1

STANDARD EQUIPMENT

Why are there so many robots in fiction, but none in real life? I would pay a lot for a robot that could put away the dishes or run simple errands. But I will not have the opportunity in this century, and probably not in the next one either. There are, of course, robots that weld or spray-paint on assembly lines and that roll through laboratory hallways; my question is about the machines that walk, talk, see, and think, often better than their human masters. Since 1920, when Karel Čapek coined the word *robot* in his play *R.U.R.*, dramatists have freely conjured them up: Speedy, Cutie, and Dave in Isaac Asimov's *I, Robot*, Robbie in *Forbidden Planet*, the flailing canister in *Lost in Space*, the daleks in *Dr. Who*, Rosie the Maid in *The Jetsons*, Nomad in *Star Trek*, Hymie in *Get Smart*, the vacant butlers and bickering haberdashers in *Sleeper*, R2D2 and C3PO in *Star Wars*, the Terminator in *The Terminator*, Lieutenant Commander Data in *Star Trek: The Next Generation*, and the wisecracking film critics in *Mystery Science Theater 3000*.

This book is not about robots; it is about the human mind. I will try to explain what the mind is, where it came from, and how it lets us see, think, feel, interact, and pursue higher callings like art, religion, and philosophy. On the way I will try to throw light on distinctively **human** quirks. Why do memories fade? How does makeup change the look of a face? Where do ethnic stereotypes come from, and when are they irrational? Why do people lose their tempers? What makes children bratty? Why do fools fall in love? What makes us laugh? And why do people believe in ghosts and spirits?

But the gap between robots in imagination and in reality is my starting point, for it shows the first step we must take in knowing ourselves: appreciating the fantastically complex design behind feats of mental life we take for granted. The reason there are no humanlike robots is not that the very idea of a mechanical mind is misguided. It is that the engineering problems that we humans solve as we see and walk and plan and make it through the day are far more challenging than landing on the moon or sequencing the human genome. Nature, once again, has found ingenious solutions that human engineers cannot yet duplicate. When Hamlet says, "What a piece of work is a man! how noble in reason! how infinite in faculty! in form and moving how express and admirable!" we should direct our awe not at Shakespeare or Mozart or Einstein or Kareem Abdul-Jabbar but at a four-year old carrying out a request to put a toy on a shelf.

In a well-designed system, the components are black boxes that perform their functions as if by magic. That is no less true of the mind. The faculty with which we ponder the world has no ability to peer inside itself or our other faculties to see what makes them tick. That makes us the victims of an illusion: that our own psychology comes from some divine force or mysterious essence or almighty principle. In the Jewish legend of the Golem, a clay figure was animated when it was fed an inscription of the name of God. The archetype is echoed in many robot stories. The statue of Galatea was brought to life by Venus' answer to Pygmalion's prayers; Pinocchio was vivified by the Blue Fairy. Modern versions of the Golem archetype appear in some of the less fanciful stories of science. All of human psychology is said to be explained by a single, omnipotent cause: a large brain, culture, language, socialization, learning, complexity, self-organization, neural-network dynamics.

I want to convince you that our minds are not animated by some godly vapor or single wonder principle. The mind, like the Apollo spacecraft, is designed to solve many engineering problems, and thus is packed with high-tech systems each contrived to overcome its own obstacles. I begin by laying out these problems, which are both design specs for a robot and the subject matter of psychology. For I believe that the discovery by cognitive science and artificial intelligence of the technical challenges overcome by our mundane mental activity is one of the great revelations of science, an awakening of the imagination comparable to learning that the universe is made up of billions of galaxies or that a drop of pond water teems with microscopic life.

THE ROBOT CHALLENGE

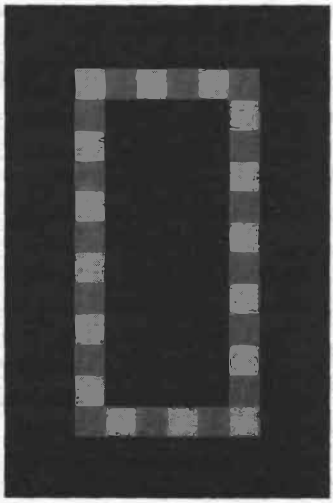
What does it take to build a robot? Let's put aside superhuman abilities like calculating planetary orbits and begin with the simple human ones: seeing, walking, grasping, thinking about objects and people, and planning how to act.

In movies we are often shown a scene from a robot's-eye view, with the help of cinematic conventions like fish-eye distortion or crosshairs. That is fine for us, the audience, who already have functioning eyes and brains. But it is no help to the robot's innards. The robot does not house an audience of little people—homunculi—gazing at the picture and telling the robot what they are seeing. If you could see the world through a robot's eyes, it would look not like a movie picture decorated with crosshairs but something like this:

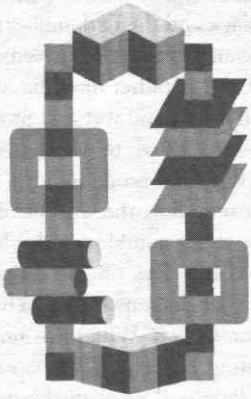
```
225 221 216 219 219 214 207 218 219 220 207 155 136 135
213 206 213 223 208 217 223 221 223 216 195 156 141 130
206 217 210 216 224 223 228 230 234 216 207 157 136 132
211 213 221 223 220 222 237 216 219 220 176 149 137 132
221 229 218 230 228 214 213 209 198 224 161 140 133 127
220 219 224 220 219 215 215 206 206 221 159 143 133 131
221 215 211 214 220 218 221 212 218 204 148 141 131 130
214 211 211 218 214 220 226 216 223 209 143 141 141 124
211 208 223 213 216 226 231 230 241 199 153 141 136 125
200 224 219 215 217 224 232 241 240 211 150 139 128 132
204 206 208 205 233 241 241 252 242 192 151 141 133 130
200 205 201 216 232 248 255 246 231 210 149 141 132 126
191 194 209 238 245 255 249 235 238 197 146 139 130 132
189 199 200 227 239 237 235 236 247 192 145 142 124 133
198 196 209 211 210 215 236 240 232 177 142 137 135 124
198 203 205 208 211 224 226 240 210 160 139 132 129 130
216 209 214 220 210 231 245 219 169 143 148 129 128 136
211 210 217 218 214 227 244 221 162 140 139 129 133 131
215 210 216 216 209 220 248 200 156 139 131 129 139 128
219 220 211 208 205 209 240 217 154 141 127 130 124 142
229 224 212 214 220 229 234 208 151 145 128 128 142 122
252 224 222 224 233 244 228 213 143 141 135 128 131 129
255 235 230 249 253 240 228 193 147 139 132 128 136 125
250 245 238 245 246 235 235 190 139 136 134 135 126 130
240 238 233 232 235 255 246 168 156 144 129 127 136 134
```

