the brain that is largely responsible for language function. Unfortunately this was the case for Lisa as well, and for the first year after her surgery she barely spoke at all. We studied her about six years after her surgery, at which point she had regained a remarkable amount of language function—far from normal for a 22-year-old, but nonetheless amazing given the fact

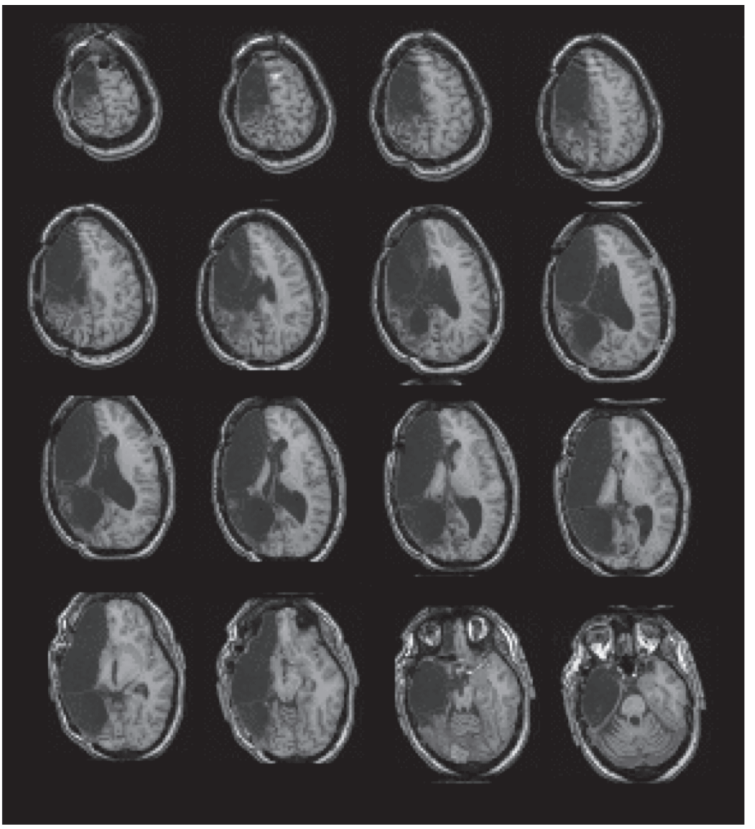

Figure 1.2. An MRI scan of Lisa's brain, showing that her left hemisphere is mostly missing, replaced by cerebrospinal fluid.

that her entire left hemisphere had been removed. This doesn't mean that half a brain is enough—after all, she was left with profound language problems—but it shows the amazing ability of the brain to recover from injury. We will return to the case of Lisa in chapter 5, where I discuss how neuroimaging allowed us to see how her brain had reorganized itself to allow her to speak and read.

There is another important way in which brains and digital computers differ. When you buy a personal computer, you have the choice of many different operating systems (such as Windows, Linux, or Mac OS) as well as a mind-numbing choice of software programs. This is because the software is fundamentally separate

