There is a lesson here about the neocortex. It isn’t made of superfast components and the rules under which it operates are not that complex. However, it does have a hierarchical structure that contains billions of neurons and trillions of synapses. If we find it hard to imagine how such a logically simple but numerically vast memory system can create our consciousness, our languages, our cultures, our art, this book, and our science and technology, I suggest it is because our intuitive sense of the capacity of the cortex and the power of its hierarchical structure is inadequate. The neocortex does work. It isn’t magic. We can understand it. And like a computer, ultimately we can build intelligent machines that work on the same principles.

When I give talks about my brain theory, audiences are usually quick to grasp the significance of prediction as it relates to a host of human activities. They ask many related questions. Where does creativity come from? What is consciousness? What is imagination? How can we separate reality from false beliefs? Although these topics have not been in the forefront of my motivations for studying brains, they are of interest to nearly everyone. I don’t pretend to be an expert in these topics, but the memory-prediction framework of intelligence can provide some answers and useful insights. In this chapter, I address some of the most frequently asked questions.

**ARE ANIMALS INTELLIGENT?**
Is a rat intelligent? Is a cat intelligent? When did intelligence begin in evolutionary time? I love this question because I find the answer surprising.

Everything I have written so far about the neocortex and how it works depends on a very basic premise—that the
world has structure and is therefore predictable. There are patterns in the world: faces have eyes, eyes have pupils, fires are hot, gravity makes objects fall, doors open and shut, and so forth. The world is not random, nor is it homogeneous. Memory, prediction, and behavior would be meaningless if the world was without structure. All behavior, whether it is the behavior of a human, a snail, a single-cell organism, or a tree, is a means of exploiting the structure of the world for the benefit of reproduction.

Imagine a one-cell animal living in a pond. The cell has a flagellum that lets it swim. On the surface of the cell are molecules that detect the presence of nutrients. Since not all areas of the pond have the same concentration of nutrients, there is a gradual change in value, or gradient, of nutrients from one side of the cell to the other. As it swims across the pond, the cell can detect the shift. This is a simple form of structure in the world of the one-cell animal. The cell exploits its chemical awareness by swimming toward places with higher concentrations of nutrients. We could say that this simple organism is making a prediction. It is predicting that by swimming in a certain way it will find more nutrients. Is there memory involved in this prediction? Yes, there is. The memory is in the DNA of the organism. The one-cell animal did not learn, in its lifetime, how to exploit this gradient. Rather, the learning occurred over evolutionary time and is stored in the animal’s DNA. If the structure of the world changed suddenly, this particular one-cell animal could not learn to adapt. It could not alter its DNA or the resulting behavior. For this species, learning can occur only through evolutionary processes over many generations.

Is this one-cell organism intelligent? Using the everyday notion of human intelligence, the answer is no. But the animal does lie at the far edge of a continuum of species that use memory and prediction to reproduce more successfully, and by that more academic measure the answer is yes. The point is not to label some species as intelligent and others as not intelligent. Memory and prediction are used by all living things. There is just a continuum of methods and sophistication in how they do it.

Plants also use memory and prediction to exploit the structure of the world. A tree makes a prediction when it sends its roots down into the soil and its branches and leaves up toward the sky. The tree is predicting where it will find water and minerals based on the experience of its ancestors. Of course a tree doesn’t think; its behavior is automatic. But the species is exploiting the structure of the world in the same way as the one-cell organism. Every plant species has a distinct set of behaviors that exploit slightly different parts of the structure of the world.

Eventually, plants evolved communication systems, based mostly on the slow release of chemical signals. If an insect damages part of a tree, the tree sends chemicals through its vascular system to other parts of the tree, which triggers a defense system, such as making toxins. Through such a communication system, the tree can exhibit slightly more complex behavior. Neurons probably evolved as a way to communicate information more quickly than a plant’s vascular system. You could think of a neuron as just a cell with its own vascular appendages. At some point, instead of slowly moving chemicals along these appendages, the neuron started using electrochemical spikes, which travel much faster. In the beginning, fast synaptic transmission and simple nervous systems probably did not involve much if any learning. The name of the game was simply faster signaling.

But then, in the march of evolutionary time, something really interesting happened. Connections between neurons became modifiable. A neuron could send a signal or not send a signal, depending on what had happened recently. Behavior could now be modified within the life of an organism. The nervous system became plastic, and so did behavior. Because memories could be rapidly formed, the animal could learn the
structure of its world during its own lifetime. If the world suddenly changed—say, a new predator arrived on the scene—the animal didn’t have to stick with its genetically determined behavior, which might no longer be appropriate. Plastic nervous systems became a tremendous evolutionary advantage and led to a burst of new species from fish to snails to humans.

As we saw in chapter 3, all mammals have an old brain, on top of which sits the neocortex. The neocortex is just the most recent neural tissue to evolve. But with its hierarchical structure, invariant representations, and prediction by analogy, the cortex allows mammals to exploit much more of the structure of the world than an animal without a neocortex can. Our cortically endowed ancestors could envision how to make a net and catch fish. The fish are not able to learn that nets mean death or to figure out how to build tools to cut nets. All mammals, from rats to cats to humans, have a neocortex. They are all intelligent, but to differing degrees.