Each number represents the brightness of one of the millions of tiny patches making up the visual field. The smaller numbers come from darker patches, the larger numbers from brighter patches. The numbers shown in the array are the actual signals coming from an electronic camera trained on a person's hand, though they could just as well be the firing rates of some of the nerve fibers coming from the eye to the brain as a person looks at a hand. For a robot brain—or a human brain—to recognize objects and not bump into them, it must crunch these numbers and guess what kinds of objects in the world reflected the light that gave rise to them. The problem is humblingly difficult.

First, a visual system must locate where an object ends and the backdrop begins. But the world is not a coloring book, with black outlines around solid regions. The world as it is projected into our eyes is a mosaic of tiny shaded patches. Perhaps, one could guess, the visual brain looks for regions where a quilt of large numbers (a brighter region) abuts a quilt of small numbers (a darker region). You can discern such a boundary in the square of numbers; it runs diagonally from the top right to the bottom center. Most of the time, unfortunately, you would not have found the edge of an object, where it gives way to empty space. The juxtaposition of large and small numbers could have come from many distinct arrangements of matter. This drawing, devised by the psychologists Pawan Sinha and Edward Adelson, appears to show a ring of light gray and dark gray tiles.



In fact, it is a rectangular cutout in a black cover through which you are looking at part of a scene. In the next drawing the cover has been removed, and you can see that each pair of side-by-side gray squares comes from a different arrangement of objects.



Big numbers next to small numbers can come from an object standing in front of another object, dark paper lying on light paper, a surface painted two shades of gray, two objects touching side by side, gray cellophane on a white page, an inside or outside corner where two walls meet, or a shadow. Somehow the brain must solve the chicken-and-egg problem of identifying three-dimensional objects from the patches on the retina *and* determining what each patch is (shadow or paint, crease or overlay, clear or opaque) from knowledge of what object the patch is part of.

The difficulties have just begun. Once we have carved the visual world into objects, we need to know what they are made of, say, snow versus coal. At first glance the problem looks simple. If large numbers come from bright regions and small numbers come from dark regions, then large number equals white equals snow and small number equals black equals coal, right? Wrong. The amount of light hitting a spot on the retina depends not only on how pale or dark the object is but also on how bright or dim the light illuminating the object is. A photographer's light meter would show you that more light bounces off a lump of coal outdoors than off a snowball indoors. That is why people are so often disappointed by their snapshots and why photography is such a complicated craft. The camera does not lie; left to its own devices, it renders outdoor

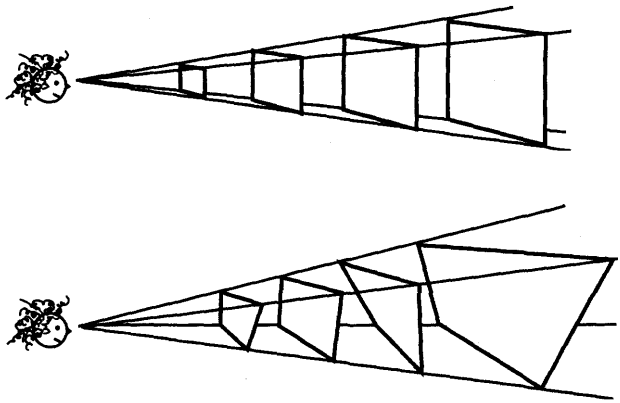
scenes as milk and indoor scenes as mud. Photographers, and sometimes microchips inside the camera, coax a realistic image out of the film with tricks like adjustable shutter timing, lens apertures, film speeds, flashes, and darkroom manipulations.

Our visual system does much better. Somehow it lets us see the bright outdoor coal as black and the dark indoor snowball as white. That is a happy outcome, because our conscious sensation of color and lightness matches the world as it is rather than the world as it presents itself to the eye. The snowball is soft and wet and prone to melt whether it is indoors or out, and we see it as white whether it is indoors or out. The coal is always hard and dirty and prone to burn, and we always see it as black. The harmony between how the world *looks* and how the world *is* must be an achievement of our neural wizardry, because black and white don't simply announce themselves on the retina. In case you are still skeptical, here is an everyday demonstration. When a television set is off, the screen is a pale greenish gray. When it is on, some of the phosphor dots give off light, painting in the bright areas of the picture. But the other dots do not suck light and paint in the dark areas; they just stay gray. The areas that you see as black are in fact just the pale shade of the picture tube when the set was off. The blackness is a figment, a product of the brain circuitry that ordinarily allows you to see coal as coal. Television engineers exploited that circuitry when they designed the screen.