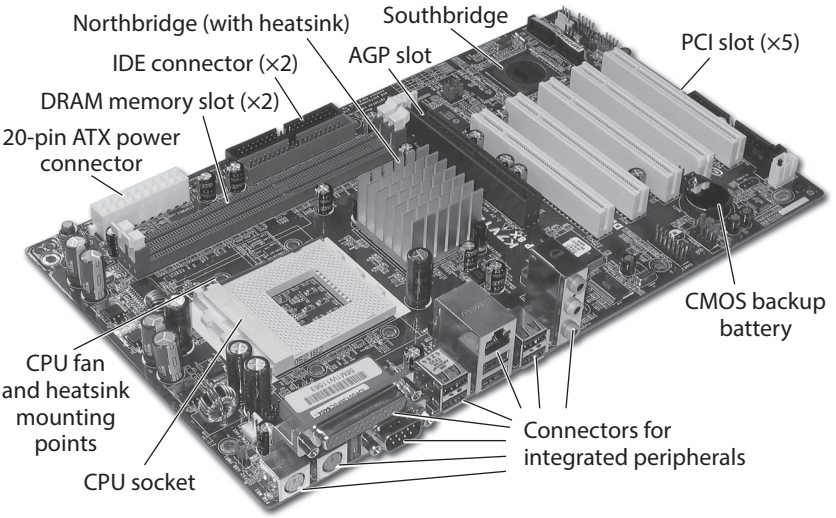

Figure 1.3. A computer motherboard with different parts labeled. Unlike the human brain, a computer is highly modular, with various parts that perform different specialized functions. By user: Moxfyre (Creative Commons license CC BY-SA 2.5 [https://creativecommons.org/licenses/by-sa/2.5]), via Wikimedia Commons.

from the hardware, for which we can also thank John von Neumann. In the brain, on the other hand, the hardware and software are inseparable; the "program" is stored in the connections between the neurons, and when we learn, it happens through changes in those connections. That is, the brain actually changes the configuration of its own hardware. We will discuss this further when we talk about brain plasticity, because it is these changes in connections between neurons that are crucial to the brain's ability to learn as well as its resilience to injury and disease.

Finally, it's important to understand how the architecture of the brain differs from the digital computers that we are familiar with, where by "architecture" I mean how the different parts function as part of the whole. Digital computers are built in a modular fashion, meaning that different parts have different specialized functions. Figure 1.3 shows the motherboard of a modern computer, with many of the different sections labeled. There are different parts of the motherboard dedicated to sound,

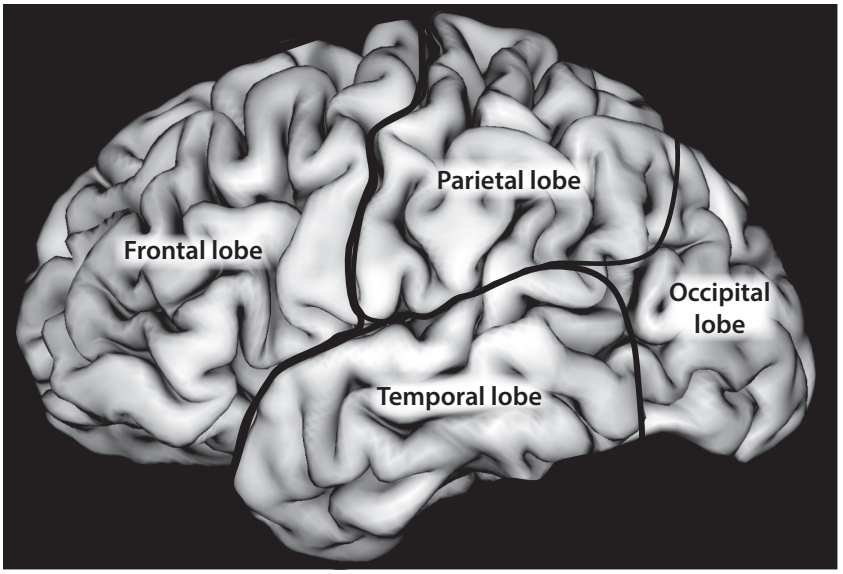

Figure 1.4. The lobes of the brain, outlined on an image of the surface of my brain that was generated from MRI data.

networking, memory, and the CPU, along with many other functions. We can tell that this system is modular in two ways. First, for many of the parts we can simply remove and replace them with a different version, as long as they are compatible. If a faster CPU or better video card comes along, I simply install it and start the computer up again, and with any luck it will just work. Second, damage to one of the parts will often have very specific effects; if I were to carefully damage the network chip (making sure not to damage any other parts), the sound functions of the computer should still work, and vice versa.

