During my family's stay in England, we rented a furnished house while the owners were away. One day, our landlady returned to the house to get some personal papers. She walked over to her filing cabinet and attempted to open the top drawer. It wouldn't open. She pushed it forward and backward, right and left, up and down, without success. I offered to help. I wiggled the drawer. Then I twisted the front panel, pushed down hard, and banged the front with the palm of one hand. The cabinet drawer slid open. "Oh," she said, "I'm sorry. I am so bad at mechanical things."

I have studied people making errors—sometimes serious ones—with mechanical devices, light switches and fuses, computer operating systems and word processors, even airplanes and nuclear power plants. Invariably people feel guilty and either try to hide the error or blame themselves for "stupidity" or "clumsiness." I often have difficulty getting permission to watch: nobody likes to be observed performing badly. I point out that the design is faulty and that others make the same errors. Still, if the task appears simple or trivial, then people blame themselves. It is as if they take perverse pride in thinking of themselves as mechanically incompetent.

I once was asked by a large computer company to evaluate a brand new product. I spent a day learning to use it and trying it out on various problems. In using the keyboard to enter data, it was necessary to differentiate between the the "return" key and the "enter" key. If the wrong key was typed, the last few minutes' work was irrevocably lost.

I pointed this problem out to the designer, explaining that I myself had made the error frequently and that my analyses indicated that this was very likely to be a frequent error among users. The designer's first response was: "Why did you make that error? Didn't you read the manual?" He proceeded to explain the different functions of the two keys.

"Yes, yes," I explained, "I understand the two keys, I simply confuse them. They have similar functions, are located in similar locations on the keyboard, and as a skilled typist, I often hit "return" automatically, without thought. Certainly others have had similar problems."

"Nope," said the designer. He claimed that I was the only person who had ever complained, and the company's secretaries had been using the system for many months. I was skeptical, so we went together to some of the secretaries and asked them whether they had ever hit the "return" key when they should have hit "enter." And did they ever lose their work as a result?

"Oh, yes," said the secretaries, "we do that a lot."

"Well, how come nobody ever said anything about it?" we asked the secretaries. After all, they were encouraged to report all problems with the system.

The reason was simple: when the system stopped working or did something strange, the secretaries dutifully reported it as a problem. But when they made the "return" versus "enter" error, they blamed themselves. After all, they had been told what to do. They had simply erred.

Of course, people do make errors. Complex devices will always require some instruction, and someone using them without instruction should expect to make errors and to be confused. But designers should take special pains to make errors as cost-free as possible. Here is my credo about errors:
If an error is possible, someone will make it. The designer must assume that all possible errors will occur and design so as to minimize the chance of the error in the first place, or its effects once it gets made. Errors should be easy to detect, they should have minimal consequences, and, if possible, their effects should be reversible.

Misconceptions of Everyday Life

Our lives are filled with misconceptions. This should not be surprising: we must frequently deal with unfamiliar situations. Psychologists love errors and misconceptions, for they give important clues about the organization and operation of our minds. Many everyday misunderstandings are classified as “naive” or “folk” understandings. And not just plain folk hold these misconceptions: Aristotle developed an entire theory of physics that physicists find quaint and amusing. Yet Aristotle’s theories correspond much better to common-sense, everyday observations than do the highly refined and abstract theories we are taught in school. Aristotle developed what we might call na"ive physics. It is only when you study the esoteric world of physics that you learn what is “correct” and are able to understand why the “naive” view is wrong.

ARISTOTLE’S NAIVE PHYSICS

For example, Aristotle thought that moving objects kept moving only if something kept pushing them. Today’s physicist says nonsense: a moving object continues to move unless some force is exerted to stop it. This is Newton’s first law of motion, and it contributed to the development of modern physics. Yet anyone who has ever pushed a heavy box along a street or, for that matter, hiked for miles into the wilderness, knows that Aristotle was right: if you don’t keep on pushing, the movement stops. Of course, Newton and his successors assume the absence of friction and air. Aristotle lived in a world where there was always friction and air resistance. Once friction is involved, then objects in motion tend to stop unless you keep pushing. Aristotle’s theory may be bad physics, but it describes reasonably well what we can see in the real world. Think about how you might answer the following questions.

1. I take a pistol and, carefully aiming it on a level, horizontal line, I fire a bullet. With my other hand, I hold a bullet so that the bullet in the pistol and the one in my hand are exactly the same distance from the ground. I drop the bullet at the same instant as I fire the pistol. Which bullet hits the ground first?

2. Imagine someone running across a field carrying a ball. As you watch, the runner drops the ball. Which path (a, b, or c in figure 2.1) does the ball take as it falls to the ground?

The physicist says the answer to the bullet problem is trivial: both bullets hit the ground at the same time. The fact that one bullet is traveling horizontally very rapidly has absolutely no effect on how fast it falls downward. Why should we accept that answer? Shouldn’t the speeding bullet develop some lift—sort of like an airplane—so that it will stay up a bit longer because it is kept up by the air? Who knows? The theory of physics is based upon a situation where there is no air. The popular misconception is that the pistol bullet will hit the ground long after the dropped bullet; yet this naive view doesn’t seem so strange.

2.1 A Running Man Drops a Ball. Which path does the ball take as it falls to the ground, path A, B, or C? When this question was asked of sixth-grade students in Boston schools, only 3 percent answered A, the right answer; the others were evenly divided between B and C. Even high school students did not do well: of forty-one students who had just studied Newtonian mechanics for a month and a half, only 20 percent got the right answer; the others were almost equally divided between B and C. (The study was performed by White & Horwitz, 1987. The figure is reprinted from Intuitive Physics by McComiskey. Copyright © 1983 by Scientific American, Inc. All rights reserved.)

36 The Design of Everyday Things

two: The Psychology of Everyday Actions 37
In the case of the falling ball, our prediction is that the ball will drop straight down. In fact, the falling ball follows trajectory A (figure 2.1). As it is carried by the runner, it is set into horizontal motion. It then maintains the same forward speed upon being released, even as it also falls to the ground. 3

Naive physics—and naive views of psychology and other fields—are often sensible, even if wrong. But at times they can get us into trouble. Yet we must have a way to digest the unfamiliar, for people are explanatory creatures.

PEOPLE AS EXPLANATORY CREATURES

Mental models, our conceptual models of the way objects work, events take place, or people behave, result from our tendency to form explanations of things. These models are essential in helping us understand our experiences, predict the outcomes of our actions, and handle unexpected occurrences. We base our models on whatever knowledge we have, real or imaginary, naive or sophisticated.

Mental models are often constructed from fragmentary evidence, with but a poor understanding of what is happening, and with a kind of naive psychology that postulates causes, mechanisms, and relationships even where there are none. Some faulty models lead to the frustrations of everyday life, as in the case of my unseetable refrigerator, where my mental model of its operation (figure 1.9 A) did not correspond to reality (figure 1.9 B). Far more serious are faulty models of such complex systems as an industrial plant or passenger airplane. Misunderstanding there can lead to devastating accidents.

Consider the room thermostat. How does it work? Here is a device that offers almost no evidence of its operation except in a highly roundabout manner. We walk into a room and feel too cold; so we walk over to the thermostat and set it higher. Eventually we feel warmer. Note that the same thing applies to the temperature control for a cooking oven (or a pottery kiln, or an air conditioner, or almost any device whose temperature is to be regulated). Want to bake a cake, but the oven is off? Set the oven thermostat and the oven gets to the desired temperature. Is the room too hot? Set the thermostat on the air conditioner. Fine, but how does the thermostat work?

If you are in a cold room, in a hurry to get warm, will the room heat more quickly if you turn the thermostat all the way up? Or if you want the oven to reach its working temperature faster, should you turn the temperature dial all the way to maximum, then turn it down once the desired temperature is reached? Or to cool a room most quickly, should you set the air conditioner thermostat to its lowest temperature setting?