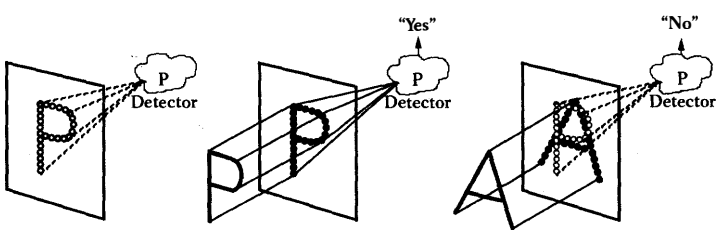
The next problem is seeing in depth. Our eyes squash the three-dimensional world into a pair of two-dimensional retinal images, and the third dimension must be reconstituted by the brain. But there are no telltale signs in the patches on the retina that reveal how far away a surface is. A stamp in your palm can project the same square on your retina as a chair across the room or a building miles away (first drawing, page 9). A cutting board viewed head-on can project the same trapezoid as various irregular shards held at a slant (second drawing, page 9).

You can feel the force of this fact of geometry, and of the neural mechanism that copes with it, by staring at a lightbulb for a few seconds or looking at a camera as the flash goes off, which temporarily bleaches a patch onto your retina. If you now look at the page in front of you, the afterimage adheres to it and appears to be an inch or two across. If you look up at the wall, the afterimage appears several feet long. If you look at the sky, it is the size of a cloud.

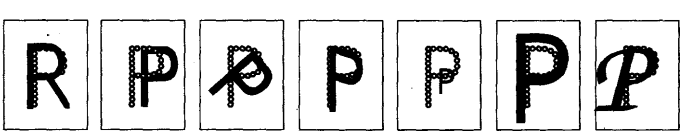
Finally, how might a vision module recognize the objects out there in the world, so that the robot can name them or recall what they do? The



obvious solution is to build a template or cutout for each object that duplicates its shape. When an object appears, its projection on the retina would fit its own template like a round peg in a round hole. The template would be labeled with the name of the shape—in this case, "the letter *P*"—and whenever a shape matches it, the template announces the name:



Alas, this simple device malfunctions in both possible ways. It sees *P*'s that aren't there; for example, it gives a false alarm to the *R* shown in the first square below. And it fails to see *P*'s that are there; for example, it misses the letter when it is shifted, tilted, slanted, too far, too near, or too fancy:



And these problems arise with a nice, crisp letter of the alphabet. Imagine trying to design a recognizer for a shirt, or a face! To be sure, after four decades of research in artificial intelligence, the technology of shape recognition has improved. You may own software that scans in a page, recognizes the printing, and converts it with reasonable accuracy to a file of bytes. But artificial shape recognizers are still no match for the ones in our heads. The artificial ones are designed for pristine, easy-to-recognize worlds and not the squishy, jumbled real world. The funny numbers at the bottom of checks were carefully drafted to have shapes that don't overlap and are printed with special equipment that positions them exactly so that they can be recognized by templates. When the first face recognizers are installed in buildings to replace doormen, they will not even try to interpret the chiaroscuro of your face but will scan in the hard-edged, rigid contours of your iris or your retinal blood vessels. Our brains, in contrast, keep a record of the shape of every face we know (and every letter, animal, tool, and so on), and the record is somehow matched with a retinal image even when the image is distorted in all the ways we have been examining. In Chapter 4 we will explore how the brain accomplishes this magnificent feat.

Let's take a look at another everyday miracle: getting a body from place to place. When we want a machine to move, we put it on wheels. The invention of the wheel is often held up as the proudest accomplishment of civilization. Many textbooks point out that no animal has evolved wheels and cite the fact as an example of how evolution is often incapable of finding the optimal solution to an engineering problem. But it is not a good example at all. Even if nature *could* have evolved a moose on wheels, it surely would have opted not to. Wheels are good only in a world with roads and rails. They bog down in any terrain that is soft, slippery, steep, or uneven. Legs are better. Wheels have to roll along an unbroken supporting ridge, but legs can be placed on a series of separate footholds, an extreme example being a ladder. Legs can also be placed to minimize lurching and to step over obstacles. Even today, when it seems as if the world has become a parking lot, only about half of the earth's land is accessible to vehicles with wheels or tracks, but most of the earth's land is accessible to vehicles with feet: animals, the vehicles designed by natural selection.

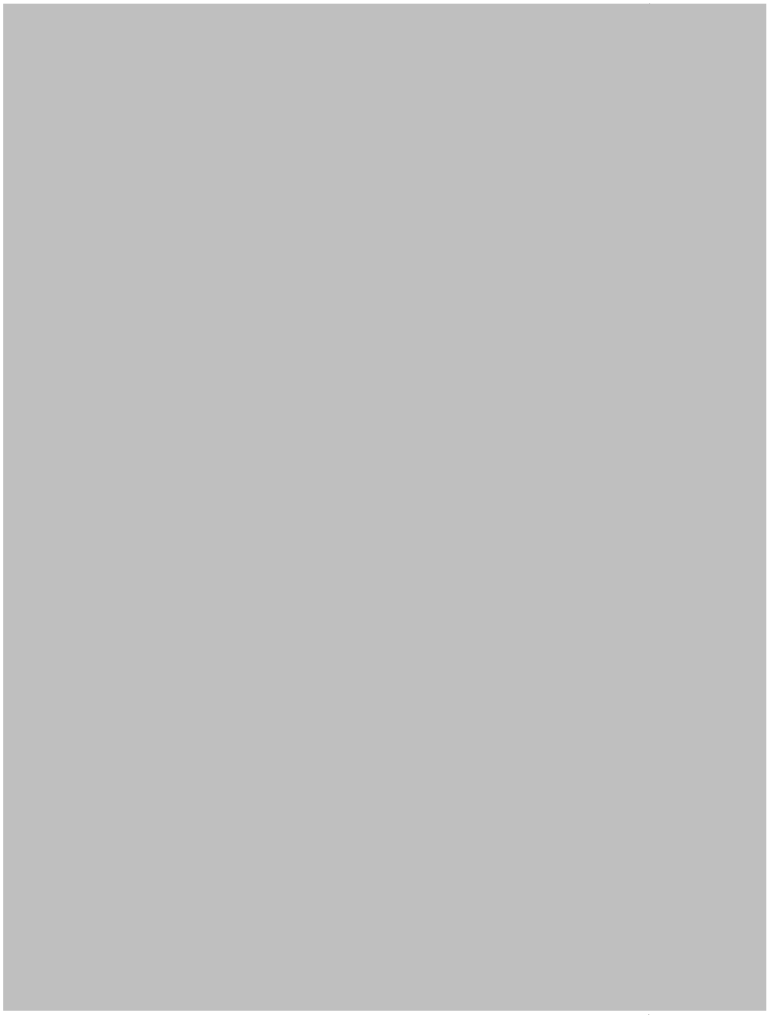
But legs come with a high price: the software to control them. A wheel, merely by turning, changes its point of support gradually and can bear weight the whole time. A leg has to change its point of support all at once, and the weight has to be unloaded to do so. The motors controlling a leg have to alternate between keeping the foot on the ground while it bears and propels the load and taking the load off to make the leg free to move. All the while they have to keep the center of gravity of the body within the polygon defined by the feet so the body doesn't topple over. The controllers also must minimize the wasteful up-and-down motion that is the bane of horseback riders. In walking windup toys, these problems are crudely solved by a mechanical linkage that converts a rotating shaft into a stepping motion. But the toys cannot adjust to the terrain by finding the best footholds.