The non-neuroscientist could be forgiven for thinking that brains are also modular. After all, we regularly hear stories about neuroimaging studies that talk about "the reward center" or "the face area" in the brain. As I will discuss further below, there is a grain of truth in these stories, which is that functions are localized to some degree in the brain. People who have a stroke in their left prefrontal lobe are much more likely to have a language problem, whereas those who have a stroke that damages their right parietal lobe (see figure 1.4) are more likely to have difficulties with

spatial awareness and attention. However, neuroscientists now realize that no single brain area does its work alone—there is no analog to the sound chip or video card in the brain. You should instead think of the brain more like a construction team. There are lots of specialist subcontractors (experts in perceiving speech, or finding locations in space, or predicting how another person will behave), along with a horde of general contractors who try to keep everything on track. However, no one of these individuals can build the building alone—it's the combination of all of them working together that creates the final product. As we will discuss in chapter 3, studies of how different regions of the brain communicate with one another have given us many new insights into how the brain functions as an integrated network rather than just a collection of specialists.

What Does the Brain Compute?

Now that we know a bit about how the brain computes, we are still left with a question: What return do we get on our outsized investment of energy in the brain? The short answer is that we get the ability to adapt. Just as humans can thrive on an amazing range of diets (from Inuits living on seal meat and whale blubber to Pacific islanders eating a diet of starchy vegetables like taro root), so too can the human brain adapt to a very wide range of cognitive environments and challenges. There are many organisms in the world that are adapted to function well within very specific niches, but few can function in as broad a range of environments as humans.

We think that one of the keys to the adaptive nature of the brain is its ability to build a predictive model of the world as it unfolds. It's easy to take for granted just how many predictions we are constantly making about the world around us. As I walk down the sidewalk, there is no guarantee that the concrete won't suddenly become liquid and swallow me, but I don't think twice as I take that next step; I assume that past experience is a good guide for the future. Similarly, when I ask one of my colleagues for advice on a problem, I assume that he or she will respond to

me in English prose and not by singing operatically in Italian. Neither of these is guaranteed, but they are both pretty good bets. It's only when our predictions are violated that we realize just how beholden we are to them.

A large body of research in the past few decades has shown that the brain is constantly making predictions about the world, and updating those predictions when they are wrong. In fact, it's these violations that are at the center of learning; if we are behaving perfectly and the world abides by our expectations exactly, then why change anything? The neurotransmitter dopamine is one of the keys that relates learning to prediction errors. Dopamine is a different kind of neurotransmitter than the ones that send signals between specific neurons. Instead, we call it a "neuromodulator" because it changes the way that other neurons act, rather than causing them to fire directly. The dopamine neurons in the brain, which reside very deep in the middle of the brain, signal whenever something happens in the world that is unexpected; this could be something novel (imagine a loud noise in the library), or something that violates our expectations, either in a good way (such as finding a $100 bill on the sidewalk) or a bad way (such as finding out that your paycheck was smaller than expected). Dopamine neurons are constantly telling the rest of the brain how good the world is in comparison to our predictions, turning up their activity when the world exceeds our expectations and turning down when the world disappoints us. One of the major successes in neuroscience has been the development of a theory that links the role of dopamine in coding of "prediction errors" to our ability to adaptively improve our predictions, through a process called "reinforcement learning." We will discuss this in more detail in chapter 7 in the context of how the brain makes decisions.

From Brain to Mind

The brain is something that we can all get our heads around—it's a piece of tissue that we can see, measure, and if we are neurosurgeons, touch. But what is the mind? We all intuitively

know what it feels like to have a mind, but what is it made of? The historical answer to this question was that the mind is on another plane. The philosopher Descartes famously argued that the mind (by which he really meant "soul") makes contact with the physical world in the pineal gland, a small structure in the middle of the brain. But how could something nonphysical interact with the physical world? That's the question that has perpetually dogged the concept of dualism—the idea that the mind is not of the physical world.