WHAT’S DIFFERENT ABOUT HUMAN INTELLIGENCE?
The memory-prediction framework offers two answers to this question. The first is pretty straightforward: our neocortex is larger than, say, a monkey’s or a dog’s. By enlarging the cortical sheet to the size of a large dinner napkin, our brains can learn a more complex model of the world and make more complex predictions. We see deeper analogies, more structure on structure, than other mammals. If we want to find a mate we don’t just look at simple attributes such as health, we interview their friends and parents, we observe how they drive and speak, and judge how honest they are. We look at these secondary and tertiary attributes to try to predict how our potential mate will behave in the future. Stock market traders look for structure in trading patterns. Mathematicians look for structure in numbers and equations. Astronomers look for structure in the motions of the planets and the stars. Our larger neocortex allows us to see our home as part of a town, which is part of a region, which is part of a planet, which is part of a large universe—structure within structure. No other mammal can ruminate to this depth. I am pretty certain my cat has no concept of a world outside our house.

The second difference between the intelligence of humans and other mammals is that we have language. Entire books have been written on the supposedly unique properties of language and how it developed. However, language fits nicely into the memory-prediction framework without any special language sauce or dedicated language machinery. Spoken and written words are just patterns in the world, as are melodies, cars, and houses. The syntax and semantics of language are not different from the hierarchical structure of other everyday objects. And in the same way that we associate the sound of a train with the visual memory image of a train, we associate spoken words with our memory of their physical and semantic counterparts. Through language one human can invoke memories and create new juxtapositions of mental objects in another human. Language is pure analogy, and through it we can cause other humans to experience and learn about things they may never actually see.

The development of language required a large neocortex capable of handling the nested structure of syntax and semantics. It also required a more fully developed motor cortex and musculature to enable us to make sophisticated, highly articulate sounds or gestures. With language, we can take patterns that we learn in a lifetime and transmit them to our children and our tribe. Language, whether it be written, spoken, or embodied in cultural traditions, became the means by which we pass on what we know about the world from generation to generation. Today, printed and electronic communications allow us to share our knowledge with millions of people around the world. Animals without language don’t transmit nearly as much information to
their offspring. A rat can learn many patterns in its lifetime, but it doesn’t pass on detailed new information—“Hey junior, here’s how my father taught me to avoid electric shocks.”

Thus, intelligence could be traced over three epochs, each using memory and prediction. The first would be when species used DNA as the medium for memory. Individuals could not learn and adapt within their lifetimes. They could only pass on the DNA-based memory of the world to their offspring through their genes.

The second epoch began when nature invented modifiable nervous systems that could quickly form memories. An individual could now learn about the structure of the world and adapt its behavior accordingly within its lifetime. But an individual still could not communicate this knowledge to its offspring other than by direct observation. The creation and expansion of the neocortex occurred within this second epoch, but did not define it.

The third and final epoch is unique to humans. It begins with the invention of language and the expansion of our large neocortex. We humans can learn a lot of the structure of the world within our lifetimes, and we can effectively communicate this to many other humans via language. You and I are participating in this process right now. I have spent a large part of my life searching for the structure in brains and how that structure leads to thought and intelligence. Through this book I am spreading what I have learned to you. Of course I couldn’t have done this if I hadn’t had access to the knowledge gathered by hundreds of scientists, who learned from others, and so on down through the ages. I was able to assimilate and add to what others have written about their own thinking and observation.

We have become the most adaptable creatures on the planet and the only ones with the ability to transfer our knowledge of the world broadly within our populace. The human population has undergone explosive growth because we can learn and exploit so much of the structure of the world and communicate it to other humans. We can thrive anywhere, be it a rain forest, a desert, the frozen tundra, or the concrete jungle. The combination of a large neocortex and language has led to the spiraling success of our species.

**WHAT IS CREATIVITY?**

I am frequently asked about creativity, I suspect because many people see creativity as something a machine couldn’t do, and therefore it is a challenge to the entire idea of building intelligent machines. What is creativity? We have already encountered the answer several times in this book. Creativity is not something that occurs in a particular region of the cortex. Nor is it like emotions or balance, which are rooted in particular structures and circuits outside of the cortex. Rather, creativity is an inherent property of every cortical region. It is a necessary component of prediction.

How can this be true? Isn’t creativity some extraordinary quality that requires high intelligence and giftedness? Not really. Creativity can be defined simply as making predictions by analogy, something that occurs everywhere in cortex and something you do continually while awake. Creativity occurs along a continuum. It ranges from simple everyday acts of perception occurring in sensory regions of the cortex (hearing a song in a new key) to difficult, rare acts of genius occurring at the highest levels in the cortex (composing a symphony in a brand-new way). At a fundamental level, everyday acts of perception are similar to the rare flights of brilliance. It is just that the everyday acts are so common we don’t notice them.

By now you have a basic understanding of how we create invariant memories, how we use invariant memories to make predictions, and how we make predictions of future events that are always somewhat different from anything we have experienced in the past. Recall also that our invariant memories are of sequences of events. We make predictions by combining the
invariant memory recall of what should happen next with the details pertaining to this moment in time (remember the parable of predicting when the train will arrive). Prediction is the application of invariant memory sequences to new situations. Therefore all cortical predictions are predictions by analogy. We predict the future by analogy to the past.