Even if we solved these problems, we would have figured out only how to control a walking insect. With six legs, an insect can always keep one tripod on the ground while it lifts the other tripod. At any instant, it is stable. Even four-legged beasts, when they aren't moving too quickly, can keep a tripod on the ground at all times. But as one engineer has put it, "the upright two-footed locomotion of the human being seems almost a recipe for disaster in itself, and demands a remarkable control to make it practicable." When we walk, we repeatedly tip over and break our fall in the nick of time. When we run, we take off in bursts of flight. These aerobatics allow us to plant our feet on widely or erratically spaced footholds that would not prop us up at rest, and to squeeze along narrow paths and jump over obstacles. But no one has yet figured out how we do it.

Controlling an arm presents a new challenge. Grab the shade of an architect's lamp and move it along a straight diagonal path from near you, low on the left, to far from you, high on the right. Look at the rods and hinges as the lamp moves. Though the shade proceeds along a straight line, each rod swings through a complicated arc, swooping rapidly at times, remaining almost stationary at other times, sometimes reversing from a bending to a straightening motion. Now imagine having to do it in reverse: without looking at the shade, you must choreograph the sequence of twists around each joint that would send the shade along a straight path. The trigonometry is frightfully complicated. But your arm is an architect's lamp, and your brain effortlessly solves the equations every time you point. And if you have ever held an architect's lamp by its clamp, you will appreciate that the problem is even harder than what I have described. The lamp flails under its weight as if it had a mind of its

own; so would your arm if your brain did not compensate for its weight, solving a near-intractable physics problem.

A still more remarkable feat is controlling the hand. Nearly two thousand years ago, the Greek physician Galen pointed out the exquisite natural engineering behind the human hand. It is a single tool that manipulates objects of an astonishing range of sizes, shapes, and weights, from a log to a millet seed. "Man handles them all," Galen noted, "as well as if his hands had been made for the sake of each one of them alone." The hand can be configured into a hook grip (to lift a pail), a scissors grip (to hold a cigarette), a five-jaw chuck (to lift a coaster), a three-jaw chuck (to hold a pencil), a two-jaw pad-to-pad chuck (to thread a needle), a two-jaw pad-to-side chuck (to turn a key), a squeeze grip (to hold a hammer), a disc grip (to open a jar), and a spherical grip (to hold a ball). Each grip needs a precise combination of muscle tensions that mold the hand into the right shape and keep it there as the load tries to bend it back. Think of lifting a milk carton. Too loose a grasp, and you drop it; too tight, and you crush it; and with some gentle rocking, you can even use the tugging on your fingertips as a gauge of how much milk is inside! And I won't even begin to talk about the tongue, a boneless water balloon controlled only by squeezing, which can loosen food from a back tooth or perform the ballet that articulates words like *thrilling* and *sixths*.



Robot design is a kind of consciousness-raising. We tend to be blasé about our mental lives. We open our eyes, and familiar articles present themselves; we will our limbs to move, and objects and bodies float into place; we awaken from a dream, and return to a comfortingly predictable

world; Cupid draws back his bow, and lets his arrow go. But think of what it takes for a hunk of matter to accomplish these improbable outcomes, and you begin to see through the illusion. Sight and action and common sense and violence and morality and love are no accident, no inextricable ingredients of an intelligent essence, no inevitability of information processing. Each is a tour de force, wrought by a high level of targeted design. Hidden behind the panels of consciousness must lie fantastically complex machinery—optical analyzers, motion guidance systems, simulations of the world, databases on people and things, goal-schedulers, conflict-resolvers, and many others. Any explanation of how the mind works that alludes hopefully to some single master force or mind-bestowing elixir like "culture," "learning," or "self-organization" begins to sound hollow, just not up to the demands of the pitiless universe we negotiate so successfully.

The robot challenge hints at a mind loaded with original equipment, but it still may strike you as an argument from the armchair. Do we actually find signs of this intricacy when we look directly at the machinery of the mind and at the blueprints for assembling it? I believe we do, and what we see is as mind-expanding as the robot challenge itself.

When the visual areas of the brain are damaged, for example, the visual world is not simply blurred or riddled with holes. Selected aspects of visual experience are removed while others are left intact. Some patients see a complete world but pay attention only to half of it. They eat food from the right side of the plate, shave only the right cheek, and draw a clock with twelve digits squished into the right half. Other patients lose their sensation of color, but they do not see the world as an arty black-and-white movie. Surfaces look grimy and rat-colored to them, killing their appetite and their libido. Still others can see objects change their positions but cannot see them move—a syndrome that a philosopher once tried to convince me was logically impossible! The stream from a teapot does not flow but looks like an icicle; the cup does not gradually fill with tea but is empty and then suddenly full.