If you think that the room or oven will heat (or cool) faster if the thermostat is turned all the way to the maximum setting, you are wrong. You hold a folk theory of thermostats. There are two commonly held folk theories about thermostats: the timer theory and the valve theory. The timer theory proposes that the thermostat simply controls the relative proportion of time that the device stays on. Set the thermostat midway, and the device is on about half the time; set it all the way up and the device is on all the time. Hence, to heat or cool something most quickly, set the thermostat so that the device is on all the time. The valve theory proposes that the thermostat controls how much heat (or cold) comes out of the device. Turn the thermostat all the way up, and you get maximum heating or cooling. 4

The correct story is that the thermostat is just an on-off switch. It treats the heater, oven, and air conditioner as all-or-nothing devices that can be either fully on or fully off, with no in-between states. The thermostat turns the heater, oven, or air conditioner completely on—at full power—until the temperature setting on the thermostat is reached. Then it turns the unit completely off. Setting the thermostat at one extreme cannot affect how long it takes to reach the desired temperature. 5

The real point of the example is not that some people have erroneous theories; it is that everyone forms theories (mental models) to explain what they have observed. In the case of the thermostat, the design gives absolutely no hint as to the correct answer. In the absence of external information, people are free to let their imaginations run free as long as the mental models they develop account for the facts as they perceive them.

Blaming the Wrong Cause

"Look at this!" my colleague exclaimed to me, "My computer terminal is broken. The library did it! Every time I connect it to the library catalog I have trouble. Now I can't even use the terminal to read my computer mail anymore."

"That doesn't make sense," I replied. "You can't even turn on the
power to the terminal. How could a computer program possibly do that kind of damage?"

"All I know," he said, "is that everything was working fine until I tried to look up an author in the library catalog using that new library program, and then my terminal stopped working. I always have trouble with that program. And this is simply too much of a coincidence to be anything else."

Well, it was a coincidence. It turns out that the power supply to the terminal had burned out, a fact that had nothing to do with the computer program. Coincidence is enough to set the causal wheels rolling.

Earlier I suggested that people have a tendency to blame themselves for difficulties with technology. Actually, the point is a bit more complicated. People do tend to find causes for events, and just what they assign as the cause varies. In part people tend to assign a causal relation whenever two things occur in succession. If I do some action A just prior to some result R, then I conclude that A must have caused R, even if, as in the example above, there really was no relationship between the two. The story is more complex when we intend an action to produce a desired result and fail, and there are problems when we have done the action through some intermediate mechanism.

Just where do we put the blame for failure? The answer is not clear. The psychology of blame (or, to be more accurate, of attribution) is complex and not fully understood. In part, there seems to have to be some perceived causal relationship between the thing being blamed and the result. The word perceived is critical: the causal relationship does not have to exist; the person simply has to think it is there. Sometimes we attribute the cause to things that had nothing to do with the action. And sometimes we ignore the real culprit.

One major aspect of the assignment of blame is that we frequently have little information on which to make the judgment, and what little we have may be wrong. As a result, blame or credit can be assessed almost independently of reality. Here is where the apparent simplicity of everyday objects causes problems. Suppose I try to use an everyday thing, but I can't: Where is the fault, in my action or in the thing? We are apt to blame ourselves. If we believe that others are able to use the device and if we believe that it is not very complex, then we conclude that any difficulties must be our own fault. Suppose the fault really lies in the device, so that lots of people have the same problems. Because everyone perceives the fault to be his or her own, nobody wants to admit to having trouble. This creates a conspiracy of silence, maintaining the feelings of guilt and helplessness among users.

Interestingly enough, the common tendency to blame ourselves for failures with everyday objects goes against the normal attributions people make. In general, it has been found that people attribute their own problems to the environment, those of other people to their personalities.

Here is a made-up example. Consider Tom, the office terror. Today Tom got to work late, slammed the door to his office, and yelled at his colleagues. "Ah," his colleagues and staff said, "there he goes again. He's so excitable—always gets mad at the slightest thing."

Now consider Tom's point of view. "I really had a hard day," Tom explains. "I woke up late because when my clock radio turned on, I tried to hit the snooze bar to give me five minutes' more sleep; instead I reset the time so that I overslept for a whole hour. That wasn't my fault—the radio's badly designed. I didn't even have time for my morning coffee. I couldn't find a close parking spot because I was late. And then because I was in such a rush I dropped my papers all over the street and got them dirty. Then when I went to get a cup of coffee from the office machine, it was all out. None of this was my fault—I had a run of really bad events. Yes, I was a bit curt with my colleagues, but who wouldn't be under the same circumstances? Surely they understand."

But Tom's colleagues see a different picture. They don't have access to his inner thoughts or even to his morning's activities. All they see is that Tom yelled at them simply because the office coffee machine was empty. And this reminds them of another time when the same thing happened. "He does that all the time," they conclude, "always blowing up over the most minor events." The events are the same events, but there are two different points of view and two different interpretations. The protagonist, Tom, views his actions as sensible responses to the trials of life. The onlooker views Tom's actions as a result of his explosive, irascible personality.

It seems natural for people to blame their own misfortunes on the environment. It seems equally natural to blame other people's misfortunes on their personalities. Just the opposite attribution, by the way, is made when things go well. When things go right, people credit their own forceful personalities and intelligence: "I really did a good job today; no wonder we finished the project so well." The onlookers do
the reverse. When they see things go well for someone else, they credit the environment: “Joan really was lucky today; she just happened to be standing there when the boss came by, so she got all the credit for the project work. Some people have all the luck.”

In all cases, whether a person is inappropriately accepting blame for the inability to work simple objects or attributing behavior to environment or personality, a faulty mental model is at work.

LEARNED HELPLESSNESS

The phenomenon called learned helplessness may help explain the self-blame. It refers to the situation in which people experience failure at a task, often numerous times. As a result, they decide that the task cannot be done, at least not by them; they are helpless. They stop trying. If this feeling covers a group of tasks, the result can be severe difficulties coping with life. In the extreme case, such learned helplessness leads to depression and to a belief that the person cannot cope with everyday life at all. Sometimes all that it takes to get such a feeling of helplessness is a few experiences that accidentally turn out bad. The phenomenon has been most frequently studied as a precursor to the clinical problem of depression, but it might easily arise with a few bad experiences with everyday objects.

TAUGHT HELPLESSNESS

Do the common technology and mathematics phobias result from a kind of learned helplessness? Could a few instances of failure in what appear to be straightforward situations generalize to every technological object, every mathematics problem? Perhaps. In fact, the design of everyday things (and the design of mathematics courses) seems almost guaranteed to cause this. We could call this phenomenon taught helplessness.

With badly designed objects—constructed so as to lead to misunderstanding—faulty mental models, and poor feedback, no wonder people feel guilty when they have trouble using objects, especially when they perceive (even if incorrectly) that nobody else is having the same problems. Or consider the normal mathematics curriculum, which continues relentlessly on its way, each new lesson assuming full knowledge and understanding of all that has passed before. Even though each point may be simple, once you fall behind it is hard to catch up. The result: mathematics phobia. Not because the material is difficult, but because it is taught so that difficulty in one stage hinders further progress. The problem is that once failure starts, it soon generalizes by self-blame to all of mathematics. Similar processes are at work with technology. The vicious cycle starts: if you fail at something, you think it is your fault. Therefore you think you can’t do that task. As a result, next time you have to do the task, you believe you can’t so you don’t even try. The result is that you can’t, just as you thought. You’re trapped in a self-fulfilling prophecy.

The Nature of Human Thought and Explanation

It isn’t always easy to tell just where the blame for a problem should be placed. A number of dramatic accidents have come about, in part, from the false assessment of blame in a situation. Highly skilled, well-trained people are using complex equipment when suddenly something goes wrong. They have to figure out what the problem is. Most industrial equipment is pretty reliable. When the instruments indicate that something is wrong, one has to consider the possibility that the instruments themselves are wrong. Often this is the correct assessment. But when operators mistakenly blame the instruments for an actual equipment failure, the situation is ripe for a major accident.

It is spectacularly easy to find examples of false assessment in industrial accidents. Analysts come in well after the fact, knowing what actually did happen; with hindsight, it is almost impossible to understand how the people involved could have made the mistake. But from the point of view of the person making decisions at the time, the sequence of events is quite natural.

At the Three Mile Island nuclear power plant, operators pushed a button to close a valve; the valve had been opened (properly) to allow excess water to escape from the nuclear core. In fact, the valve was deficient, so it didn’t close. But a light on the control panel indicated that the valve position was closed. The light actually didn’t monitor the valve, only the electrical signal to the valve, a fact known by the operators. Still, why suspect a problem? The operators did look at the temperature in the pipe leading from the valve; it was high, indicating that fluid was still flowing through the closed valve. Ah, but the opera-
tors knew that the valve had been leaky, so the leak would explain the high temperature; but the leak was known to be small, and operators assumed that it wouldn't affect the main operation. They were wrong, and the water that was able to escape from the core added significantly to the problems of that nuclear disaster. I think the operators' assessment was perfectly reasonable: the fault was in the design of the lights and in the equipment that gave false evidence of a closed valve.