Imagine you are about to have dinner in an unfamiliar restaurant and you want to wash your hands. Even though you have never been in this building before, your brain predicts that there will be a restroom somewhere in the restaurant with a basin suitable for hand washing. How does it know this? Other restaurants you have been in have a restroom, and by analogy this restaurant will likely have one, too. Further, you know where and what to look for. You predict there will be a door or sign with some type of symbol associated with men or women. You predict it will be toward the back of the restaurant, either by the bar or down a hall, but generally not in plain view of the eating areas. Again, you have never been in this particular restaurant before, but by analogy to other eating establishments you are able to find what you need. You don’t look around randomly. You look for expected patterns that let you find the restroom quickly. This kind of behavior is a creative act; it is predicting the future by analogy to the past. We don’t normally think of this as being creative, but it very much is.

Recently I bought a vibraphone. We have a piano, but I had never played the vibraphone before. The day we brought it home, I took a sheet of music from the piano, placed it on the stand over the vibraphone, and started playing simple melodies. My ability to do this was not remarkable. But in a fundamental way, it was a creative act. Think about what was involved. I have an instrument that is very different from a piano. The vibraphone has gold metal bars; the piano has black and white keys. The gold bars are big and gradually change in size; the keys are small and of two different sizes. The gold bars are arranged in two different rows; the black and white keys are interleaved. On one instrument I use my fingers, and on the other I swing mallets. For this I’m standing up, and for that I’m sitting down. The particular muscles and motions needed to play the vibraphone are completely different from those needed to play the piano.

So how was I able to play a melody on an unfamiliar instrument? The answer is that my cortex sees an analogy between the keys on a piano and the bars on a vibraphone. Using this similarity allowed me to play a tune. It isn’t really any different from singing a song in a new key. In both cases, we know what to do by analogy to past learning. I realize that to you the similarity between these two instruments may appear obvious, but that is only because our brains automatically see analogies. Try to program a computer to find similarities between objects such as pianos and vibraphones and you will see how incredibly difficult this is. Prediction by analogy—creativity—is so pervasive we normally don’t notice it.

We do, however, believe we are being creative when our memory-prediction system operates at a higher level of abstraction, when it makes uncommon predictions, using uncommon analogies. For example, most people would agree that a mathematician who proves a difficult conjecture is being creative. But let’s take a close look at what’s involved with her mental efforts. Our mathematician stares hard at an equation and says, “How am I going to tackle this problem?” If the answer isn’t readily obvious she may rearrange the equation. By writing it down in a different fashion, she can look at the same problem from a different perspective. She stares some more. Suddenly she sees a part of the equation that looks familiar. She thinks, “Oh, I recognize this. There’s a structure to this equation that is similar to the structure of another equation I worked on several years ago.” She then makes a prediction by analogy. “Maybe I can solve this new equation using the same techniques I used successfully on the old equation.” She is able to solve the problem by analogy to a previously learned problem. It is a creative act.
My father had a mysterious blood disorder that his physicians could not diagnose. So how did they know what treatment to offer? One of the things they did was to look at months of data taken from analyses of my father’s blood to see if they could identify patterns. (My father printed a beautiful chart so the doctors could see the data clearly.) While his symptoms did not closely match those of known diseases, there were some similarities. The doctors ended up basing his treatment on a mixture of strategies that had worked for other blood disorders. The treatments used were guesses based on analogies to diseases the physicians had previously treated. Recognizing these patterns required extensive exposure to other uncommon diseases.

Shakespeare’s metaphors are the paragon of creativity. “Love is a smoke made with the fume of sighs.” “Adversity’s sweet milk, philosophy.” “There’s daggers in men’s smiles.” Such metaphors become obvious when you see them but they’re very hard to invent, which is one reason why Shakespeare is regarded as a literary genius. To create such metaphors he had to see a succession of clever analogies. When he writes “There’s daggers in men’s smiles,” he is not talking about daggers or smiles. Daggers are analogous to ill intent, and men’s smiles are analogous to deceit. Two clever analogies in only five words! At least that is how I interpret it. Poets have the gift of correlating seemingly unrelated words or concepts in manners that illuminate the world in new ways. They create unexpected analogies as a means of teaching higher-level structure.

In fact, highly creative works of art are appreciated because they violate our predictions. When you see a film that breaks the familiar mold of a character, story line, or cinematography (including special effects), you like it because it is not the same old same old. Paintings, music, poetry, novels—all creative artistic forms—strive to break convention and violate the expectations of an audience. There is a contradictory tension in what makes a work of art great. We want art to be familiar yet at the same time to be unique and unexpected. Too much familiarity is retread or kitsch; too much uniqueness is jarring and difficult to appreciate. The best works break some expected patterns while simultaneously teaching us new ones. Consider a great piece of classical music. The best music has an appeal at a simple level—good beat, simple melody and phrasing. Anyone can understand and appreciate it. However, it is also a little different and unexpected. But the more you listen to it, the more you see there is pattern in the unexpected parts, such as repeated unusual harmonies or key changes. The same is true with great literature or great movies. The more you read or see them, the more creative detail and complexity of structure you observe.

You’ve probably had the experience of looking at something when a twinge of recognition goes off in your head: “Hmmm, I’ve seen this pattern before, someplace else . . .” You may not have been trying to solve a problem, it’s just that an invariant representation in your brain was activated by a novel situation. You saw an analogy between two normally unrelated events. I might recognize that promoting a scientific idea is similar to selling a business idea or that bringing about political reform is like raising children. If I’m a poet, voilà! I have a new metaphor. If I’m a scientist or engineer, I have a new solution to a longstanding problem. Creativity is mixing and matching patterns of everything you’ve ever experienced or come to know in your lifetime. It’s saying “this is kinda like that.” The neural mechanism for doing this is everywhere in the cortex.