Other patients cannot recognize the objects they see: their world is like handwriting they cannot decipher. They copy a bird faithfully but identify it as a tree stump. A cigarette lighter is a mystery until it is lit. When they try to weed the garden, they pull out the roses. Some patients can recognize inanimate objects but cannot recognize faces. The patient deduces that the visage in the mirror must be his, but does not viscerally recognize himself. He identifies John F. Kennedy as Martin Luther King,

and asks his wife to wear a ribbon at a party so he can find her when it is time to leave. Stranger still is the patient who recognizes the face but not the person: he sees his wife as an amazingly convincing impostor.

These syndromes are caused by an injury, usually a stroke, to one or more of the thirty brain areas that compose the primate visual system. Some areas specialize in color and form, others in where an object is, others in what an object is, still others in how it moves. A seeing robot cannot be built with just the fish-eye viewfinder of the movies, and it is no surprise to discover that humans were not built that way either. When we gaze at the world, we do not fathom the many layers of apparatus that underlie our unified visual experience, until neurological disease dissects them for us.

Another expansion of our vista comes from the startling similarities between identical twins, who share the genetic recipes that build the mind. Their minds are astonishingly alike, and not just in gross measures like IQ and personality traits like neuroticism and introversion. They are alike in talents such as spelling and mathematics, in opinions on questions such as apartheid, the death penalty, and working mothers, and in their career choices, hobbies, vices, religious commitments, and tastes in dating. Identical twins are far more alike than fraternal twins, who share only half their genetic recipes, and most strikingly, they are almost as alike when they are reared apart as when they are reared together. Identical twins separated at birth share traits like entering the water backwards and only up to their knees, sitting out elections because they feel insufficiently informed, obsessively counting everything in sight, becoming captain of the volunteer fire department, and leaving little love notes around the house for their wives.

People find these discoveries arresting, even incredible. The discoveries cast doubt on the autonomous "I" that we all feel hovering above our bodies, making choices as we proceed through life and affected only by our past and present environments. Surely the mind does not come equipped with so many small parts that it could predestine us to flush the toilet before and after using it or to sneeze playfully in crowded elevators, to take two other traits shared by identical twins reared apart. But apparently it does. The far-reaching effects of the genes have been documented in scores of studies and show up no matter how one tests for them: by comparing twins reared apart and reared together, by comparing identical and fraternal twins, or by comparing adopted and biological children. And despite what critics sometimes claim, the effects are not

products of coincidence, fraud, or subtle similarities in the family environments (such as adoption agencies striving to place identical twins in homes that both encourage walking into the ocean backwards). The findings, of course, can be misinterpreted in many ways, such as by imagining a gene for leaving little love notes around the house or by concluding that people are unaffected by their experiences. And because this research can measure only the ways in which people *differ*, it says little about the design of the mind that all normal people share. But by showing how many ways the mind can vary in its innate structure, the discoveries open our eyes to how much structure the mind must have.

REVERSE-ENGINEERING THE PSYCHE

The complex structure of the mind is the subject of this book. Its key idea can be captured in a sentence: The mind is a system of organs of computation, designed by natural selection to solve the kinds of problems our ancestors faced in their foraging way of life, in particular, understanding and outmaneuvering objects, animals, plants, and other people. The summary can be unpacked into several claims. The mind is what the brain does; specifically, the brain processes information, and thinking is a kind of computation. The mind is organized into modules or mental organs, each with a specialized design that makes it an expert in one arena of interaction with the world. The modules' basic logic is specified by our genetic program. Their operation was shaped by natural selection to solve the problems of the hunting and gathering life led by our ancestors in most of our evolutionary history. The various problems for our ancestors were subtasks of one big problem for their genes, maximizing the number of copies that made it into the next generation.

On this view, psychology is engineering in reverse. In forward-engineering, one designs a machine to do something; in reverse-engineering, one figures out what a machine was designed to do. Reverse-engineering is what the boffins at Sony do when a new product is announced by Panasonic, or vice versa. They buy one, bring it back to the lab, take a screwdriver to it, and try to figure out what all the parts are for and how they combine to make the device work. We all engage in reverse-engineering when we face an interesting new gadget. In rummaging through

an antique store, we may find a contraption that is inscrutable until we figure out what it was designed to do. When we realize that it is an olive-pitter, we suddenly understand that the metal ring is designed to hold the olive, and the lever lowers an X-shaped blade through one end, pushing the pit out through the other end. The shapes and arrangements of the springs, hinges, blades, levers, and rings all make sense in a satisfying rush of insight. We even understand why canned olives have an X-shaped incision at one end.

In the seventeenth century William Harvey discovered that veins had valves and deduced that the valves must be there to make the blood circulate. Since then we have understood the body as a wonderfully complex machine, an assembly of struts, ties, springs, pulleys, levers, joints, hinges, sockets, tanks, pipes, valves, sheaths, pumps, exchangers, and filters. Even today we can be delighted to learn what mysterious parts are for. Why do we have our wrinkled, asymmetrical ears? Because they filter sound waves coming from different directions in different ways. The nuances of the sound shadow tell the brain whether the source of the sound is above or below, in front of or behind us. The strategy of reverse-engineering the body has continued in the last half of this century as we have explored the nanotechnology of the cell and of the molecules of life. The stuff of life turned out to be not a quivering, glowing, wondrous gel but a contraption of tiny jigs, springs, hinges, rods, sheets, magnets, zippers, and trapdoors, assembled by a data tape whose information is copied, downloaded, and scanned.

The rationale for reverse-engineering living things comes, of course, from Charles Darwin. He showed how "organs of extreme perfection and complication, which justly excite our admiration" arise not from God's foresight but from the evolution of replicators over immense spans of time. As replicators replicate, random copying errors sometimes crop up, and those that happen to enhance the survival and reproduction rate of the replicator tend to accumulate over the generations. Plants and animals are replicators, and their complicated machinery thus appears to have been engineered to allow them to survive and reproduce.