Similar misinterpretations take place all the time. I have studied a number of airline accidents. Consider the flight crew of the Lockheed L-1011 flying from Miami, Florida, to Nassau, Bahamas. The plane was over the Atlantic Ocean, about 110 miles from Miami, when the low oil pressure light for one of the three engines went on. The crew turned off the engine and turned around to go back to Miami. Eight minutes later, the low pressure lights for the remaining two engines also went on, and the instruments showed zero oil pressure and quantity in all three engines. What did the crew do now? They didn't believe it! After all, the pilot correctly said later, the likelihood of simultaneous oil exhaustion in all three engines was "one in millions I would think." At the time, sitting in the airplane, simultaneous failure did seem most unlikely. Even the National Transportation Safety Board declared, "The analysis of the situation by the flight crew was logical, and was what most pilots probably would have done if confronted by the same situation."

What happened? The second and third engines were indeed out of oil, and they failed. So there were no operating engines: one had been turned off when its gauge registered low, the other two had failed. The pilots prepared the plane for an emergency landing on the water. The pilots were too busy to instruct the flight crew properly, so the passengers were not prepared. There was semi-hysteria in the passenger cabin. At the last minute, just as the plane was about to ditch in the ocean, the pilots managed to restart the first engine and to land safely at Miami. Then that engine failed at the end of the runway.

Why did all three engines fail? Three missing O-rings, one missing from each of three oil plugs, allowed all the oil to seep out. The O-rings were put in by two different people who worked on the three engines (one for the two plugs on the wings, the other for the plug on the tail). How did both workers make the same mistake? Because the normal method by which they got the oil plugs had been changed that day. The whole tale is very instructive, for there were four major failures of different sorts, from the omission of the O-rings, to the inadequacy of the maintenance procedures, to the false assessment of the problem, to the poor handling of the passengers. Fortunately, nobody was injured. The analysts of the National Transportation Safety Board got to write a fascinating report.

I've misinterpreted signals, as I'm sure most people have. My family was driving from San Diego to Mammoth, California, a ski area about 500 miles north; a ten- to twelve-hour drive. As we drove, we noticed more and more signs advertising the hotels and gambling casinos of Las Vegas, Nevada. "Strange," we said, "Las Vegas always did advertise a long way off—there is even a billboard in San Diego—but this seems excessive, advertising on the road to Mammoth." We stopped for gasoline and continued on our journey. Only later, when we tried to find a place to eat supper, did we discover that we had taken the wrong turn nearly two hours earlier, before we had stopped for gasoline, and that we were on the road to Las Vegas, not the road to Mammoth. We had to backtrack the entire two-hour segment, wasting four hours of driving. It's humorous now; it wasn't then.

Find an explanation, and we are happy. But our explanations are based on analogy with past experience, experience that may not apply in the current situation. In the Three Mile Island incident, past experience with the leaky valve explained away the discrepant temperature reading; on the flight from Miami to Nassau, the pilots' lack of experience with simultaneous oil pressure failure triggered their belief that the instruments must be faulty; in the driving story, the prevalence of billboards for Las Vegas seemed easily explained. Once we have an explanation—correct or incorrect—for otherwise discrepant or puzzling events, there is no more puzzle, no more discrepancy. As a result, we are complacent, at least for a while.

How People Do Things:
The Seven Stages of Action

I am in Italy, at a conference. I watch the next speaker attempt to thread a film onto a projector that he has never used before. He puts the reel into place, then takes it off and reverses it. Another person comes to help. Jointly they thread the film through the projector and hold the free end, discussing how to put it on the takeup reel. Two
more people come over to help, and then another. The voices grow louder, in three languages: Italian, German, and English. One person investigates the controls, manipulating each and announcing the result. Confusion mounts. I can no longer observe all that is happening. The conference organizer comes over. After a few moments he turns and faces the audience, which has been waiting patiently in the auditorium. “Ahem,” he says, is anybody expert in projectors?” Finally, fourteen minutes after the speaker had started to thread the film (and eight minutes after the scheduled start of the session) a blue-coated technician appears. He scowls, then promptly takes the entire film off the projector, rethreads it, and gets it working.

What makes something—like threading the projector—difficult to do? To answer this question, the central one of this book, we need to know what happens when someone does something. We need to examine the structure of an action.

The basic idea is simple. To get something done, you have to start with some notion of what is wanted—the goal that is to be achieved. Then, you have to do something to the world, that is, take action to move yourself or manipulate someone or something. Finally, you check to see that your goal was made. So there are four different things to consider: the goal, what is done to the world, the world itself, and the check of the world. The action itself has two major aspects: doing something and checking. Call these execution and evaluation (figure 2.2).

Real tasks are not quite so simple. The original goal may be imprecisely specified—perhaps “get something to eat,” “get to work,” “get dressed,” “watch television.” Goals do not state precisely what to do—where and how to move, what to pick up. To lead to actions goals must be transformed into specific statements of what is to be done, statements that I call intentions. A goal is something to be achieved, often vaguely stated. An intention is a specific action taken to get to the goal. Yet even intentions are not specific enough to control actions.

Suppose I am sitting in my armchair, reading a book. It is dusk, and the light has gotten dimmer and dimmer. I decide I need more light (that is the goal: get more light). My goal has to be translated into the intention that states the appropriate action in the world: push the switch button on the lamp. There’s more: I need to specify how to move my body, how to stretch to reach the light switch, how to extend my finger to push the button (without knocking over the lamp). The goal
has to be translated into an intention, which in turn has to be made into a specific action sequence, one that can control my muscles. Note that I could satisfy my goal with other action sequences, other intentions. If someone walked into the room and passed by the lamp, I might alter my intention from pushing the switch button to asking the other person to do it for me. The goal hasn’t changed, but the intention and resulting action sequence have.

The specific actions bridge the gap between what we would like to have done (our goals and intentions) and all possible physical actions. After we specify what actions to make, we must actually do them—the stage of execution. All in all, there are three stages that follow from the goal: intention, action sequence, and execution (figure 2.3).

The evaluation side of things, checking up on what happened, has three stages: first, perceiving what happened in the world; second, trying to make sense of it (interpreting it); and, finally, comparing what happened with what was wanted (figure 2.4).

There we have it. Seven stages of action: one for goals, three for execution, and three for evaluation.

- Forming the goal
- Forming the intention
- Specifying an action
- Executing the action
- Perceiving the state of the world
- Interpreting the state of the world
- Evaluating the outcome

The seven stages form an approximate model, not a complete psychological theory. In particular, the stages are almost certainly not discrete entities. Most behavior does not require going through all stages in sequence, and most activities will not be satisfied by single actions. There must be numerous sequences, and the whole activity may last hours or even days. There is a continual feedback loop, in which the results of one activity are used to direct further ones, in which goals lead to subgoals, intentions lead to subintentions. There are activities in which goals are forgotten, discarded, or reformulated.

For many everyday tasks, goals and intentions are not well specified: they are opportunistic rather than planned. Opportunistic actions are those in which the behavior takes advantage of the circumstances. Rather than engage in extensive planning and analysis, the person goes about the day’s activities and performs the intended actions if the relevant opportunity arises. Thus, we may not go out of our way to go to a shop, or to the library, or to ask a question of a friend. Rather, we go through the day’s activities, and if we find ourselves at the shop, near the library, or encountering the friend, then we allow the opportunity to trigger the relevant activity. Otherwise, the task remains undone. Only in the case of crucial tasks do we make special efforts to ensure that they get done. Opportunistic actions are less precise and certain than specified goals and intentions, but they result in less mental effort, less inconvenience, and perhaps more interest.

The seven-stage process of action can be started at any point. People do not always behave as full, logical, reasoning organisms, starting with high-level goals and working to achieve them. Our goals are often ill-formed and vague. We may respond to the events of the world (in what is called data-driven behavior) rather than to think out plans and goals. An event in the world may trigger an interpretation and a resulting response. Actions may be executed before they are fully developed. In fact, some of us adjust our lives so that the environment can control our behavior. For example, sometimes when I must do an important task, I make a formal, public promise to get it done by a certain date. I make sure that I will be reminded of the promise. And then, hours before the deadline, I actually get to work and do the job. This kind of behavior is fully compatible with the seven-stage analysis.