ARE SOME PEOPLE MORE CREATIVE THAN OTHERS?
A related question I often hear is, “If all brains are inherently creative, why are there differences in our creativity?” The memory-prediction framework points to two possible answers. One has to do with nature and the other with nurture.

On the nurture side, everyone has different life experiences.
Therefore everyone develops different models and memories of the world in his or her cortex, and will make different analogies and predictions. If I have been exposed to music, I will be able to sing songs in new keys and play simple melodies on new instruments. If I have never been exposed to music, I will not be able to make these predictive leaps. If I have studied physics, I will be able to explain the behavior of everyday objects via analogy to the laws of physics. If I grew up with dogs, I am apt to see analogies about dogs and will be better at predicting their behavior. Some people are more creative in social situations or in language, math, or diplomacy, all based on the environment they grew up in. Our predictions, and thus our talents, are built upon our experiences.

In chapter 6, I described how memories are pushed down the cortical hierarchy. The more you are exposed to certain patterns, the more the memory of these patterns are re-formed at lower levels. This allows you to learn the relationships among higher-order abstract objects at the top. It’s the essence of expertise. An expert is someone who through practice and repeated exposure can recognize patterns that are more subtle than can be recognized by a nonexpert, such as the shape of a fin on a late-fifties car or the size of a spot on a seagull’s beak. Experts can recognize patterns on top of patterns. Ultimately there is a physical limit to what we can learn constrained by the size of our cortex. But as humans, our cortex is large compared to other species and we have a tremendous flexibility in what we can learn. It all depends on what we are exposed to throughout our lives.

On the nature side, brains exhibit physical variation. Certainly some of the differences are genetically determined such as the size of regions (individuals can show as much as a three-fold difference in the gross area of V1) and hemispheric laterality (women tend to have thicker cables connecting the left and right sides of the brain than men do). Among humans, some brains probably have more cells or different kinds of connections. It’s unlikely that Albert Einstein’s creative genius was purely a function of the stimulating environment in the patent office where he worked as a young man. Recent analyses of his brain—which had been thought lost, but was found preserved in a jar a few years ago—reveal that his brain was measurably unusual. It had more support cells, called glia, per neuron than average. It showed an unusual pattern of grooves, or sulci, in the parietal lobes—a region thought to be important for mathematical abilities and spatial reasoning. It was also 15 percent wider than most other brains. We may never know why Einstein was as creative and smart as he was, but it is a safe bet that part of his talent derived from genetic factors.

Whatever the difference between brilliant and average brains, we are all creative. And through practice and study we can enhance our skills and talents.

**CAN YOU TRAIN YOURSELF TO BE MORE CREATIVE?**

Yes, most definitely. I have found there are ways to foster finding useful analogies when working on problems. First, you need to assume up front that there is an answer to what you are trying to solve. People give up too easily. You need confidence that a solution is waiting to be discovered and you must persist in thinking about the problem for an extended period of time.

Second, you need to let your mind wander. You need to give your brain the time and space to discover the solution. Finding a solution to a problem is literally finding a pattern in the world, or a stored pattern in your cortex that is analogous to the problem you are working on. If you are stuck on a problem, the memory-prediction model suggests that you should find different ways to look at it to increase the likelihood of seeing an analogy with a past experience. If you just sit there and stare at it over and over, you won’t get very far. Try taking the parts of your problem and rearranging them in different ways—literally and
figuratively. When I play Scrabble, I constantly shuffle the order of the tiles. It isn’t that I hope the letters will by chance spell a new word, but that different letter combinations will remind me of words or parts of words that might be part of a solution. If you are looking at a drawing of something that just doesn’t make sense, try drawing it upside down, changing colors, or changing perspectives. For example, when I was thinking about how different patterns in V1 could lead to invariant representations in IT, I was stuck. So I flipped the problem around and asked how a constant pattern in IT could lead to different predictions in V1. Inverting the problem was immediately helpful, ultimately leading to my belief that V1 should not be viewed as a single cortical region.

If you get stuck on a problem, go away for a little while. Do something else. Then start again, rephrasing the problem anew. If you do this enough times something will click sooner or later. It may take days or weeks, but eventually it will happen. The goal is to find an analogous situation somewhere in your past or present experience. To succeed you must ponder the problem often but also do other things so the cortex will have the opportunity to find an analogous memory.

Here is another example of how rearranging a problem led to a novel solution. In 1994, my colleagues and I were trying to figure out how to enter text on handheld computers. Everyone was focused on handwriting recognition software. They said, “Look, you write things on pieces of paper, you should be able to write the same way on a computer screen.” Unfortunately, this turns out to be really hard. It’s another one of those things that computers aren’t very good at, even though brains find it quite simple. The reason is that the brain uses memory and current context to predict what is written. Words and letters that are unrecognizable on their own are easily recognized in context. Pattern matching with computers is not sufficient to the task. I had designed several computers that used traditional handwriting recognition but it was never good enough.

