your friend’s eyes and nose are, and you know the structure of her face, then you can predict exactly where her lips should be. If you know her skin is being tinged orange by the light of sunset, then you know what color her hair should appear. Once again, your brain does this by combining a memory of the invariant structure of her face with the particulars of your immediate experience.

The train schedule example is just an analogy of what is going on in your cortex, but the melody and face examples are not. The combining of invariant representations and current input to make detailed predictions is exactly what is happening. It is a ubiquitous process that happens in every region of cortex. It is how you make specific predictions about the room you are sitting in right now. It is how you are able to predict not only the words others will say, but also in what tone of voice they will say them, the accent they will use, and where in the room you expect to hear the voice come from. It is how you know precisely when your foot will hit the floor, and what it will feel like when you climb a set of stairs. It is how you can sign your name with your foot, or catch a thrown ball.

The three properties of cortical memory discussed in this chapter (storing sequences, auto-associative recall, and invariant representations) are necessary ingredients to predict the future based on memories of the past. In the next chapter I propose that making predictions is the essence of intelligence.

---

A NEW FRAMEWORK OF INTELLIGENCE

One day in April 1986 I was contemplating what it means to “understand” something. For months I had been struggling with the fundamental question What do brains do if they aren’t generating behavior? What does a brain do when it is passively listening to speech? What is your brain doing right now while it is reading? Information goes into the brain but doesn’t come out. What happens to it? Your behaviors at the moment are probably basic—such as breathing and eye movements—yet, as you are aware, your brain is doing a lot more than that as you read and understand these words. Understanding must be the result of neural activity. But what? What are the neurons doing when they understand?

As I looked around my office that day, I saw familiar chairs, posters, windows, plants, pencils, and so on. There were hundreds of items and features all around me. My eyes saw them as I glanced around, yet just seeing them didn’t cause me to perform any action. No behavior was invoked or required, yet somehow I “understood” the room and its
nts. I was doing what Searle’s Chinese Room couldn’t do, didn’t have to pass anything back through a slot. I under-
stood, but had no action to prove it. What did it mean to “un-
derstand?”

It was while pondering this dilemma that I had an “aha”
 moment, one of those emotionally powerful moments when sud-
denly what was a tangle of confusion becomes clear and under-
standable. All I did was ask what would happen if a new object, one
never seen before, appeared in the room—say, a blue cof-
fey.

The answer seemed simple. I would notice the new object
belonging. It would catch my attention as being new. I
wouldn’t consciously ask myself if the coffee cup was new. It
would jump out as being new. Underlying that seem-
trivial answer is a powerful concept. To notice that some-
thing is different, some neurons in my brain that weren’t active
would have to become active. How would these neurons
know that the blue coffee cup was new and the hundreds of
objects in the room were not? The answer to this question
surprises me. Our brains use stored memories to constantly
make predictions about everything we see, feel, and hear. When
we walk around the room, my brain is using memories to form
predictions about what it expects to experience before I experi-
ex. The vast majority of predictions occur outside of aware-
ness. It’s as if different parts of my brain were saying, “Is the
chair in the middle of the desk? Yes. Is it black? Yes. Is the
book on the right-hand corner of the desk? Yes. Is the dictionary
here? I left it? Yes. Is the window rectangular and the walls ver-
tical? Yes. Is sunlight coming from the correct direction for the
time of day? Yes.” But when some visual pattern comes in that I
am not memorized in that context, a prediction is violated. And
attention is drawn to the error.

Of course, the brain doesn’t talk to itself while making pre-
cisions, and it doesn’t make predictions in a serial fashion. It
also doesn’t just make predictions about distinct objects like
coffee cups. Your brain constantly makes predictions about the
very fabric of the world we live in, and it does so in a parallel
fashion. It will just as readily detect an odd texture, a misshapen
nose, or an unusual motion. It isn’t immediately apparent how
pervasive these mostly unconscious predictions are, which is
perhaps why we missed their importance for so long. They
happen so automatically, so easily, we fail to fathom what is hap-
pening inside our skulls. I hope to impress on you the power of
this idea. Prediction is so pervasive that what we “perceive”—
that is, how the world appears to us—does not come solely from
our senses. What we perceive is a combination of what we sense
and of our brains’ memory-derived predictions.