The Gulfs of Execution and Evaluation

Remember the movie projector story? People’s problems threading the projector did not come from a lack of understanding of the goal or the task. It did not come from deep, subtle complexity. The difficulty lay entirely in determining the relationship between the intended actions and the mechanisms of the projector, in determining the functions of each of the controls, in determining what specific manipulation of each control enabled each function, and in deciding by the sights, sounds, lights, and movements of the projector whether the intended actions were being done successfully. The users had a problem with mappings and feedback, as they would have with the projector in figure 2.6.

The projector story is only an extreme case of the difficulties faced in the conduct of many tasks. For a surprisingly large number of every-
day tasks, the difficulty resides entirely in deriving the relationships between the mental intentions and interpretations and the physical actions and states. There are several gulf that separate mental states from physical ones. Each gulf reflects one aspect of the distance between the mental representations of the person and the physical components and states of the environment. And these gulf present major problems for users.6

THE GULF OF EXECUTION

Does the system provide actions that correspond to the intentions of the person? The difference between the intentions and the allowable actions is the Gulf of Execution. One measure of this gulf is how well the system allows the person to do the intended actions directly, without extra effort: Do the actions provided by the system match those intended by the person?

Consider the movie projector example: one problem resulted from the Gulf of Execution. The person wanted to set up the projector. Ideally, this would be a simple thing to do. But no, a long, complex sequence was required. It wasn’t at all clear what actions had to be done to accomplish the intentions of setting up the projector and showing the film.

Self-threading projectors do exist. These nicely bridge the gulf. Or look at VCRs. They have the same mechanical problem as film projectors: the videotape has to be threaded through their mechanism. But the solution is to hide this part of the system, to put the task on the machine, not the person. So the machinery bridges the gulf. All the user has to do is to plop in the cartridge and push the start button. It’s a pity the film companies are so far behind. Well, in a while it won’t matter. There won’t be any film, just videotape.

THE GULF OF EVALUATION

Does the system provide a physical representation that can be directly perceived and that is directly interpretable in terms of the intentions and expectations of the person? The Gulf of Evaluation reflects the amount of effort that the person must exert to interpret the physical state of the system and to determine how well the expectations and intentions have been met. The gulf is small when the system provides information about its state in a form that is easy to get, is easy to interpret, and matches the way the person thinks of the system.

In the movie projector example there was also a problem with the Gulf of Evaluation. Even when the film was in the projector, it was
difficult to tell if it had been threaded correctly. With VCRs all you have to know is whether the cartridge is properly inserted into the machine. If it isn't, usually it won't fit right: it sticks out obviously, and you know that things are not right.

But VCRs aren't perfect, either. I remember a conference speaker who pushed the start button on the VCR and told the audience to watch the screen. No picture. She fiddled with the machine, then called for help. One, then two, then three technicians appeared on the scene. They carefully checked the power connections, the leads to the VCR, the circuits. The audience waited impatiently, giggling. Finally the problem was found: there wasn't any tape in the VCR. No tape, no picture. The problem was that once the cartridge door to that particular VCR was shut there was no visible way to tell whether it contained a tape. Bad design. That Gulf of Evaluation sunk another user.

The gulf are present to an amazing degree in a variety of devices. Usually the difficulties are unremarked and invisible. The users either take the blame themselves (in the case of things they believe they should be capable of using, such as water faucets, refrigerator temperature controls, stove tops, radio and television sets) or decide that they are incapable of operating the pesky devices (sewing machines, washing machines, digital watches, digital controls on household appliances, VCRs, audio sets). These are indeed the gadgets of everyday household use. None of them has a complex structure, yet many of them defeat the otherwise capable user.

The Seven Stages of Action as Design Aids

The seven-stage structure can be a valuable design aid, for it provides a basic checklist of questions to ask to ensure that the Gulls of Evaluation and Execution are bridged (figure 2.7).

In general, each stage of action requires its own special design strategies and, in turn, provides its own opportunity for disaster. It would be fun, were it not also so frustrating, to look over the world and gleefully analyze each deficiency. On the whole, as you can see in figure 2.7, the questions for each stage are relatively simple. And these, in turn, boil down to the principles of good design introduced in chapter 1.

- **Visibility**. By looking, the user can tell the state of the device and the alternatives for action.

<table>
<thead>
<tr>
<th>How Easily Can One:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determine The Function of the Device?</td>
</tr>
<tr>
<td>Tell What Actions Are Possible?</td>
</tr>
<tr>
<td>Determine Mapping from Intention to Physical Movement?</td>
</tr>
<tr>
<td>Perform the Action?</td>
</tr>
</tbody>
</table>

- **A good conceptual model**. The designer provides a good conceptual model for the user, with consistency in the presentation of operations and results and a coherent, consistent system image.

- **Good mappings**. It is possible to determine the relationships between actions and results, between the controls and their effects, and between the system state and what is visible.

- **Feedback**. The user receives full and continuous feedback about the results of actions.

Each point provides support for one or more of the seven stages of action. The next time you can't immediately figure out the shower control in a motel or work an unfamiliar television set or stove, remember that the problem is in the design. And the next time you pick up an unfamiliar object and use it smoothly and effortlessly on the first try, stop and examine it: the ease of use did not come about by accident. Someone designed the object carefully and well.
A friend kindly let me borrow his car. Just before I was about to leave, I found a note waiting for me: "I should have mentioned that to get the key out of the ignition the car needs to be in reverse." The car needs to be in reverse! If I hadn't seen the note, I never could have figured that out. There was no visible cue in the car: the knowledge needed for this trick had to reside in the head. If the driver lacks that knowledge, the key stays in the ignition forever.

It is easy to show the faulty nature of human knowledge and memory. A common classroom exercise in the United States demonstrates that students cannot recall the pairing of letters and numbers on their telephones. One of my graduate students found that when professional typists were given caps for typewriter keys, they could not arrange them in the proper configuration.1 American students dial telephones properly, and all those typists could type rapidly and accurately. Why the apparent discrepancy between the precision of behavior and the imprecision of knowledge? Because not all of the knowledge required for precise behavior has to be in the head. It can be distributed—partly in the head, partly in the world, and partly in the constraints of the world. Precise behavior can emerge from imprecise knowledge for four reasons.

1. Information is in the world. Much of the information a person needs to do a task can reside in the world. Behavior is determined by combining the information in memory (in the head) with that in the world.

2. Great precision is not required. Precision, accuracy, and completeness of knowledge are seldom required. Perfect behavior will result if the knowledge describes the information or behavior sufficiently to distinguish the correct choice from all others.

3. Natural constraints are present. The world restricts the allowed behavior. The physical properties of objects constrain possible operations: the order in which parts can go together and the ways in which an object can be moved, picked up, or otherwise manipulated. Each object has physical features—projections, depressions, screwthreads, appendages—that limit its relationships to other objects, operations that can be performed to it, what can be attached to it, and so on.

4. Cultural constraints are present. In addition to natural, physical constraints, society has evolved numerous artificial conventions that govern acceptable social behavior. These cultural conventions have to be learned, but once learned they apply to a wide variety of circumstances.

Because of these natural and artificial constraints, the number of alternatives for any particular situation is reduced, as are the amount and specificity of knowledge required within human memory.

In everyday situations, behavior is determined by the combination of internal knowledge and external information and constraints. People routinely capitalize on this fact. They can minimize the amount of material they must learn or the completeness, precision, accuracy, or depth of the learning. People can deliberately organize the environment to support their behavior. Some people with brain damage can function so well that even their co-workers may not be aware of their handicap. Nonreaders have been known to fool others, even in situations where their job presumably requires reading skills. They know what is expected of them, follow the behavior of their co-workers, and set up situations so that they do not need to read or so that their co-workers do the reading for them.

What is true in these extreme cases must certainly also be true of
ordinary people in ordinary situations: it is only the amount of reliance upon the external world that differs. There is a tradeoff between the amount of mental knowledge and the amount of external knowledge required in performing tasks. People are free to operate variously in allowing for this tradeoff.

**Precise Behavior from Imprecise Knowledge**

**INFORMATION IS IN THE WORLD**

Whenever information needed to do a task is readily available in the world, the need for us to learn it diminishes. For example, we lack knowledge about common coins, even though we recognize them just fine (figure 3.1). Or consider typing. Many typists have not memorized the keyboard. Usually each letter is labeled, so nontypists can hunt and peck letter by letter, relying on knowledge in the world and minimizing the time required for learning. The problem is that such typing is slow and difficult. With experience, of course, hunt-and-peck typists learn the positions of many of the letters on the keyboard, even without instruction, and typing speed increases notably, quickly surpassing handwriting speeds and, for some, reaching quite respectable rates. Peripheral vision and the feel of the keyboard provide some information about key locations. Frequently used keys become completely learned, infrequently used keys are not learned well, and the other keys are partially learned. But as long as the typist needs to watch the keyboard, the speed is limited. The knowledge is still mostly in the world, not in the head.