I struggled with how to make the recognition software work better for several years and was stuck. One day I stepped back and decided to look at the problem from a different perspective. I looked for analogous problems. I said to myself, “How do we enter text into desktop computers? We type on a keyboard. How do we know how to type on a keyboard? Well, actually, it’s not easy. It’s a recent invention and it takes a long time to learn. Touch-typing on a typewriter-style keyboard is hard and not intuitive, it isn’t at all like writing—yet millions of people learn how. Why? Because it works.” My thinking continued by analogy, “Maybe I can come up with a text input system that is not necessarily intuitive, that you have to learn, but people will use it because it works.”

Literally, that’s the process I went through. I used the act of typing on a keyboard as an analogy to figure out how to enter text with a stylus on a display. I recognized that people were willing to learn a difficult task (typing) because it was a reliable and fast way to enter text into a machine. Therefore if we could create a new method of entering text with a stylus that was fast and reliable, people would use it even though it required learning. So I designed an alphabet that would reliably translate what you wrote into computer text; we called it Graffiti. With traditional handwriting recognition systems, when the computer misinterprets your writing you don’t know why. But the Graffiti system always produces the correct letter unless you make a mistake in writing. Our brains hate unpredictability, which is why people hate traditional handwriting recognition systems.

Many people thought Graffiti was a sensationally stupid idea. It went against everything they believed about how computers were supposed to work. The mantra in those days was that computers should adapt to the user, not the other way around. But I was confident that people would accept this new
way of entering text by analogy to the keyboard. Graffiti turned out to be a good solution and was widely adopted. To this day I still hear people claim that computers should adapt to users. This isn’t always true. Our brains prefer systems that are consistent and predictable, and we like learning new skills.

**CAN CREATIVITY LEAD ME ASTRAY?** CAN I FOOL MYSELF?

False analogy is always a danger. The history of science is rife with examples of beautiful analogies that turned out to be wrong. For example, the celebrated astronomer Johannes Kepler convinced himself that the orbits of the six known planets were defined by the Platonic solids. The Platonic solids are the only three-dimensional shapes that can be constructed entirely out of regular polygons. There are exactly five of them: tetrahedron (four equilateral triangles), hexahedron (six squares, aka a cube), octahedron (eight equilateral triangles), dodecahedron (twelve regular pentagons), and icosahedron (twenty equilateral triangles). They were discovered by the ancient Greeks, who were obsessed with the relationship of mathematics and the cosmos.

Like all Renaissance scholars, Kepler was heavily influenced by Greek thought. It seemed to him that it couldn’t possibly be coincidence that there were five Platonic solids and six plants. As he put it in his book *The Cosmic Mystery* (1596): “The dynamic world is represented by the flat-faced solids. Of these there are five: when viewed as boundaries, however, these five determine six distinct things: hence the six planets that revolve about the sun. This is also the reason why there are but six plants.” He saw a beautiful but entirely false analogy.

Kepler went on to account for the orbits of the planets in terms of nested Platonic solids that were all centered on the sun. He took the sphere defined by Mercury’s orbit as his baseline and circumscribed it with an octahedron. The tips of the octahedron defined a larger sphere, which gave the orbit of Venus. Around Venus’s orbit he circumscribed an icosahedron whose outer tips yielded the orbit of the Earth. The progression continued: a dodecahedron drawn around the Earth’s orbit gave the orbit of Mars, a tetrahedron around Mars’s orbit gave the orbit of Jupiter, and a cube around Jupiter’s orbit gave the orbit of Saturn. It was elegant and beautiful. Given the limited precision of astronomical data in his day, he was able to convince himself that this scheme worked! (Years later, Kepler realized he had been mistaken after he got hold of the high-precision astronomical data of his deceased colleague Tycho Brahe, which proved that planetary orbits are ellipses, not circles.)

Kepler’s excitement serves as a cautionary tale for scientists, and indeed for all thinkers. The brain is an organ that builds models and makes creative predictions, but its models and predictions can as easily be specious as valid. Our brains are always looking at patterns and making analogies. If correct correlations cannot be found, the brain is more than happy to accept false ones. Pseudoscience, bigotry, faith, and intolerance are often rooted in false analogy.

**WHAT IS CONSCIOUSNESS?**

This is one of those questions neuroscientists dread, unnecessarily so in my opinion. Some scientists, such as Christof Koch, are willing to tackle the issue of consciousness, but most consider it a question of philosophy bordering on pseudoscience. I think it deserves consideration if for no other reason than many people are curious about it. I cannot provide a completely satisfactory answer, but I think memory and prediction can address part of it. First, here’s a flavor of the conundrum as it comes up in conversation.

Not long ago I was at a science conference in a lovely spot on Long Island Sound. It was early evening when a dozen of us took our glasses of wine down to a pier to sit by the water and
chat before dinner and the evening session. After a while, the conversation turned to the topic of consciousness. As I said, neuroscientists normally don’t talk about this, but we were in a beautiful setting, some wine had been consumed, and the topic was brought up.

A British scientist was holding forth on her ideas about consciousness and said, “Of course, we’ll never understand consciousness.” I disagreed, “Consciousness is not a big problem. I think consciousness is simply what it feels like to have a cortex.” A silence fell on the group, then an argument quickly ensued as several scientists tried to educate me on my obvious error. “You must admit that the world seems so alive and beautiful. How can you deny your consciousness that perceives the world? You must admit you feel like something special.” To make a point, I said, “I don’t know what you are talking about. Given the way you are talking about consciousness, I have to conclude I am different from you. I don’t feel what you are feeling, so maybe I am not a conscious being. I must be a zombie.” Zombies are often invoked when philosophers talk about consciousness. A zombie is defined as physically identical to a human, but lacking consciousness. They are walking breathing meat machines, but with no one home.