Dualism has not fared well in the age of neuroscience. The more we learn about the brain's workings, the clearer it is that our minds really are one and the same with our brains. An amazing example of this comes from research using electrical brain stimulation. It's common for researchers to stimulate and record from the brains of nonhuman animals, but those animals can't directly tell us about their experiences. In rare cases, however, researchers are able to examine the effects of direct electrical stimulation in the human brain. When someone suffers from severe epilepsy, surgeons will sometimes try to remove the part of the brain from which the seizures arise—in the case of Lisa this involved an entire hemisphere, but often a small part can be removed, which has a much less detrimental impact on the person's function. The goal is to remove only the part of the brain that causes the seizure, but it's usually not possible to tell which part that is from the outside, so surgeons will sometimes implant electrodes into the person's brain and then record from those electrodes for an extended period of a week or more. During this time, the person is sitting in the hospital with a bandaged head, waiting for the next seizure, and is often willing to participate in research to help alleviate the boredom.

One of the things that researchers can do is to stimulate the brain by running very small amounts of electrical current through the electrodes in the patient's brain. This is not enough to cause a seizure or damage the brain, but it is enough to change the patient's conscious experience in radical and sometimes bizarre ways. My colleagues Josef Parvizi and Kalanit Grill-Spector did an experiment in one such patient where they stimulated a part of the brain involved in face processing. The

video of the patient's experience, published along with the paper, shows how a tiny bit of electricity can alter our experience:

> PARVIZI: Just look at my face and tell me what happens when I do this, alright? 1, 2, 3. [No stimulation is given.]
> PATIENT: Nothing.
> PARVIZI: I'm going to do it one more time, look at my face. 1, 2, 3. [Delivers 4 milliamp stimulation to face-sensitive area.]
> PATIENT: You just turned into somebody else. Your face metamorphosed. Your nose got saggy, went to the left. You almost looked like somebody I'd seen before, but somebody different. That was a trip.[4]

It is also possible to stimulate the brain in humans without surgery, though the stimulation is much less specific, using a technique called transcranial magnetic stimulation, or TMS for short. TMS involves putting a powerful electromagnetic coil up against the skull, and pulsing electricity through it for a very brief period (less than one thousandth of a second). When the electromagnet is pulsed, it causes a rapid change in the magnetic field underneath the coil, including in the brain tissue just below. Remember that neurons are conductors of electrical current, like little biological wires. We know from physics that a changing magnetic field will induce an electrical current in a conductor, and this is what happens underneath the coil: electrical activity is induced in the neurons below. If the pulses are strong enough they can actually cause a seizure, but researchers use much weaker pulses, which can alter brain activity without the risk of a dangerous seizure. Depending on how the pulses are applied, they can either stimulate or deactivate the neurons underneath the coil.

I experienced TMS firsthand in the 1990s when I volunteered for an experiment being run by my fellow postdoctoral researcher John Desmond. John wanted to use TMS to examine whether disruption of a specific brain area in the parietal lobe would affect the ability to hold information in mind, which we call working memory. In the experiment, I was shown a set of letters and had to hold them in memory, so that after a few seconds I could say which letters had been shown. On some of

these trials he would zap my parietal lobe with TMS, hoping to scramble the activity there and disrupt my memory. I don't think the TMS had a very strong effect on my memory, but it did have another very striking effect: every time he zapped me, I experienced a strong metallic taste on half of my tongue. It turns out that the TMS pulse was probably also stimulating nerves in my face that are involved in the perception of taste, and that's why I had this experience. This highlights one of the challenges with using TMS to study the brain—its effects are fairly widespread, at best targeting an area roughly the volume of a golf ball, and often stimulating nerves or muscles outside the brain as well.

Studying the Mind

I have just made the argument that the brain and the mind are identical, which might lead you to think that studying the mind and studying the brain are the same thing, but that's not quite right. We call people who study the brain "neuroscientists" and those who study the mind "psychologists." In particular I am referring to experimental psychologists, who use experimental methods to try to understand how the mind works by testing hypotheses about how people will behave in certain situations. This is the field that I initially trained in; it was only after I came to Stanford in 1995 as a postdoctoral fellow that I began to use neuroimaging to study the brain.