If a person needs to type large amounts of material regularly, further investment is worthwhile: a course, a book, or an interactive computer program. The important thing is to learn the proper placement of fingers on the keyboard, to learn to type without looking, to get knowledge of the keyboard from the world into the head. It takes several hours to learn the system and several months to become expert. But the payoff of all this effort is increased typing speed, increased accuracy, and decreased mental load and effort at the time of typing.

There is a tradeoff between speed and quality of performance and mental effort. Thus, in finding your way through a city, locating items in a store or house, or working complex machinery, the tradeoff can determine what needs to be learned. Because you know that the inform-
impossible to write down and difficult to teach. It is best taught by
demonstration and best learned through practice. Even the best teach-
ers cannot usually describe what they are doing. Procedural knowledge
is largely subconscious.

Knowledge from the world is usually easy to come by. Designers
provide a large number of memory aids. The letters on the typewriter
keyboard are one example. The lights and labels on controls act as
external memory aids, reminding the user of the purpose and state of
the control. Industrial equipment is replete with signal lights, indica-
tors, and other reminders. We make extensive use of written notes. We
place items in specific locations as reminders. In general, people struc-
ture the environment to provide a considerable amount of the informa-
tion required for something to be remembered.

Many people organize their lives in the world, creating a pile here,
a pile there, each indicating some activity to be done, some event in
progress. Probably everybody uses such a strategy to some extent. Look
around you at the variety of ways people structure their rooms and
desks. Many styles of organization are possible, but the physical
arrangement and visibility of the items frequently convey information
about relative importance. Want to do your friends a nasty turn? Do
them a favor—clean up their desks or rooms. Do this to some people
and you can completely destroy their ability to function.2

GREAT PRECISION IS NOT REQUIRED

Normally, people do not need precise memory information. People
can remember enough to distinguish one familiar coin from another
although they may be unable to remember the faces, pictures, and
words on the coins.3 But make more precise memory necessary and
you get havoc. Three countries have rediscovered this fact in recent
years: the United States, when it introduced the Susan B. Anthony
one-dollar coin; Great Britain, when it introduced the one-pound
coin; and France, when it introduced a new ten-franc coin. The new
U.S. dollar coin was confused with the existing twenty-five-cent
piece (the quarter), and the British pound coin was confused with the
existing five-pence piece. (The one-pound coin has the same diameter
as the five-pence piece, but is considerably thicker and heavier.) Here
is what happened in France:

“PARIS . . .” With a good deal of fanfare, the French government
released the new 10-franc coin (worth a little more than $1.50) on Oct.
22 [1986]. The public looked at it, weighed it, and began confusing it
so quickly with the half-franc coin (worth only 8 cents) that a cre-
sendo of fury and ridicule fell on both the government and the coin.

“Five weeks later, Minister of Finance Edouard Balladur suspended
circulation of the coin. Within another four weeks, he canceled it
altogether.

“In retrospect, the French decision seems so foolish that it is hard to
fathom how it could have been made. . . . After much study, designers
came up with a silver-colored coin made of nickel and featuring a
modernistic drawing by artist Joaquin Jimenez of a Gallic rooster on
one side and of Marianne, the female symbol of the French republic,
on the other. The coin was light, sported special ridges on its rim for
easy reading by electronic vending machines and seemed tough to
counterfeit.

“But the designers and bureaucrats were obviously so excited by
their creation that they ignored or refused to accept the new coin’s
similarity to the hundreds of millions of silver-colored, nickel-based
half-franc coins in circulation . . . [whose] size and weight were peri-
ously similar.”4

The confusions probably occurred because the users of coins formed
representations in their memory systems that were sufficiently precise
only to distinguish among the coins that they actually had to use. It is
a general property of memory that we store only partial descriptions
of the things to be remembered, descriptions that are sufficiently pre-
cise to work at the time something is learned, but that may not work
later on, when new experiences have also been encountered and en-
tered into memory. The descriptions formed to distinguish among the
old coins were not precise enough to distinguish between the new one
and at least one of the old ones.5

Suppose I keep all my notes in a small red notebook. If this is my
only notebook, I can describe it simply as my notebook. If I buy several
more notebooks, the earlier description will no longer work. Now I
must call the first one small or red, or maybe both small and red,
whichever allows me to distinguish it from the others. But what if I
acquire several small, red notebooks? Now I must find some other
means of describing the first book, adding to the richness of the de-
scription and thereby to its ability to discriminate among the several similar items. Descriptions need discriminate only among the choices in front of me, but what works for one purpose may not for another.6

THE POWER OF CONSTRAINTS

Back in the good old days of oral tradition (and even today for some cultures), performers traveled around reciting epic poems thousands of lines long. How did they do it? Do some people have huge amounts of knowledge in their heads? Not really. It turns out that external constraints exert powerful control over the permissible choice of words, thus dramatically reducing the memory load.

Consider the constraints of rhyming. If you wish to rhyme one word with another in English, there are usually ten to twenty alternatives. But if you must have a word with a particular meaning to rhyme with another, there are usually no candidates at all. And if there are any, in most cases there is only one. Combining the two constraints of rhyme and meaning can therefore reduce the information about the particular word that must be kept in memory to nothing; as long as the constraints are known, the choice of word can be completely determined. The learning of material like poetry is greatly aided by these kinds of constraints, which work on the general schema for the class of poem, meter, and topic.

Here is an example. I am thinking of three words: one means “a mythical being,” the second is “the name of a building material,” and the third is “a unit of time.” What words do I have in mind? Although you can probably think of three words that fit the descriptions, you are not likely to get the same three that I have in mind. There simply are not enough constraints.

Now try a second task, this time looking for rhyming words. I am thinking of three words: one rhymes with “post,” the second with “eel,” and the third with “ear.” What words am I thinking of?

Suppose I now tell you that the words I seek are the same in both tasks: What is a word that means a mythical being and rhymes with “post”? What word is the name of a building material and rhymes with “eel”? And what word is a unit of time and rhymes with “ear”? Now the task is easy: the joint specification of the words completely constrains the selection.

In the psychology laboratory, people almost never got the correct meanings or rhymes for the first two tasks, but they correctly answered “ghost,” “steel,” and “year” in the combined task almost always.7

The classic study of memory for epic poetry was done by Albert Bates Lord. He went to Yugoslavia and found people who still followed the oral tradition. He demonstrated that the “singer of tales,” the person who learns epic poems and goes from village to village reciting them, is really recreating them, composing poetry on the fly in such a way that it obeys the rhythm, theme, story line, structure, and other characteristics of the poem. This is a prodigious feat, but it is not an example of rote memory. Rather, the practice illustrates the immense power of the multiple constraints that allow the singer to listen to another singer tell a lengthy tale once, and then (after a delay of a few hours or a day) apparently recite “the same song, word for word, and line for line.”9 In fact, as Lord points out, the original and new recitations are not the same word for word. But the listener would perceive them as the same, even if the second version were twice as long as the first. They are the same in the ways that matter to the listener: they tell the same story, express the same ideas, and follow the same rhyme and meter. They are the same in all senses that matter to the culture. Lord shows just how the combination of memory for poetics, theme, and style combine with cultural structures into what he calls a formula for producing an appropriate poem, perceived as identical to earlier recitations. The notion that someone should be able to recite word for word is relatively modern. Such a notion can be held only after printed texts become available; otherwise who could judge the accuracy of a recitation? Perhaps more important, who would care? All this is not to detract from the feat. Learning and reciting an epic poem such as Homer’s Odyssey or Iliad is clearly difficult even if the singer is recreating it: there are 27,000 lines of verse in the written version.9

Most of us do not learn epic poems. But we do make use of strong constraints that serve to simplify what must be retained in memory. Consider an example from a completely different domain: taking apart and reassembling a mechanical device. Typical items in the home that an adventuresome person might attempt to repair include a door lock, toaster, and washing machine. The device is apt to have tens of parts. What has to be remembered in order to put the parts together again in proper order? Not as much as might appear from an initial analysis. In the extreme case, if there are ten parts, there are 10! (10 factorial: 10

0 The Design of Everyday Things

THREE: Knowledge in the Head and in the World 61
different ways in which to reassemble them—a little over 3.5 million alternatives. But never can all possible orderings be produced: there will be a number of physical constraints on the ordering. Some pieces must be assembled before it is even possible to assemble the others. Some pieces are physically constrained from fitting into the spots reserved for others: bolts must fit into holes of an appropriate diameter and depth; nuts and washers must be paired with bolts and screws of appropriate sizes; and washers must always be put on before nuts. There are even cultural constraints: we turn screws clockwise to tighten, counterclockwise to loosen; the heads of screws tend to go on the visible part (front or top) of a piece, bolts on the less visible part (bottom, side, or interior) of a piece; wood screws and machine screws look different and are inserted into different kinds of materials. In the end, the apparently large number of decisions is reduced to only a few choices that should have been learned or otherwise noted during the disassembly. The constraints by themselves are often not sufficient to determine the proper reassembly of the device—mistakes do get made—but the constraints reduce the amount that must be learned to a reasonable quantity.