The British scientist looked at me. “Of course you are conscious.”

“No, I don’t think so. I may look that way to you, but I’m not a conscious human being. Don’t worry about it, I’m okay with it.”

She said, “Well, don’t you perceive the wonder?” and swept her arm toward the glistening water as the sun began to sink and the sky turned iridescent salmon-pink.

“Yes, I see all this stuff. So?”

“How do you explain your subjective experience?”

I replied, “Yes, I know I’m here. I have memories of things like this evening. But I don’t feel anything special is going on, so if you feel something special maybe I’m just not conscious.” I was trying to pin her down about what she thought was so miraculous and unexplainable about consciousness. I was trying to get her to define consciousness.

We continued this line of argument—a yes you are, no I’m not kind of thing—until it was time to head up to dinner. I don’t think I changed anyone’s mind about the existence and meaning of consciousness. But I was trying to get them to realize that most people think consciousness is some kind of magical sauce that is added on top of the physical brain. You’ve got a brain, made of cells, and you pour consciousness, this magical sauce, on it, and that’s the human condition. In this view, consciousness is a mysterious entity separate from brains. That’s why zombies have brains but they don’t have consciousness. They have all the mechanical stuff, neurons and synapses, but they don’t have the special sauce. They can do everything a human can do. From the outside you can’t tell a zombie from a human.

The idea that consciousness is something extra stems from earlier beliefs in elan vital—a special force once thought to animate living things. People believed you needed this life force to explain the difference between rocks and plants or metals and maidens. Few people believe this anymore. Nowadays we know enough about the differences between inanimate and animate matter to understand that there isn’t a special sauce. We now know a great deal about DNA, protein folding, gene transcription, and metabolism. While we don’t yet know all the mechanisms of living systems, we know enough about biology to leave out magic. Similarly, no longer do people suggest it takes magic or spirits to make muscles move. We have folding proteins that pull long molecules past one another. You can read all about it.

Nevertheless, many people persist in believing that consciousness is different and can’t be explained in reductionist biological terms. Again, I am not a student of consciousness. I
haven’t read all the philosophers’ opinions. But I have some ideas about what I think people are confusing in this debate. I believe consciousness is simply what it feels like to have a neocortex. But we can do better than that. We can break consciousness into two major categories. One is similar to self-awareness—the everyday notion of being conscious. This is relatively easy to understand. The second is *qualia*—the idea that feelings associated with sensation are somehow independent of sensory input. Qualia is the harder part.

When most people say the word *conscious*, they are referring to the first category. “Were you conscious that you walked past me without saying hello?” “Were you conscious when you fell out of bed last night?” “You aren’t conscious when you sleep.” Some people say this form of consciousness is exactly the same as awareness. The two are close, but I don’t think awareness quite captures it correctly. I suggest this meaning of consciousness is synonymous with forming declarative memories. Declarative memories are memories that you can recall and talk about to someone else. You can express them verbally. If you ask me where I went last weekend, I can tell you. That is a declarative memory. If you ask me how to balance a bicycle, I can tell you to hold the handle bar and push the pedals, but I can’t explain exactly how to do it. How to balance a bicycle has mostly to do with neural activity in the old brain, so it is not a declarative memory.

I have a little thought experiment to show how our everyday notion of consciousness is the same as forming declarative memories. Recall that all memory is believed to reside in physical changes to synapses and the neurons they connect to. Therefore, if I had a method to reverse those physical changes, your memory would be erased. Now imagine I could flip a switch and return your brain to the exact physical state it was in at some point in the past. It could be an hour ago, twenty-four hours ago, whatever. I just flip the switch in my way-back machine and your synapses and neurons return to a previous state in time. By doing so, I erase all your memory of what occurred since that time.

Let’s assume you go through today and wake up tomorrow. But just as you’re waking up, I flip the switch and erase the last twenty-four hours. You would have absolutely zero memory of the previous day. From your brain’s perspective, yesterday never happened. I would tell you it’s Wednesday and you’d protest, “No, it’s Tuesday. I’m certain of it. The calendar has been altered. No way, this is Tuesday. Why are you pulling this trick on me?” But everyone whom you had met on Tuesday would say that you had been conscious throughout the day. They saw you, had lunch with you, and talked with you. Don’t you remember it? You’d say no, it didn’t happen. Finally, shown a video of you having lunch, you gradually become convinced that the day did happen, even though you have no memory of it. It’s as if you were a zombie for a day, not conscious. However, you were conscious at the time. Your belief that you were conscious disappeared only when your declarative memory was erased.

This thought experiment captures the equivalence between declarative memory and our everyday notion of being conscious. If during and at the end of a game of tennis I asked you if you are conscious, you would, of course, say yes. If I then erased your memory of the last two hours, you would claim to have been unconscious and not responsible for your actions during that time. In either case, you played the same game of tennis. The only difference is whether you have a memory of it at the time I ask you. Therefore, this meaning of consciousness is not absolute. It can be changed after the fact by memory erasure.