A nice example of this kind of experimental psychology research comes from a study by the psychologists Henry "Roddy" Roediger and Jeff Karpicke of Washington University, who examined how we can best learn and remember new information.[5] In their studies, the experimental participants are presented with short paragraphs about various topics, such as sea otters or the sun. Participants were split into three groups: one group was told to read the paragraph four times (without getting tested at all), another was told to read it three times and was then tested once on the content, and the third was allowed to read it once, and was then tested on it several times. All of the groups were then asked how well they thought they had learned the

materials, and the results were clear that the members of the first group (who had read the passage four times) were much more confident in their knowledge. The three groups were also tested on their knowledge of the material in the paragraph, either five minutes or one week later. The results of the immediate test were in line with the subjects' own predictions: memory was better for the people who had read the passage multiple times. However, a week later the results were strikingly opposite: The people who had been so confident immediately after learning had forgotten the most, and the people who had only read the paragraph once now had the best memories. Based on this research, the researchers proposed a theory that says that bringing information back from memory is actually one of the most powerful ways to cement the information into memory for the long term.

What's essential to note is that even though the research by Roediger and Karpicke is studying the workings of the brain, the study didn't measure the brain directly and the paper itself never mentions the brain. One can study the workings of the brain by measuring behavior, without actually looking at the brain itself. However, most psychology researchers now believe that the best way to understand the mind is to study both behavior and the brain simultaneously. This is the fundamental idea of the field known as *cognitive neuroscience*, of which I consider myself a member.

Cognitive Neuroscience before Neuroimaging

The focus of this book is on neuroimaging, which today is the most important tool of cognitive neuroscience. However, the field itself existed well before the advent of neuroimaging. Lore has it that the name "cognitive neuroscience" was hatched by Michael Gazzaniga and George Miller while they shared a taxi in the late 1970s. Miller was a famous experimental psychologist, perhaps best known for his 1956 paper titled "The Magical Number Seven, plus or minus Two," which pointed out that humans are limited to processing a small amount of information at once (usually about seven things) across many

different domains. Gazzaniga, who most would consider to be the father of cognitive neuroscience, was famous for his studies of "split-brain" patients, which had shown how the two hemispheres of the brain can act independently. The field that these two researchers envisioned would combine psychology and neuroscience research approaches to provide a better understanding of how the brain gives rise to the mind.

Before neuroimaging, the only way to understand human brain function was to study people with brain damage and examine how specific damage leads to specific cognitive problems. This method first took hold in the nineteenth century, when European neurologists like Paul Broca and Carl Wernicke examined the postmortem brains of patients who had suffered from stroke, and noted that the location of the stroke corresponded with different types of language impairments. This is, in a sense, relying on Nature to do our experiments for us. However, Nature is an unreliable lab partner: Strokes can be large and messy, and results are often difficult to interpret for this reason. However, in rare cases the natural experiment can be much more specific. A fascinating example comes from a small set of individuals with a disease called Urbach–Wiethe syndrome, who have been studied by Ralph Adolphs from Caltech and his colleagues for a number of years. This is a disease that primarily affects the skin, but it also has a very specific effect on the brain, causing degeneration in a part of the brain called the amygdala, which has long been associated with emotion and fear. These patients have normal intelligence and mostly normal cognitive function, but they do show a very specific deficit: they largely do not experience fear. In one study, these researchers exposed an Urback–Wiethe patient (known by her initials, "S.M.") to stimuli that would make most of us shiver: live snakes and spiders, a haunted house, and clips from scary films like *The Blair Witch Project* and *The Shining*.[6] None of these things fazed her; in fact, the researchers reported that at the haunted house, "She reacted to the monsters by smiling, laughing, or trying to talk to them."[7] The only kind of fear that has been identified in these individuals is the fear of suffocation. This kind of research provides us with important clues about the brain systems that are involved in experiencing fear, and many

other psychological functions have also been studied in this way, using patients with different kinds of lesions or brain disorders. Lesion studies still play a critical role in cognitive neuroscience because they allow us to ask a specific question: Is a particular brain region necessary for a particular cognitive function? Neuroimaging can't answer this question—sometimes brain areas can be active when a person does a task, but lesions to that area don't actually impair his or her ability to perform the task.