Memory Is Knowledge in the Head

Remember the story of 'Ali Baba and the forty thieves? 'Ali Baba discovered the secret words that opened the thieves’ cave. His brother-in-law, Kasim, forced him to reveal the secret. Kasim then went to the cave.

"When he reached the entrance of the cavern, he pronounced the words, Open Simsim!"

"The door immediately opened, and when he was in, closed on him. In examining the cave he was greatly astonished to find much more riches than he had expected from 'Ali Baba's relation. He quickly lade at the door of the cavern as many bags of gold as his ten mules could carry, but his thoughts were now so full of the great riches he should possess, that he could not think of the necessary words to make the door open. Instead of Open Simsim! he said Open Barley! and was much amazed to find that the door remained shut. He named several sorts of grain, but still the door would not open.

"Kasim never expected such an incident, and was so alarmed at the danger he was in that the more he endeavoured to remember the word Simsim the more his memory was confounded, and he had as much forgotten it as if he had never heard it mentioned."

Kasim never got out. The thieves returned, cut off Kasim’s head, and quartered his body.10

THE CONSPIRACY AGAINST MEMORY

Most of us will not get our heads cut off if we fail to remember a secret code, but it can still be very hard to do. It is one thing to have to memorize one or two secrets: a combination, or a password, or the secret to opening the door. But when the number of secret codes gets too large, memory fails. There seems to be a conspiracy, one calculated to destroy our sanity by overloading our memory. Consider what we are asked to remember in our “convenient” world. A simple search through my own wallet and papers reveals the following things.

- Postal codes ranging in the United States from the “short form” of five digits to the “long form” of nine. Human short-term memory can comfortably retain only a five- to seven-digit number, yet here I am asked to use nine. I need to know the code for where I live, the code for where I work, the codes for my parents and for my children, the codes for my friends, and the codes for anyone with whom I correspond regularly. American codes, such as 92042-6207; British codes, such as WC1N 3BG; Canadian codes, such as M6P2V8. All for the sake of the machinery, and despite the fact that addresses are perfectly sensible and normally unambiguous. But machines have trouble with addresses, whereas they can deal with simple postal codes.

- Telephone numbers, sometimes with area codes and extensions. A seven-digit number becomes ten when the area code is added, and then fourteen when there is a four-digit extension. International codes, with country code and city code, add more digits. How many telephone numbers must I know? More than I wish to contemplate. All my personal contacts. Numbers for information, time, and weather; the special number for emergencies. And I mustn’t forget to dial 9 (or, in some cases, 8) so that the call will go outside the institution or company.
Access numbers for telephone budget cards, so that when I make a long distance call from my university, I can cause the correct account to pay the bill: a five-digit number for each account (and I have four of them). Don't show these to anyone, I am warned. Keep them hidden in a secret place.

Access numbers for telephone credit cards, so when I travel I can have the bill automatically put on my home telephone number. The codes consist of my home telephone number plus four secret digits. The secret digits aren't even printed on the card: memorize and destroy. But I have six of them (two home phone accounts and four different university phone accounts). If I want to dial a long distance number from a hotel using one of my telephone credit cards, I must dial as many as thirty-six digits.

Passwords or numbers for bank automatic teller machines, those clever machines that let you put in a card, type in your secret password, and get money. Two bank accounts, two secret passwords. Don't write them down, a thief might see them. Memorize.

Secret passwords for my computer accounts: can't let people steal my valuable data, or perhaps change their course grades, or peek at the examination questions. Make the password at least six characters long, we are told. And no words—words are too easy for someone to discover—make it nonsense. (I cheat and make all my computer accounts use the same password.)

Driver's license number. When I lived briefly in Texas I couldn't do anything without my driver's license number: not pay for food at the supermarket, not pay the telephone bill, not even open up a bank account. That was one letter, seven digits. Other states have longer numbers.

Social security numbers for me, my wife, and my children. Nine digits each.

Passport numbers, again for my whole family.

My employee number.

License plate numbers for our cars.

Birthdays.

Ages.

Clothing sizes.

Addresses.

Credit card numbers.

Bah and humbug.

So many of these numbers and codes must be kept secret. Apparently, thieves are everywhere, just waiting for me to write down my secret password or number, anxious to make that phone call on my account or to purchase items with my charge card. There is no way that I can learn all those numbers. And they keep changing, anyway, some of them annually. I even have trouble remembering how old I am: it changes every year too. (Quick: what magic phrase was Kasim trying to remember to open the cavern door?)

How can we remember all these things? Most of us can't, even with the use of mnemonics to make some sense of nonsensical material. Books and courses on improving memory can work, but the methods are laborious to learn and need continued practice to maintain. So we put the memory in the world, writing things down in books, on scraps of paper, even on the backs of our hands. But we disguise them to thwart would-be thieves. That creates another problem: How do we disguise the items, how do we hide them, and how do we remember what the disguise was or where we put them? Ah, the foibles of memory.

Where should you hide something so that nobody else will find it? In unlikely places, right? Money is hidden in the freezer, jewelry in the medicine cabinet or in shoes in the closet. The key to the front door is hidden under the mat or just below the window ledge. The car key is under the bumper. The love letters are in a flower vase. The problem is, there aren't that many unlikely places in the home. You may not remember where the love letters or keys are hidden, but your burglar will. Two psychologists who examined the issue described the problem this way:

"There is often a logic involved in the choice of unlikely places. For example, a friend of ours was required by her insurance company to acquire a safe if she wished to insure her valuable gems. Recognizing that she might forget the combination to the safe, she thought carefully about where to keep the combination. Her solution was to write it in her personal phone directory under the letter S next to 'Mr. and Mrs. Safe,' as if it were a telephone number. There is a clear logic here: Store numerical information with other numerical information. She was appalled, however, when she heard a reformed burglar on a daytime
television talk show say that upon encountering a safe, he always headed for the phone directory because many people keep the combination there."11

All these numbers to remember add up to unwitting tyranny. It is time for a revolt.

THE STRUCTURE OF MEMORY

"Say aloud the numbers 1, 7, 4, 2, 8. Next, without looking back, repeat them. Try again if you must, perhaps closing your eyes, the better to 'hear' the sound still echoing in mental activity. Have someone read a random sentence to you. What were the words? The memory of the just present is available immediately, clear and complete, without mental effort.

"What did you eat for dinner three days ago? Now the feeling is different. It takes time to recover the answer, which is neither as clear nor as complete a remembrance as that of the just present, and the recovery is likely to require considerable mental effort. Retrieval of the past differs from retrieval of the just present. More effort is required, less clarity results. Indeed, the 'past' need not be so long ago. Without looking back, what were those digits? For some people, this retrieval now takes time and effort."12

Psychologists distinguish between two major classes of memory: short-term memory and long-term memory (abbreviated STM and LTM, respectively). The two are quite different. Short-term memory is the memory of the just present. Information is retained in it automatically and retrieved without effort; but the amount of information that can be retained this way is severely limited. Something like five to seven items is the limit of STM, with the number going to ten or twelve if a person also rehearse, mentally repeating the items to be retained. Short-term memory is invaluable in the performance of everyday tasks, in letting us remember words, names, phrases, and parts of tasks. It acts as a working or temporary memory. But the memory is quite fragile. Get distracted by some other activity and, poof, the stuff in STM disappears. It is capable of holding a five-digit postal code or seven-digit telephone number from the time you look them up until the time they are used—as long as no distractions occur. Nine- or ten-digit numbers give trouble, and when the number starts to exceed that—don’t bother. Write it down. Or divide the number into several shorter segments.