Darwin insisted that his theory explained not just the complexity of an animal's body but the complexity of its mind. "Psychology will be based on a new foundation," he famously predicted at the end of *The Origin of Species*. But Darwin's prophecy has not yet been fulfilled. More than a century after he wrote those words, the study of the mind is still mostly Darwin-free, often defiantly so. Evolution is said to be irrelevant,

sinful, or fit only for speculation over a beer at the end of the day. The allergy to evolution in the social and cognitive sciences has been, I think, a barrier to understanding. The mind is an exquisitely organized system that accomplishes remarkable feats no engineer can duplicate. How could the forces that shaped that system, and the purposes for which it was designed, be irrelevant to understanding it? Evolutionary thinking is indispensable, not in the form that many people think of—dreaming up missing links or narrating stories about the stages of Man—but in the form of careful reverse-engineering. Without reverse-engineering we are like the singer in Tom Paxton's "The Marvelous Toy," reminiscing about a childhood present: "It went ZIP! when it moved, and POP! when it stopped, and WHIRRR! when it stood still; I never knew just what it was, and I guess I never will."

Only in the past few years has Darwin's challenge been taken up, by a new approach christened "evolutionary psychology" by the anthropologist John Tooby and the psychologist Leda Cosmides. Evolutionary psychology brings together two scientific revolutions. One is the cognitive revolution of the 1950s and 1960s, which explains the mechanics of thought and emotion in terms of information and computation. The other is the revolution in evolutionary biology of the 1960s and 1970s, which explains the complex adaptive design of living things in terms of selection among replicators. The two ideas make a powerful combination. Cognitive science helps us to understand how a mind is possible and what kind of mind we have. Evolutionary biology helps us to understand *why* we have the kind of mind we have.

The evolutionary psychology of this book is, in one sense, a straightforward extension of biology, focusing on one organ, the mind, of one species, *Homo sapiens*. But in another sense it is a radical thesis that discards the way issues about the mind have been framed for almost a century. The premises of this book are probably not what you think they are. Thinking is computation, I claim, but that does not mean that the computer is a good metaphor for the mind. The mind is a set of modules, but the modules are not encapsulated boxes or circumscribed swatches on the surface of the brain. The organization of our mental modules comes from our genetic program, but that does not mean that there is a gene for every trait or that learning is less important than we used to think. The mind is an adaptation designed by natural selection, but that does not mean that everything we think, feel, and do is biologically adaptive. We evolved from apes, but that does not mean we have the same minds as

apes. And the ultimate goal of natural selection is to propagate genes, but that does not mean that the ultimate goal of people is to propagate genes. Let me show you why not.

This book is about the brain, but I will not say much about neurons, hormones, and neurotransmitters. That is because the mind is not the brain but what the brain does, and not even everything it does, such as metabolizing fat and giving off heat. The 1990s have been named the Decade of the Brain, but there will never be a Decade of the Pancreas. The brain's special status comes from a special thing the brain does, which makes us see, think, feel, choose, and act. That special thing is information processing, or computation.

Information and computation reside in patterns of data and in relations of logic that are independent of the physical medium that carries them. When you telephone your mother in another city, the message stays the same as it goes from your lips to her ears even as it physically changes its form, from vibrating air, to electricity in a wire, to charges in silicon, to flickering light in a fiber optic cable, to electromagnetic waves, and then back again in reverse order. In a similar sense, the message stays the same when she repeats it to your father at the other end of the couch after it has changed its form inside her head into a cascade of neurons firing and chemicals diffusing across synapses. Likewise, a given program can run on computers made of vacuum tubes, electromagnetic switches, transistors, integrated circuits, or well-trained pigeons, and it accomplishes the same things for the same reasons.

This insight, first expressed by the mathematician Alan Turing, the computer scientists Alan Newell, Herbert Simon, and Marvin Minsky, and the philosophers Hilary Putnam and Jerry Fodor, is now called the computational theory of mind. It is one of the great ideas in intellectual history, for it solves one of the puzzles that make up the "mind-body problem": how to connect the ethereal world of meaning and intention, the stuff of our mental lives, with a physical hunk of matter like the brain. Why did Bill get on the bus? Because he wanted to visit his grandmother and knew the bus would take him there. No other answer will do. If he hated the sight of his grandmother, or if he knew the route had changed, his body would not be on that bus. For millennia this has been

a paradox. Entities like "wanting to visit one's grandmother" and "knowing the bus goes to Grandma's house" are colorless, odorless, and tasteless. But at the same time they are *cause*s of physical events, as potent as any billiard ball clacking into another.

The computational theory of mind resolves the paradox. It says that beliefs and desires are *information*, incarnated as configurations of symbols. The symbols are the physical states of bits of matter, like chips in a computer or neurons in the brain. They symbolize things in the world because they are triggered by those things via our sense organs, and because of what they do once they are triggered. If the bits of matter that constitute a symbol are arranged to bump into the bits of matter constituting another symbol in just the right way, the symbols corresponding to one belief can give rise to new symbols corresponding to another belief logically related to it, which can give rise to symbols corresponding to other beliefs, and so on. Eventually the bits of matter constituting a symbol bump into bits of matter connected to the muscles, and behavior happens. The computational theory of mind thus allows us to keep beliefs and desires in our explanations of behavior while planting them squarely in the physical universe. It allows meaning to cause and be caused.