The Unlikely Success of fMRI

Since its invention in the early 1990s (which is discussed in much more detail in the next chapter), fMRI has overtaken all other methods in cognitive neuroscience, including lesion studies and other neuroimaging methods. However, in retrospect it's amazing that fMRI works at all. Its success relies upon a set of chemical and biological dominoes that all had to fall into place for it to have any chance of working, almost as if nature conspired to help make it just a bit easier for us to understand how the brain works (though just the tiniest little bit).

The first biological fact that makes fMRI possible is that the firing of neurons is relatively localized across the brain. Take the portion of the brain that processes visual information, which neuroscientists very creatively call the "visual cortex." Within this part of the brain, different sections respond to information coming from different parts of the visual world. Another part of the brain, in the temporal lobe (the auditory cortex), responds to sounds, and yet another (the motor cortex) makes my fingers move as I type words on the page. Different parts of the brain seem to do different things (that is, there is some degree of modularity, as we discussed above), and as we will see it is this localization of function that will ultimately allow us to decode what a person is doing or thinking of simply by looking at brain activity—the concept of decoding that I introduced earlier. It's possible to imagine that evolution could have constructed the brain very differently, with every function involving every part of the brain equally. In fact, until the middle of the twentieth century some very famous neuroscientists (such as Karl Lashley)

believed that this was the case. However, the demonstration of clear effects of specific brain lesions on specific functions finally convinced the field that function is localized in the brain, at least to some degree.

Another aspect of the brain that makes fMRI possible is that brains are organized in a relatively similar way across individuals. Every human (and in fact nearly every mammal, except for monotremes such as the duck-billed platypus or spiny anteater) has a visual cortex that sits at the back of the brain, receives input from the eyes, and shows activity that is related to vision. Similarly, most mammals have a motor strip at the rear of the frontal lobe that controls the hands, whiskers, paws, or paddles. Again, we might imagine that evolution could have given us a random, haphazard organization of brain areas that varies from one individual to another, like spots on a calico cat. In this case, it would be very difficult to combine neuroimaging data across individuals, which we often need to do in order to gain statistical power through averaging. We also would not be able to compare the results with those from animal research, which can give us better insight into exactly what is happening in a particular area. Instead, research using nonhuman animals has provided important validation for results from fMRI research. The alignment across people is far from perfect, but it's good enough that we can warp together the brains of different people in order to analyze them as a group.

A third crucial biological fact is that the firing of neurons results in changes in blood flow that happen in a localized fashion as well. When neurons become active in a particular part of the brain, blood flow increases within the very close vicinity of those neurons (though we don't yet fully understand how this works). Without such tight localization, we would be able to see changes in blood flow but wouldn't be able to tie them closely to the neurons that caused them.

The final biological domino is the fact that this blood flow response to the area of active neurons is, in an important sense, an overreaction—at least with regard to oxygen. Blood brings with it a number of important things that neurons need, two of the most important being glucose and oxygen. What we know is that the brain seems to deliver about the right amount of glucose

to make up for the energy used by the neurons when they fire, but it sends *too much* oxygen relative to the small amount that is used by neurons. The details of exactly how this works are still the fodder for spirited academic arguments, but what we know for sure is that it is this overflow of oxygenated blood that lets us detect the activity of neurons using fMRI.