Long-term memory is memory for the past. As a rule, it takes time to put stuff away in LTM and time and effort to get to get it out again. This is how we maintain our experiences, not as an exact recording of the events, but as interpreted through our understanding of them, subject to all the distortions and changes that the human explanatory mechanism imposes upon life. How well we can ever recover experiences and knowledge from LTM is highly dependent upon how the material was interpreted in the first place. What is stored in LTM under one interpretation probably cannot be found later on when sought under some other interpretation. As for how large the memory is, nobody really knows: billions of items, probably. One informed scientist estimates the capacity as a billion (10⁹) bits or about 100 million (10⁸) items.13 Whatever the size, it is so large as not to impose any practical limit. The difficulty with LTM is in organization—in getting material in and in figuring out how to retrieve it—not in capacity. Storage and retrieval are easier when the material makes sense, when it fits into what is already known. When the material makes no sense, it will have to be worked on, structured, and interpreted, until finally it can be retained.

Human memory is essentially knowledge in the head, or internal knowledge. If we examine how people use their memories and how they retrieve information, we discover a number of categories. Three are important for us now:

1. Memory for arbitrary things. The items to be retained seem arbitrary, with no meaning and no particular relationship to one other or to things already known.
2. Memory for meaningful relationships. The items to be retained form meaningful relationships with themselves or with other things already known.
3. Memory through explanation. The material does not have to be remembered, but rather can be derived from some explanatory mechanism.

MEMORY FOR ARBITRARY THINGS

Arbitrary knowledge can be classified as the simple remembering of what is to be done, without reliance on an understanding of why or on internal structure. This is how we learned the alphabet and how to tie
a shoelace. It is even how we learned the multiplication tables, that 3 times 2 is 6, although for that we could refer to an external structure. This is how we are expected to learn arbitrary codes to operate the modern, misbegotten telephone system. It is also how we are forced to learn many procedures required of modern technology: "To load this program, put the floppy diskette into drive A and type ALT MODE, CONTROL-SHIFT-X, DELETE." This is rote learning, the bane of modern existence.

Rote learning creates problems. First, because what is being learned is arbitrary, the learning is difficult; it can take considerable time and effort. Second, when a problem arises, the memorized sequence of actions gives no hint of what has gone wrong, no suggestion of what might be done to fix the problem. Although some things are appropriate to learn by rote (the letters of the alphabet, for example), most are not. Alas, it is still the dominant method of instruction in many school systems, and even for much adult training. This is how some people are taught to use computers, or to cook. It is how we have to learn to use some of the new (poorly designed) gadgets of our technology.

Most psychologists would argue that it is not really possible to learn arbitrary associations or sequences. Even where there appears to be no structure, people manufacture some artificial and usually rather unsatisfactory one, which is why the learning is so bad. For our purposes it does not matter whether arbitrary learning is impossible or simply very difficult, the end result is the same: it is not the best way to go, not if there is any choice in the matter. Thus, in teaching the alphabet, we try to make it into a tune, using the natural constraints of rhyme and rhythm to simplify the memory load. People who have learned to use computers or cook by rote are probably not very good. Since they do not understand the reasons for their actions, they must find tasks arbitrary and strange. When something goes wrong, they don't know what to do (unless they've memorized solutions). Although rote learning is at times necessary or efficient—so that emergency procedures for things like high-speed military jet aircraft are handled quickly, automatically when the need arises—on the whole, it is most unsatisfactory.

MEMORY FOR MEANINGFUL RELATIONSHIPS

Most things in the world have a sensible structure, which tremendously simplifies the memory task. When things make sense, they correspond to knowledge that we already have, so the new material can be understood, interpreted, and integrated with previously acquired material. Now we can use rules and constraints to help understand what things go together. Meaningful structure can organize apparent chaos and arbitrariness.

Remember the discussion of mental models in chapter 2? Part of the power of a good mental model lies in its ability to provide meaning to things. Let's look at an example to show how a meaningful interpretation transforms an apparently arbitrary task into a natural one. Note that the appropriate interpretation may not at first be obvious; it, too, is knowledge and has to be discovered.

A Japanese colleague, call him Mr. Tanaka, had difficulty remembering how to use the turn-signal switch on his motorcycle's left handlebar. Moving the switch forward signaled a right turn, backward a left turn. The meaning of the switch was clear and unambiguous, but the direction in which it should be moved was not. Tanaka kept thinking that because the switch was on the left handlebar, pushing it forward should signal a left turn. That is, he was trying to map the action "push the left switch forward" to the intention "turn left," which was wrong. As a result, he had trouble remembering which switch direction should be used for which turning direction. Most motorcycles have the turn-signal switch mounted differently, rotated 90º, so that moving it left signals a left turn, moving it right a right turn. This mapping is easy to learn (it is a natural mapping). But the turn switch on Tanaka's motorcycle moved forward and back, not left and right. How could he learn it?

Mr. Tanaka solved the problem by reinterpreting the action. Consider the way the handlebars of the motorcycle turn. For a left turn, the left handlebar moves backward. For a right turn, the left handlebar moves forward. The required switch movements exactly paralleled the handlebar movements. If the task is reconceptualized as signaling the direction of motion of the handlebars rather than the direction of the motorcycle, the switch motion can be seen to mimic the desired motion; finally we have a natural mapping. At first, the motion of the switch seemed arbitrary, indirect, and difficult to remember. With the proper interpretation, the switch motion is direct and logical, and, as a result, easy to learn and to use. A meaningful relationship can be indispensable, but you have to have the right one.14

Without the proper interpretation, it was difficult to remember the switch directions. With it, both the remembering and the performance
of the task became trivial. Note that Tanaka's interpretation of the switch movement did not explain anything. It simply let him relate the proper direction to move the switch with the direction in which he was turning the motorcycle. The interpretation is essential, but it should not be confused with understanding.

MEMORY THROUGH EXPLANATION

Now we come to a different, more powerful form of internal memory: understanding. People are explanatory creatures, as I showed in chapter 2. Explanations and interpretations of events are fundamental to human performance, both in understanding the world and in learning and remembering. Here mental models play a major role. Mental models simplify learning, in part because the details of the required behavior can be derived when needed. They can be invaluable in dealing with unexpected situations. Note that the use of mental models to remember (in this case, derive) behavior is not ideal for tasks that must be done rapidly and smoothly. The derivation takes time and requires mental resources, neither of which may be in great supply during critical incidents. Mental models let people derive appropriate behavior for situations that are not remembered (or never before encountered). People probably make up mental models for most of the things they do. This is why designers should provide users with appropriate models: when they are not supplied, people are likely to make up inappropriate ones.¹⁵

The sewing machine provides a good example of the power of a mental model. A sewing machine is a mysterious beast, managing to loop an upper thread through a lower thread, even though each thread is always connected to its spool or bobbin, respectively. The mental model has to explain how the upper thread goes through the material being sewn, dips under the surface plate, and then loops around the lower thread.

The proper model, it turns out, is something like this. Picture the lower bobbin held gently in the machine by a kind of cup with sloping sides. The cup keeps the bobbin stable, allowing it to rotate so its thread can be unwound. Yet the cup is loose enough so that the upper thread can go inside the cup and around the bobbin—and therefore around the bottom thread. When the upper needle goes through the material and under the plate, a rotating hook grabs its thread and guides it between the inner walls of the cup and the outer walls of the bobbin case. This helps explain why the machine won't work properly if the bobbin is bent, even if the bobbin still appears to fit and the bottom thread unrolls properly. It explains why dirt on the bobbin or in the cup will mess things up, and why certain kinds of upper thread might cause more trouble than others. (A thick upper thread, especially one that was rough or sticky, might not go smoothly around the bobbin.)

To be honest, I don't know if anything I just said about the failures of bobbins is true. I derived each example from my mental model of a sewing machine. I can't sew. But when Naomi Miyake did her research for her doctoral thesis in my laboratory, she studied people's understanding of sewing and of the machines. The result was twofold: a fine piece of research for her and a mental model for me. So now I can derive what would happen, even if it has never happened to me.

The power of mental models is that they let you figure out what would happen in novel situations. Or, if you are actually doing the task and there is a problem, they let you figure out what is happening. If the model is wrong, you will be wrong too. Am I right about the sewing machine? Decide for yourself: go look at one.

After word got out that I was collecting instances of design peculiarities, a friend reported the following about the sunroof of his new car, an Audi. Supposedly, if the ignition is not on, the sunroof cannot be operated. However, a mechanic explained that you could close the sunroof even without the ignition key if you turned on the headlights and then (1) pulled back on the turn-signal stalk (which normally switches the headlights to high beam), and (2) pushed the close control for the sunroof.