Minutes later I conceived a thought experiment to help con-
vey what I understood at that moment. I call it the altered door
experiment. Here is how it goes.

When you come home each day, you usually take a few sec-
onds to go through your front door or whichever door you use.
You reach out, turn the knob, walk in, and shut it behind you.
It’s a firmly established habit, something you do all the time and
pay little attention to. Suppose while you are out, I sneak over to
your home and change something about your door. It could be
almost anything. I could move the knob over by an inch, change
a round knob into a thumb latch, or turn it from brass to chrome.
I could change the door’s weight, substituting solid oak for a hollow
door, or vice versa. I could make the hinges squeaky and stiff, or make them glide frictionless. I could
widen or narrow the door and its frame. I could change its
color, add a knocker where the peephole used to be, or add a
window. I can imagine a thousand changes that could be made
to your door, unbeknownst to you. When you come home that
day and attempt to open the door, you will quickly detect that
something is wrong. It might take you a few seconds’ reflection to realize exactly what is wrong, but you will notice the change very quickly. As your hand reaches for the moved knob, you will realize that it is not in the correct location. Or when you see the door’s new window, something will appear odd. Or if the door’s weight has been changed, you will push with the wrong amount of force and be surprised. The point is that you will notice any of a thousand changes in a very short period of time.

How do you do that? How do you notice these changes? The AI or computer engineer's approach to this problem would be to create a list of all the door’s properties and put them in a database, with fields for every attribute a door can have and specific entries for your particular door. When you approach the door, the computer would query the entire database, looking at width, color, size, knob position, weight, sound, and so on. While this may sound superficially similar to how I described my brain checking each of its myriad predictions as I glanced around my office, the difference is real and far-reaching. The AI strategy is implausible. First, it is impossible to specify in advance every attribute a door can have. The list is potentially endless. Second, we would need to have similar lists for every object we encounter every second of our lives. Third, nothing we know about brains and neurons suggests that this is how they work. And finally, neurons are just too slow to implement computer-style databases. It would take you twenty minutes instead of two seconds to notice the change as you go through the door.

There is only one way to interpret your reaction to the altered door: your brain makes low-level sensory predictions about what it expects to see, hear, and feel at every given moment, and it does so in parallel. All regions of your neocortex are simultaneously trying to predict what their next experience will be. Visual areas make predictions about edges, shapes, objects, locations, and motions. Auditory areas make predictions about tones, direction to source, and patterns of sound. Somatosensory areas make predictions about touch, texture, contour, and temperature.

“Prediction” means that the neurons involved in sensing your door become active in advance of them actually receiving sensory input. When the sensory input does arrive, it is compared with what was expected. As you approach the door, your cortex is forming a slew of predictions based on past experience. As you reach out, it predicts what you will feel on your fingers, when you will feel the door, and at what angle your joints will be when you actually touch the door. As you start to push the door open, your cortex predicts how much resistance the door will offer and how it will sound. When your predictions are all met, you’ll walk through the door without consciously knowing these predictions were verified. But if your expectations about the door are violated, the error will cause you to take notice. Correct predictions result in understanding. The door is normal. Incorrect predictions result in confusion and prompt you to pay attention. The door latch is not where it’s supposed to be. The door is too light. The door is off center. The texture of the knob is wrong. We are making continuous low-level predictions in parallel across all our senses.

But that’s not all. I am arguing a much stronger proposition. Prediction is not just one of the things your brain does. It is the primary function of the neocortex, and the foundation of intelligence. The cortex is an organ of prediction. If we want to understand what intelligence is, what creativity is, how your brain works, and how to build intelligent machines, we must understand the nature of these predictions and how the cortex makes them. Even behavior is best understood as a by-product of prediction.

I don’t know who was the first person to suggest that prediction is key to understanding intelligence. In science and industry no one invents anything completely new. Rather, people see
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.