The more difficult question about consciousness concerns qualia. Qualia is often phrased in Zen-like queries, such as “Why is red red and green green? Does red look the same to me as it does to you? Why is red emotionally laden with certain feelings? It has a certain inextricable quality or feelingness to me. What feelingness does it cause in you?”
I find such descriptions difficult to relate to neurobiology, so I'd like to rephrase the question. For me, an equivalent question, but one I still find hard to explain, is, Why do different senses seem qualitatively different? Why does sight seem different from hearing and why does hearing seem different from touch? If the cortex is the same everywhere, if it works with the same processes, if it is just dealing with patterns, if no sound or light enters the brain, just patterns, then why does vision seem so different from hearing? I find it difficult to describe how sight differs from hearing, but it self-evidently is. I assume it is for you, too. Yet an axon representing sound and another representing light are, for all practical purposes, identical. “Lightness” and “soundness” are not carried down the axon of a sensory neuron.

People with a condition called synesthesia have brains that blur the distinction between the senses—certain sounds have a color, or certain textures have a color. This tells us that the qualitative aspect of a sense is not immutable. Through some sort of physical modification, a brain can impart a qualitative aspect of vision to an auditory input.

So what is the explanation for qualia? I can think of two possibilities, neither of which I find completely satisfactory. One is that although hearing, touch, and vision work under similar principles in the neocortex, they are handled differently below the cortex. Hearing relies on a set of audition-specific subcortical structures that process auditory patterns before they reach the cortex. Somatosensory patterns also travel through a set of subcortical areas that are unique to somatic senses. Perhaps qualia, like emotions, are not mediated purely by the neocortex. If they are somehow bound up with subcortical parts of the brain that have unique wiring, perhaps tied to emotion centers, this might explain why we perceive them differently, even if it doesn't help explain why there is any sort of qualia sensation in the first place.

The other possibility I can think of is that the structure of the inputs—differences in the patterns themselves—dictates how you experience qualitative aspects of the information. The nature of the spatial-temporal pattern on the auditory nerve is different from the nature of the spatial-temporal pattern on the optic nerve. The optic nerve has a million fibers and carries quite a bit of spatial information. The auditory nerve has only thirty thousand fibers and carries more temporal information. These differences may be related to what we call qualia.

We can be certain that however consciousness is defined, memory and prediction play crucial roles in creating it.

Related to consciousness are the notions of mind and soul.

As a child I used to wonder what it would have been like if “I” had been born in another child's body in another country, as if “I” was somehow independent of my body. These feelings of a mind independent of physicalness are common and a natural consequence of how the neocortex works. Your cortex creates a model of the world in its hierarchical memory. Thoughts are what occur when this model runs on its own; memory recall leads to predictions, which act like sensory inputs, which lead to new memory recall, and so on. Our most contemplative thoughts are not driven by or even connected to the real world; they are purely a creation of our model. We close our eyes and seek quiet so that our thinking will not be interrupted by sensory input. Of course our model was originally created by exposure to the real world through our senses, but when we plan and think about the world, we do so via the cortical model, not the world itself.

To the cortex, our bodies are just part of the external world. Remember, the brain is in a quiet and dark box. It knows about the world only via the patterns on the sensory nerve fibers. From the brain's perspective as a pattern device, it doesn't know about your body any differently than it knows about the rest of
the world. There isn’t a special distinction between where the body ends and the rest of the world begins. But the cortex has no ability to model the brain itself because there are no senses in the brain. Thus we can see why our thoughts appear independent of our bodies, why it feels like we have an independent mind or soul. The cortex builds a model of your body but it can’t build a model of the brain itself. Your thoughts, which are located in the brain, are physically separate from the body and the rest of the world. Mind is independent of body, but not of brain.

We can clearly see this differentiation through trauma and disease. If someone loses a limb, his brain’s model of the limb may nevertheless remain intact, resulting in a so-called phantom limb, which he can still feel attached to his body. On the flip side, if he suffers cortical trauma he may lose his model of the arm even though he retains the arm itself. In this case he may suffer what’s known as alien limb syndrome and have the uncomfortable, perhaps intolerable, feeling that the arm is not his own and is being controlled by someone else. Some even insist that the limb should be amputated! If our brain stays intact while the rest of our body becomes ill, we have the feeling of a healthy mind trapped in a dying body, although what we really have is a healthy brain trapped in a dying body. It is natural to imagine that our mind will continue after the death of our body, but when the brain dies so does the mind. The truth of this is evident if our brains fail before our bodies. People with Alzheimer’s disease or with serious brain damage lose their minds even if their bodies stay healthy.

WHAT IS IMAGINATION?
Conceptually, imagination is rather simple. Patterns flow into each cortical area either from your senses or from lower areas of the memory hierarchy. Each cortical area creates predictions, which are sent back down the hierarchy. To imagine something, you merely let your predictions turn around and become inputs. Without physically doing anything, you can follow the consequences of your predictions. “If this happens, then this will happen, then this will happen,” and so on. We do this when preparing for a business meeting, playing a game of chess, preparing for a sports event, or doing a thousand other things.

In chess you imagine moving your knight to a certain position and then visualize what the board will look like after the move. With this image in mind, you predict what your opponent will do and what the board will look like following that move. Then you predict what you will do, and so on. You walk through the imagined steps and their consequences. Ultimately you decide, based on this imagined sequence of events, whether the initial move was a good one or not. Certain athletes, such as downhill skiers, can improve their performance if they mentally rehearse the racecourse over and over in their head. By closing their eyes and imagining each and every turn, every obstacle, and even being on the winning stand, they increase their chances of success. Imagining is just another word for planning. This is where the predictive ability of our cortex pays off. It permits us to know what the consequences of our actions will be before we do them.