The chemical fact that makes fMRI possible was discovered by the Nobel Prize-winning chemist Linus Pauling in the 1930s. He was studying the magnetic properties of the hemoglobin molecule, which is the molecule that carries oxygen in the blood. What he discovered was that oxygenated hemoglobin (which is what makes fresh blood red) was not magnetic, but deoxygenated hemoglobin was "paramagnetic." A paramagnetic substance is not a magnet itself, but it takes on magnetic properties in the presence of a magnetic field. Think of a paper clip, which is not magnetic on its own, but when put next to a bar magnet will become magnetic. The invention of fMRI took advantage of the relationship between oxygen level and the magnetic characteristics of blood, by developing particular ways to use the MRI scanner to detect these differences.

What Can't Neuroimaging Tell Us?

While fMRI has shown itself to be incredibly powerful, it has also been used in ways that go beyond what it can actually tell us, which was illustrated well in an event from 2007. On November 11 of that year, an op-ed piece titled "This Is Your Brain on Politics" was published in the *New York Times*.[8] The authors, well-known neuroscientists and politicial scientists, reported results from a study in which they used fMRI to measure brain activity while so-called "swing voters" viewed video clips of candidates in the then-ongoing US presidential primaries. Based on these data, they drew a number of broad conclusions about the state of the electorate, which were based on the brain areas that were active while viewing the videos. One of the claims in the op-ed was that:

> Emotions about Hillary Clinton are mixed. Voters who rated Mrs. Clinton unfavorably on their questionnaire appeared not entirely comfortable with their assessment. When viewing

> images of her, these voters exhibited significant activity in the anterior cingulate cortex, an emotional center of the brain that is aroused when a person feels compelled to act in two different ways but must choose one. It looked as if they were battling unacknowledged impulses to like Mrs. Clinton. Subjects who rated her more favorably, in contrast, showed very little activity in this brain area when they viewed pictures of her.

Here was the verdict on Barack Obama:

> Mr. Obama was rated relatively high on the pre-scan questionnaire, yet both men and women exhibited less brain activity while viewing the pre-video set of still pictures of Mr. Obama than they did while looking at any of the other candidates. Among the male subjects, the video of Mr. Obama provoked increased activity in some regions of the brain associated with positive feeling, but in women it elicited little change.

As I read this piece, my blood began to boil. My research has focused on what kinds of things we can and cannot learn from neuroimaging data, and one of the clearest conclusions to come from this work is that activity in a particular region in the brain *cannot* tell us on its own whether a person is experiencing fear, reward, or any other psychological state. In fact, when people claim that activation in a particular brain area signals something like fear or reward, they are committing a basic logical fallacy, which is now referred to commonly as *reverse inference*. My ultimate fear was that the kind of fast-and-loose interpretation of fMRI data seen in the *New York Times* op-ed would lead readers to think erroneously that this kind of reasoning was acceptable, and would also lead other scientists to ridicule our field.

What's the problem with reverse inference? Take the example of a fever. If we see that our child has a fever, we can't really tell what particular disease he or she has, because there are so many different diseases that cause a fever (flu, pneumonia, and bacterial infections, just to name a few). On the other hand, if we see a round red rash with raised bumps, we can be fairly sure that it is caused by ringworm, because there are few other diseases that cause such a specific symptom. When we are interpreting