CONSCIOUSNESS AND CREATIVITY

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
first grasp. The telephone has evolved into a wireless voice and data communications network permitting any two people on the planet to communicate with each other, no matter where they are, via voice, text, and images. The transistor was invented by Bell Labs in 1947. It was instantly clear to people that the device was a breakthrough, but the initial applications were just improvements on old applications: transistors replaced vacuum tubes. This led to smaller and more reliable radios and computers, which was important and exciting in its day, but the main differences were the size and reliability of the machines. The transistor's most revolutionary applications weren't discovered until later. A period of gradual innovation was necessary before anyone could conceive of the integrated circuit, the microprocessor, the digital signal processor, or the memory chip. The microprocessor, likewise, was first developed, in 1970, with desktop calculators in mind. Again, the first applications were just replacements of existing technologies. The electronic calculator was a replacement for the mechanical desktop calculator. Microprocessors were also clear candidates to replace the solenoids that were then used in certain kinds of industrial control, such as switching traffic lights. However, it was years before the true power of the microprocessor began to be manifest. No one at the time could foresee the modern personal computer, the cell phone, the Internet, the Global Positioning System, or any other piece of today's bread-and-butter information technology.

By the same token, we would be foolish to think we can predict the revolutionary applications of brainlike memory systems. I fully expect these intelligent machines to improve life in all sorts of ways. We can be sure of it. But predicting the future of technology more than a few years out is impossible. To appreciate this you need only read some of the absurd prognostications futurists have confidently made over the years. In the 1950s, it was predicted that by the year 2000 we'd all have atomic reactors in our basements and take our vacations on the moon. But as long as we keep these cautionary tales in mind, there's a lot to be gained by speculating about what intelligent machines will be like. At a minimum, there are certain broad and useful conclusions we can draw about the future.

The questions are intriguing ones. Can we build intelligent machines, and, if so, what will they look like? Will they be closer to the humanlike robots seen in popular fiction, the black or beige box of a personal computer, or something else? How will they be used? Is this a dangerous technology that can harm us or threaten our personal liberties? What are the obvious applications for intelligent machines, and is there any way we can know what the fantastic applications will be? What will the ultimate impact of intelligent machines be on our lives?

**CAN WE BUILD INTELLIGENT MACHINES?**

Yes, we can build intelligent machines, but they may not be what you expect. Although it may seem like the obvious thing to do, I don't believe we will build intelligent machines that act like humans, or even interact with us in humanlike ways.

One popular notion of intelligent machines comes to us from movies and books—they are the lovable, evil, or occasionally bumbling humanoid robots that converse with us about feelings, ideas, and events, and play a role in endless science-fiction plots. A century of science fiction has trained people to view robots and androids as an inevitable and desirable part of our future. Generations have grown up with images of Robbie the Robot from *Forbidden Planet*, R2D2 and C3PO from *Star Wars*, and Lieutenant Commander Data from *Star Trek*. Even HAL in the movie *2001: A Space Odyssey*, although not possessing a body, was very humanlike, designed to be as much a companion as a programmed copilot for the humans on their long space journey. Limited-application robots—things like smart cars, autonomous minisubmarines to explore the deep ocean, and self-guided vacuum cleaners or lawn mowers—are feasible and
may well grow more common someday. But androids and robots like Commander Data and C3PO are going to remain fictional for a very long time. There are a couple of reasons for this.

First, the human mind is created not only by the neocortex but also by the emotional systems of the old brain and by the complexity of the human body. To be human you need all of your biological machinery, not just a cortex. To converse like a human on all matters (to pass the Turing Test) would require an intelligent machine to have most of the experiences and emotions of a real human, and to live a humanlike life. Intelligent machines will have the equivalent of a cortex and a set of senses, but the rest is optional. It might be entertaining to watch an intelligent machine shuffle around in a humanlike body, but it will not have a mind that is remotely humanlike unless we imbue it with humanlike emotional systems and humanlike experiences. That would be extremely difficult and, it seems to me, quite pointless.

Second, given the cost and effort that would be necessary to build and maintain humanoid robots, it is difficult to see how they could be practical. A robot butler would be more expensive and less helpful than a human assistant. While the robot might be “intelligent,” it would not have the kind of rapport and easy understanding a human assistant would have by virtue of being a fellow human being.

Both the steam engine and the digital computer evoked robotic visions, which never came to fruition. Similarly, when we think of building intelligent machines, many people find it natural to imagine humanlike robots once again, but it is unlikely to happen. Robots are a concept born of the industrial revolution and refined by fiction. We should not look to them for inspiration in developing genuinely intelligent machines.