Imagining requires a neural mechanism for turning a prediction into an input. In chapter 6 I proposed that cells in layer 6 are where precise prediction occurs. Cells in this layer project down to lower levels of the hierarchy, but they also project back up to the input cells in layer 4. Thus a region’s outputs can become its own inputs. As I mentioned earlier, longtime cortical modeler Stephen Grossberg calls this circuit for imagination “folded feedback.” If you close your eyes and imagine a hippopotamus, the visual area of your cortex will become active, just as it would if you actually were looking at a hippo. You see what you imagine.
WHAT IS REALITY?
People ask, with an expression of worry and astonishment, “Do you mean to say that our brains create a model of the world? And that the model can be more important than actual reality?”

“Well, yes, to some extent that would be true,” I say.

“But doesn’t the world exist outside my head?”

Of course it does. People are real, trees are real, my cat is real, the social situations you find yourself in are real. But your understanding of the world and your responses to it are based on predictions coming from your internal model. At any moment in time, you can directly sense only a tiny part of your world. That tiny part dictates what memories will be invoked, but it isn’t sufficient on its own to build the whole of your current perception. For example, I am typing in my office right now and hear a knock on my front door. I know my mother has come to visit and I imagine she is downstairs, even though I haven’t actually seen or heard her. There was nothing in the sensory input that was specifically tied to my mother. It is my memory model of the world that predicts she is here by analogy to past experience. Most of what you perceive is not coming through your senses; it is generated by your internal memory model.

So the question “What is reality?” is largely a matter of how accurately our cortical model reflects the true nature of the world.

Many aspects of the world around us are so consistent that nearly every human has the same internal model of them. As a baby, you learned that the light falling on a round object produces a certain shadow, and that you can assess the shape of most objects by cues from the natural world. You learned that if you flung a cup off your highchair, gravity always pulled it to the floor. You learned textures, geometry, colors, and the rhythms of day and night. The simple physical properties of the world are learned consistently by all people.

But much of our world model is based on custom, culture, and what our parents teach us. These parts of our model are less consistent and might be totally different for different people. A child who is raised in a loving, caring home with parents who respond to his or her emotional needs will probably grow to adulthood predicting that the world is safe and loving. Children abused by one or both parents are more likely to see future events as dangerous and cruel, and believe that no one is to be trusted—no matter how well they are treated later. Much of psychology is based on the consequences of early life experience, attachment, and nurturance because that is when the brain first lays down its model of the world.

Your culture thoroughly shapes your world model. For example, studies show that Asians and Westerners perceive space and objects differently. Asians attend more to the space between objects, whereas Westerners mostly attend to objects—a difference that translates into separate aesthetics and ways of solving problems. Research has shown that some cultures, such as tribes in Afghanistan and some communities in the American South, are built on principles of honor and, as a result, are more likely to accept the naturalness of violence. Differing religious beliefs learned early in life can lead to completely different models of morality, how men and women are to be treated, and even the value of life itself. Clearly these differing models of the world can’t all be correct in some absolute, universal way, even though they may seem correct to an individual. Moral reasoning, both the good and the bad, is learned.

Your culture (and family experience) teaches you stereotypes, which are unfortunately an unavoidable part of life. Throughout this book, you could substitute the word stereotype for invariant memory (or invariant representation) without substantially altering the meaning. Prediction by analogy is pretty much the same as judgment by stereotype. Negative stereotyping has terrible social consequences. If my theory of
intelligence is right, we cannot rid people of their propensity to think in stereotypes, because stereotypes are how the cortex works. Stereotyping is an inherent feature of the brain.

The way to eliminate the harm caused by stereotypes is to teach our children to recognize false stereotypes, to be empathetic, and to be skeptical. We need to promote these critical-thinking skills in addition to instilling the best values we know. Skepticism, the heart of the scientific method, is the only way we know how to ferret out fact from fiction.

By now, I hope I have convinced you that mind is just a label of what the brain does. It isn’t a separate thing that manipulates or coexists with the cells in the brain. Neurons are just cells. There is no mystical force that makes individual nerve cells or collections of nerve cells behave in ways that differ from what they would normally do. Given this fact, we can now turn our attention to how we might implement the ability of brain cells to remember and predict—our cortical algorithm—in silicon.

It’s hard to predict the ultimate uses of a new technology. As we’ve seen throughout this book, brains make predictions by analogy to the past. So our natural inclination is to imagine that a new technology will be used to do the same kinds of things as a previous technology. We imagine using a new tool to do something familiar, only faster, more efficiently, or more cheaply.

Examples are abundant. People called the railroad the “iron horse” and the automobile the “horseless carriage.” For decades the telephone was viewed in the context of the telegraph, something that should be used only to communicate important news or emergencies; it wasn’t until the 1920s that people started using it casually. Photography was at first used as a new form of portraiture. And motion pictures were conceptualized as a variation on stage plays, which is why movie theaters had retracting curtains over the screens for much of the twentieth century.

Yet the ultimate uses of a new technology are often unexpected and more far-reaching than our imaginations can at