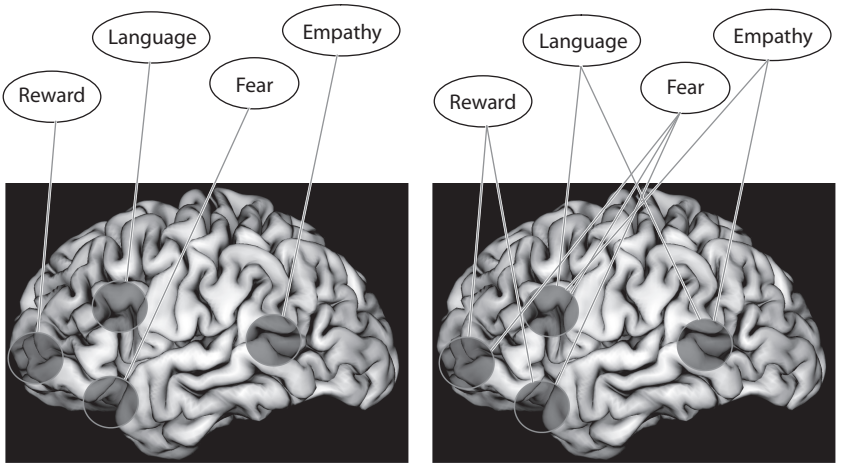

Figure 1.5. Can you infer cognitive function from areas of brain activation? If there was a one-to-one mapping between brain areas and cognitive functions, as shown in the *left panel*, then reverse inference based on activation in those areas would be possible—activation in the amygdala would imply fear, and activation in the ventromedial prefrontal cortex would imply reward. However, the brain is actually organized more like the *right panel*—any mental function involves a combination of many different brain regions, that are combined in different ways to support different mental functions.

brain activation, we need to ask the analogous question: How many different psychological processes could have caused the activation? If we knew, for example, that mental conflict was the only thing that causes the anterior cingulate cortex to be active, then we would be fairly safe in concluding from anterior cingulate activity that the person is experiencing conflict when viewing images of Hillary Clinton. On the other hand, if many different things can cause the region to be active, then we can't safely draw that conclusion. Figure 1.5 shows an example of each of these two different cases. Work that I published in 2006 showed that activity of individual brain regions was not very specific for different psychological functions (that is, it's more like a fever than a round rash), and thus that this kind of simple reverse inference is problematic.[9] The anterior cingulate cortex is a prime example of this. When we looked across many thousands of published neuroimaging studies in a later study, we found that this area was active in about one-quarter of all those studies,

which involved many different types of cognitive tasks.[10] This means that we cannot tell very much at all about what a person is doing from the fact that the anterior cingulate cortex was active.

Throughout this book, I will return to the fact that there is no simple one-to-one mapping between psychological states and activity in specific brain areas. As we will see, it is possible sometimes to decode the contents of a person's mind using fMRI, but it requires sophisticated statistical analyses along with careful interpretation.

A Road Map for the Book

The goal of this chapter was to give some background in the kinds of questions that cognitive neuroscientists ask, as a prelude to discussing how we use neuroimaging to ask them. In the rest of the book I will lay out the story of how neuroimaging came to be, what it can and can't tell us, and where it is going.

The first section of the book focuses on the development of neuroimaging as a tool for studying the mind and brain. In chapter 2, I will describe how researchers in the 1980s began to use a method called positron emission tomography (PET) to study how brain activity related to mental function, building on a century of prior ideas about the relation between brain activity and blood flow. Their discoveries led to the development of functional magnetic resonance imaging (fMRI), which is now the dominant technique for measuring brain activity in humans, relying on an amazing confluence of biology, chemistry, and physics. In chapter 3, I discuss how fMRI grew from being a new technique to the most powerful tool in human neuroscience. We will see how it was validated as a measure of brain activity, and how it was used to ask specific questions about how the brain is organized. In chapter 4 we will dig more deeply into how fMRI has been used to decode the contents of the mind and to attempt to achieve "mind reading." In chapter 5, I will discuss how fMRI has shown us how experiences change the brain, and how individual human brains change over time.

The second section of the book will focus on the ways in which neuroimaging has begun to influence the world outside of the

laboratory. In chapter 6, I discuss the ongoing attempts to use neuroimaging evidence in the courts, including the attempt to use fMRI for lie detection and why this is currently problematic. In chapter 7, I discuss the use of neuroscience tools to better understand how humans make choices and, ultimately, use them to find ways to sell us things more effectively, via the new field of consumer neuroscience. In chapter 8, I discuss how cognitive neuroscience has improved our understanding of mental illness, and discuss the ethical and social challenges of thinking of mental illnesses and addictions as "brain diseases."

Finally, in chapter 9 I discuss the future of fMRI and how its limitations may be addressed using other new methods.