So what will intelligent machines look like if not walking talking robots? Evolution discovered that if it attached a hierarchical memory system to our senses, the memory would model the world and predict the future. Borrowing from nature, we should build intelligent machines along the same lines. Here, then, is the recipe for building intelligent machines. Start with a set of senses to extract patterns from the world. Our intelligent machine may have a set of senses that differ from a human’s, and may even “exist” in a world unlike our own (more on this later). So don’t assume that it has to have a set of eyeballs and a pair of ears. Next, attach to these senses a hierarchical memory system that works on the same principles as the cortex. We will then have to train the memory system much as we teach children. Over repetitive training sessions, our intelligent machine will build a model of its world as seen through its senses. There will be no need or opportunity for anyone to program in the rules of the world, databases, facts, or any of the high-level concepts that are the bane of artificial intelligence. The intelligent machine must learn via observation of its world, including input from an instructor when necessary. Once our intelligent machine has created a model of its world, it can then see analogies to past experiences, make predictions of future events, propose solutions to new problems, and make this knowledge available to us.

Physically, our intelligent machine might be built into planes or cars, or sit stoically on a rack in a computer room. Unlike humans, whose brains must accompany their bodies, the memory system of an intelligent machine might be located remotely from its sensors (and “body,” if it had one). For example, an intelligent security system might have sensors located throughout a factory or a town, but the hierarchical memory system attached to those sensors could be locked in a basement of one building. Therefore, the physical embodiment of an intelligent machine could take many forms.

There is no reason that an intelligent machine should look, act, or sense like a human. What makes it intelligent is that it can understand and interact with its world via a hierarchical
memory model and can think about its world in a way analogous to how you and I think about our world. As we will see, its thoughts and actions might be completely different from anything a human does, yet it will still be intelligent. Intelligence is measured by the predictive ability of a hierarchical memory, not by humanlike behavior.

Let's turn our attention to the largest technical challenge we will face when building intelligent machines, creating the memory. To build intelligent machines, we will need to construct large memory systems that are hierarchically organized and that work like the cortex. We will confront challenges with capacity and connectivity.

Capacity is the first issue. Let's say the cortex has 32 trillion synapses. If we represented each synapse using only two bits (giving us four possible values per synapse) and each byte has eight bits (so one byte could represent four synapses), then we would need roughly 8 trillion bytes of memory. A hard drive on a personal computer today has 100 billion bytes, so we would need about eighty of today's hard drives to have the same amount of memory as a human cortex. (Don't worry about the exact numbers because they are all rough guesses.) The point is, this amount of memory is definitely buildable in the lab. We aren't off by a factor of a thousand, but it is also not the kind of machine you could put in your pocket or build into your toaster. What is important is that the amount of memory required is not out of the question, whereas only ten years ago it would have been. Helping us is the fact that we don't have to re-create an entire human cortex. Much less may suffice for many applications.

Our intelligent machines will need lots of memory. We will probably start building them using hard drives or optical disks, but eventually we will want to build them out of silicon as well.

Silicon chips are small, low power, and rugged. And it is only a matter of time before silicon memory chips could be made with enough capacity to build intelligent machines. In fact, there is an advantage intelligent memory has over conventional computer memory. The economics of the semiconductor industry is based on the percentage of chips that have errors. For many chips even a single error will make the chip useless. The percentage of good chips is called the yield. It determines whether a particular chip design can be manufactured and sold at a profit. Because the chance of an error increases as the size of the chip does, most chips today are no bigger than a small postage stamp. The industry has boosted the amount of memory on a single chip not by making the chip larger but, mostly, by making the individual features on the chip smaller.

But intelligent memory chips will be inherently tolerant of faults. Remember, no single component of your brain holds any indispensable item of data. Your brain loses thousands of neurons each day, yet your mental capacity decays at only a slow pace throughout your adult life. Intelligent memory chips will work on the same principles as cortex, so even if a percentage of the memory elements come out defective, the chip will still be useful and commercially viable. Most likely, the inherent tolerance to errors of brainlike memory will allow designers to build chips that are significantly larger and denser than today's computer memory chips. The result is that we may be able to put a brain in silicon sooner than current trends might indicate.

The second problem we have to overcome is connectivity. Real brains have large amounts of subcortical white matter. As we noted earlier, the white matter is made up of the millions of axons streaming this way and that just beneath the thin cortical sheet, connecting the different regions of the cortical hierarchy with each other. An individual cell in the cortex may connect to five or ten thousand other cells. This kind of massively parallel wiring is difficult or impossible to implement using traditional