The computational theory of mind is indispensable in addressing the questions we long to answer. Neuroscientists like to point out that all parts of the cerebral cortex look pretty much alike—not only the different parts of the human brain, but the brains of different animals. One could draw the conclusion that all mental activity in all animals is the same. But a better conclusion is that we cannot simply look at a patch of brain and read out the logic in the intricate pattern of connectivity that makes each part do its separate thing. In the same way that all books are physically just different combinations of the same seventy-five or so characters, and all movies are physically just different patterns of charges along the tracks of a videotape, the mammoth tangle of spaghetti of the brain may all look alike when examined strand by strand. The content of a book or a movie lies in the *pattern* of ink marks or magnetic charges, and is apparent only when the piece is read or seen. Similarly, the content of brain activity lies in the patterns of connections and patterns of activity among the neurons. Minute differences in the details of the connections may cause similar-looking brain patches to implement very different programs. Only when the program is run does the coherence become evident. As Tooby and Cosmides have written,

> There are birds that migrate by the stars, bats that echolocate, bees that compute the variance of flower patches, spiders that spin webs, humans that speak, ants that farm, lions that hunt in teams, cheetahs that hunt alone, monogamous gibbons, polyandrous seahorses, polygynous gorillas. . . . There are millions of animal species on earth, each with a different set of cognitive programs. *The same basic neural tissue embodies all of these programs*, and it could support many others as well. Facts about the properties of neurons, neurotransmitters, and cellular development cannot tell you which of these millions of programs the human mind contains. Even if all neural activity is the expression of a uniform process at the cellular level, it is the arrangement of neurons—into bird song templates or web-spinning programs—that matters.

That does not imply, of course, that the brain is irrelevant to understanding the mind! Programs are assemblies of simple information-processing units—tiny circuits that can add, match a pattern, turn on some other circuit, or do other elementary logical and mathematical operations. What those microcircuits can do depends only on what they are made of. Circuits made from neurons cannot do exactly the same things as circuits made from silicon, and vice versa. For example, a silicon circuit is faster than a neural circuit, but a neural circuit can match a larger pattern than a silicon one. These differences ripple up through the programs built from the circuits and affect how quickly and easily the programs do various things, even if they do not determine exactly which things they do. My point is not that prodding brain tissue is irrelevant to understanding the mind, only that it is not enough. Psychology, the analysis of mental software, will have to burrow a considerable way into the mountain before meeting the neurobiologists tunneling through from the other side.

The computational theory of mind is not the same thing as the despised "computer metaphor." As many critics have pointed out, computers are serial, doing one thing at a time; brains are parallel, doing millions of things at once. Computers are fast; brains are slow. Computer parts are reliable; brain parts are noisy. Computers have a limited number of connections; brains have trillions. Computers are assembled according to a blueprint; brains must assemble themselves. Yes, and computers come in putty-colored boxes and have AUTOEXEC.BAT files and run screen-savers with flying toasters, and brains do not. The claim is not that the brain is like commercially available computers. Rather, the claim is that brains and computers embody intelligence for some of the same

reasons. To explain how birds fly, we invoke principles of lift and drag and fluid mechanics that also explain how airplanes fly. That does not commit us to an Airplane Metaphor for birds, complete with jet engines and complimentary beverage service.

Without the computational theory, it is impossible to make sense of the evolution of the mind. Most intellectuals think that the human mind must somehow have escaped the evolutionary process. Evolution, they think, can fabricate only stupid instincts and fixed action patterns: a sex drive, an aggression urge, a territorial imperative, hens sitting on eggs and ducklings following hulks. Human behavior is too subtle and flexible to be a product of evolution, they think; it must come from somewhere else—from, say, "culture." But if evolution equipped us not with irresistible urges and rigid reflexes but with a neural computer, everything changes. A program is an intricate recipe of logical and statistical operations directed by comparisons, tests, branches, loops, and subroutines embedded in subroutines. Artificial computer programs, from the Macintosh user interface to simulations of the weather to programs that recognize speech and answer questions in English, give us a hint of the finesse and power of which computation is capable. Human thought and behavior, no matter how subtle and flexible, could be the product of a very complicated program, and that program may have been our endowment from natural selection. The typical imperative from biology is not "Thou shalt . . . ," but "If . . . then . . . else."

The mind, I claim, is not a single organ but a system of organs, which we can think of as psychological faculties or mental modules. The entities now commonly evoked to explain the mind—such as general intelligence, a capacity to form culture, and multipurpose learning strategies—will surely go the way of protoplasm in biology and of earth, air, fire, and water in physics. These entities are so formless, compared to the exacting phenomena they are meant to explain, that they must be granted near-magical powers. When the phenomena are put under the microscope, we discover that the complex texture of the everyday world is supported not by a single substance but by many layers of elaborate machinery. Biologists long ago replaced the concept of an all-powerful protoplasm with the concept of functionally specialized

mechanisms. The organ systems of the body do their jobs because each is built with a particular structure tailored to the task. The heart circulates the blood because it is built like a pump; the lungs oxygenate the blood because they are built like gas exchangers. The lungs cannot pump blood and the heart cannot oxygenate it. This specialization goes all the way down. Heart tissue differs from lung tissue, heart cells differ from lung cells, and many of the molecules making up heart cells differ from those making up lung cells. If that were not true, our organs would not work.

A jack-of-all-trades is master of none, and that is just as true for our mental organs as for our physical organs. The robot challenge makes that clear. Building a robot poses many software engineering problems, and different tricks are necessary to solve them.

Take our first problem, the sense of sight. A seeing machine must solve a problem called inverse optics. Ordinary optics is the branch of physics that allows one to predict how an object with a certain shape, material, and illumination projects the mosaic of colors we call the retinal image. Optics is a well-understood subject, put to use in drawing, photography, television engineering, and more recently, computer graphics and virtual reality. But the brain must solve the *opposite* problem. The input is the retinal image, and the output is a specification of the objects in the world and what they are made of—that is, what we know we are seeing. And there's the rub. Inverse optics is what engineers call an "ill-posed problem." It literally has no solution. Just as it is easy to multiply some numbers and announce the product but impossible to take a product and announce the numbers that were multiplied to get it, optics is easy but inverse optics impossible. Yet your brain does it every time you open the refrigerator and pull out a jar. How can this be?

The answer is that *the brain supplies the missing information*, information about the world we evolved in and how it reflects light. If the visual brain "assumes" that it is living in a certain kind of world—an evenly lit world made mostly of rigid parts with smooth, uniformly colored surfaces—it can make good guesses about what is out there. As we saw earlier, it's impossible to distinguish coal from snow by examining the brightnesses of their retinal projections. But say there is a module for perceiving the properties of surfaces, and built into it is the following assumption: "The world is smoothly and uniformly lit." The module can solve the coal-versus-snow problem in three steps: subtract out any gradient of brightness from one edge of the scene to the other; estimate the average level of brightness of

the whole scene; and calculate the shade of gray of each patch by subtracting its brightness from the average brightness. Large positive deviations from the average are then seen as white things, large negative deviations as black things. If the illumination really is smooth and uniform, those perceptions will register the surfaces of the world accurately. Since Planet Earth has, more or less, met the even-illumination assumption for eons, natural selection would have done well by building the assumption in.

The surface-perception module solves an unsolvable problem, but at a price. The brain has given up any pretense of being a general problem-solver. It has been equipped with a gadget that perceives the nature of surfaces in typical earthly viewing conditions because it is specialized for that parochial problem. Change the problem slightly and the brain no longer solves it. Say we place a person in a world that is not blanketed with sunshine but illuminated by a cunningly arranged patchwork of light. If the surface-perception module assumes that illumination is even, it should be seduced into hallucinating objects that aren't there. Could that really happen? It happens every day. We call these hallucinations slide shows and movies and television (complete with the illusory black I mentioned earlier). When we watch TV, we stare at a shimmering piece of glass, but our surface-perception module tells the rest of our brain that we are seeing real people and places. The module has been unmasked; it does not apprehend the nature of things but relies on a cheat-sheet. That cheat-sheet is so deeply embedded in the operation of our visual brain that we cannot erase the assumptions written on it. Even in a lifelong couch potato, the visual system never "learns" that television is a pane of glowing phosphor dots, and the person never loses the illusion that there is a world behind the pane.

Our other mental modules need their own cheat-sheets to solve their unsolvable problems. A physicist who wants to figure out how the body moves when muscles are contracted has to solve problems in kinematics (the geometry of motion) and dynamics (the effects of forces). But a brain that has to figure out how to contract muscles to get the body to move has to solve problems in *inverse* kinematics and *inverse* dynamics—what forces to apply to an object to get it to move in a certain trajectory. Like inverse optics, inverse kinematics and dynamics are ill-posed problems. Our motor modules solve them by making extraneous but reasonable assumptions—not assumptions about illumination, of course, but assumptions about bodies in motion.

Our common sense about other people is a kind of intuitive psychol-

ogy—we try to infer people's beliefs and desires from what they do, and try to predict what they will do from our guesses about their beliefs and desires. Our intuitive psychology, though, must make the assumption that other people *have* beliefs and desires; we cannot sense a belief or desire in another person's head the way we smell oranges. If we did not see the social world through the lens of that assumption, we would be like the Samaritan I robot, which sacrificed itself for a bag of lima beans, or like Samaritan II, which went overboard for any object with a human-like head, even if the head belonged to a large wind-up toy. (Later we shall see that people suffering from a certain syndrome lack the assumption that people have minds and *do* treat other people as wind-up toys.) Even our feelings of love for our family members embody a specific assumption about the laws of the natural world, in this case an inverse of the ordinary laws of genetics. Family feelings are designed to help our genes replicate themselves, but we cannot see or smell genes. Scientists use forward genetics to deduce how genes get distributed among organisms (for example, meiosis and sex cause the offspring of two people to have fifty percent of their genes in common); our emotions about kin use a kind of inverse genetics to guess which of the organisms we interact with are likely to share our genes (for example, if someone appears to have the same parents as you do, treat the person as if their genetic well-being overlaps with yours). I will return to all these topics in later chapters.

The mind has to be built out of specialized parts because it has to solve specialized problems. Only an angel could be a general problem-solver; we mortals have to make fallible guesses from fragmentary information. Each of our mental modules solves its unsolvable problem by a leap of faith about how the world works, by making assumptions that are indispensable but indefensible—the only defense being that the assumptions worked well enough in the world of our ancestors.

The word "module" brings to mind detachable, snap-in components, and that is misleading. Mental modules are not likely to be visible to the naked eye as circumscribed territories on the surface of the brain, like the flank steak and the rump roast on the supermarket cow display. A mental module probably looks more like roadkill, sprawling messily over the bulges and crevasses of the brain. Or it may be broken into regions that are interconnected by fibers that make the regions act as a unit. The beauty of information processing is the flexibility of its demand for real estate. Just as a corporation's management can be scattered across sites

linked by a telecommunications network, or a computer program can be fragmented into different parts of the disk or memory, the circuitry underlying a psychological module might be distributed across the brain in a spatially haphazard manner. And mental modules need not be tightly sealed off from one another, communicating only through a few narrow pipelines. (That is a specialized sense of "module" that many cognitive scientists have debated, following a definition by Jerry Fodor.) Modules are defined by the special things they do with the information available to them, not necessarily by the kinds of information they have available.

So the metaphor of the mental module is a bit clumsy; a better one is Noam Chomsky's "mental organ." An organ of the body is a specialized structure tailored to carry out a particular function. But our organs do not come in a bag like chicken giblets; they are integrated into a complex whole. The body is composed of systems divided into organs assembled from tissues built out of cells. Some kinds of tissues, like the epithelium, are used, with modifications, in many organs. Some organs, like the blood and the skin, interact with the rest of the body across a widespread, convoluted interface, and cannot be encircled by a dotted line. Sometimes it is unclear where one organ leaves off and another begins, or how big a chunk of the body we want to call an organ. (Is the hand an organ? the finger? a bone in the finger?) These are all pedantic questions of terminology, and anatomists and physiologists have not wasted their time on them. What is clear is that the body is not made of Spam but has a heterogeneous structure of many specialized parts. All this is likely to be true of the mind. Whether or not we establish exact boundaries for the components of the mind, it is clear that it is not made of mental Spam but has a heterogeneous structure of many specialized parts.