My friend said that it was thoughtful of Audi to provide this override of the ignition key in case the sunroof was open when it started raining. You could close it even if you didn't have your key. But we both wondered why the sequence was so peculiar.

Ever skeptical, I asked to see the manual for the car. The manual was explicit: "You cannot work the sunroof if the ignition is off." A similar statement appeared in the discussion of the electrically powered windows. My friend's mental model was functional: it explained why you would want such a feature, but not how it worked. If the feature was so desirable, why was it not mentioned in the manual?
We searched for another explanation. Perhaps it wasn't a design feature, after all. Perhaps it was an accident of design. Perhaps turning on the lights and pulling back on the stalk connected the electrical power to the car, overriding the fact that the ignition key was off. This would allow the sunroof to work, but only as a by-product of the way the lights were wired.

This model was more specific. It explained what was happening and allowed us to predict that all electrical items should work. So we checked. Turning on the light switch without engaging the ignition did not turn on the headlights; only the parking lights went on. But when we also pulled back on the turn-signal stalk, the headlights did turn on, even though the ignition was off. With the stalk pulled back, the sunroof would close and open. The windows would close and open. The fan on the heating system worked. So did the radio. This was an effective mental model. Now we could understand better what was happening, predict new results, and more easily remember the peculiar set of operations required for the task.

Memory Is Also Knowledge in the World

As we have seen, knowledge in the world, external knowledge, can be very valuable. But it, too, has drawbacks. For one, it is available only if you are there, in the appropriate situation. When you are somewhere else, or if the world has changed meanwhile, the knowledge is gone. The critical memory aids provided by the external information are absent, and so the task or item may not be remembered. A folk saying captures this situation well: "Out of sight, out of mind."

REMINING

One of the most important and interesting aspects of the role of external memory is reminding, a good example of the interplay between knowledge in the head and in the world. Suppose a neighboring family asks you to take them to the airport. You agree to take them next Saturday at 3:30 P.M. Now the knowledge is in your head, but how are you going to remember it at the proper time? You will need to be reminded. There are many strategies for reminding. One is simply to keep the information in your head. If the event is important enough, you count on having it come repeatedly to mind—what psychologists call rehearsal—so that you can simply assume that there will be no difficulty at all remembering when to leave on Saturday. You can keep the information in your head especially when the event is of great personal importance; suppose you are catching the plane for your first trip to Paris. You won't have any problem remembering. But keeping the knowledge in your head is not ordinarily a good reminding technique.

Suppose the event is not personally important, it is several days away, and you are leading a very busy life. Now you'd better transfer some of the burden of remembering to the outside world. Here is where you use notes to yourself, or pocket and desk calendars or diaries, or electronic alarm clocks that can be set for time of day and date. Or you can ask a friend to remind you. Those of us with secretaries put the burden on them. They, in turn, write notes, enter events on calendars, or set an alarm on the computer system (if it is well enough designed that they can figure out how to work it).

A good reminding method is to put the burden on the thing itself. Do my neighbors want me to take them to the airport? Fine, but they have to call me up the night before and remind me. Do I want to remember to take a book to the university to give to a colleague? I put the book someplace where I cannot fail to see it when I leave the house. A good spot is against the front door of the house. I can't leave without tripping over the book. If I am at a friend's house and I borrow a paper or a book, I remember to take it by putting my car keys on it. Then when I leave, I am reminded. Even if I forget and go out to my car; I can't drive away without the keys.

There are two different aspects to a reminder: the signal and the message. Just as in doing an action we can distinguish between knowing what can be done and knowing how to do it, in reminding we must distinguish between knowing that something is to be remembered and remembering what it is. Most popular reminding devices provide only one of these two critical aspects. The famous "tie a string around your finger" reminder provides only the signal. It gives no hint of what is to be remembered. Writing a note to yourself provides only the message; it doesn't remind you ever to look at it. (Tying a knot in your handkerchief—Carrman's device in figure 3.2—provides neither signal nor message.) The ideal reminder has to have both components: the signal that something is to be remembered, the message of what it is.
3.2 Carelman's Preknotted Handkerchief. What an aid to the forgetful—except that the act of tying the knot is probably just as useful a memory cue as the knot itself. (Jacques Carelman: “Preknotted Handkerchief” Copyright © 1969–76-80 Jacques Carelman and A. D. A. G. F. Paris. From Jacques Carelman, Catalog of Unfindable Objects. Balland, éditeur, Paris-France. Used by permission of the artist.)

The need for timely reminders has created loads of products that make it easier to put the knowledge in the world—alarm clocks, diaries, calendars. A variety of sophisticated watches and small, calculator-sized reminding devices are starting to appear. So far they are limited in power and difficult to use. But I believe there is a need for them. They just need more work, better technology, and better design.

Would you like a pocket-size device that reminded you of each appointment and daily event? I would. I am waiting for the day when portable computers become small enough that I can keep one with me at all times. I will definitely put all my reminding burdens upon it. It has to be small. It has to be convenient to use. And it has to be relatively powerful, at least by today's standards. It has to have a full, standard typewriter keyboard and a reasonably large display. It needs good graphics, because that makes a tremendous difference in usability, and a lot of memory—a huge amount, actually. And it should be easy to hook up to the telephone; I need to connect it to my home and laboratory computers. Of course, it should be relatively inexpensive.

What I ask for is not unreasonable. The technology I need is available today. It’s just that the full package has never been put together, partly because the cost in today’s world would be prohibitive. But it will exist in imperfect form in five years, possibly in perfect form in ten.

NATURAL MAPPINGS

The arrangement of burners and controls on the kitchen stove provides a good example of the power of natural mappings to reduce the need for information in memory. Without a good mapping, the user cannot readily determine which burner goes with which control. Consider the standard stove with four burners, arranged in the traditional rectangle. If the four controls were truly arbitrary, as in figure 3.3, the user would have to learn each control separately: twenty-four possible arrangements. Why twenty-four? Start with the leftmost control: it could work any of the four burners. That leaves three possibilities for the next leftmost. So there are 12 \((4 \times 3)\) possible arrangements of the first two controls: four for the first, three for the second. The third control could work either of the two remaining burners, and then there is only one burner left for the last control. This makes twenty-four possible mappings between the controls and burners: \(4 \times 3 \times 2 \times 1 = 24\).

With the completely arbitrary arrangement, the stove is unworkable unless each control is fully labeled to indicate which burner it controls.

Most stoves have controls arranged in a line, even though the burners are arranged rectangularly. Controls are not mapped naturally to burners. As a result, you have to learn which control goes with which burner. Consider how the use of spatial analogies can relieve the memory burden. Start with a partial mapping that is in common use today: the controls are segregated into left and right halves, as in figure 3.4. Now we need know only which left burner each of the two left controls affects and which right burner each right control affects—two alternatives for each of the four burners. The number of possible arrangements is now only four—two possibilities for each side: quite a reduction from the twenty-four. But the controls must still be labeled, which indicates that the mapping is still imperfect. Since some of the information is now in the spatial arrangement, each control need only be labeled back or front; the left and right labels are no longer needed.

What about a proper, full, natural mapping, with the controls spatially arranged in the same pattern as the burners, as in figure 3.5? The organization of the controls now carries all the information required. We know immediately which control goes with which burner. Such is the power of natural mapping. We can see that the number of possible sequences has been reduced from twenty-four to one.16 If all possible
3.3 Arbitrary Arrangement of Stove Controls (top of opposite page). Couple the usual rectangular arrangement of burners with this arbitrary row of controls, and there is trouble: which control goes with which burner? You don’t know unless the controls are labeled. The memory load for this arrangement is high: there are twenty-four possible arrangements, and you have to remember which of the twenty-four this one is. Fortunately, the controls are seldom arranged quite this arbitrarily.

3.4 Paired Stove Controls (bottom of opposite page). This is the type of partial mapping of controls to burners in common use today. The two controls on the left work the left burners, and the two controls on the right work the right burners. Now there are only four possible arrangements (two for each side). Even so, confusion is possible (and, I can assure you, it occurs often).

3.5 Full Natural Mapping of Controls and Burners (below). Two of the Possible Ways. There is no ambiguity, no need for learning or remembering, no need for labels. Why can’t all stoves be like these?
natural mappings were applied in our lives, the cumulative effect would be enormous.

The problem of the stove top may seem trivial, but in fact it is a cause of great frustration for many homeowners. Why do stove designers insist on arranging the burners in a rectangular pattern and the controls in a row? We have known for forty years just how bad such an arrangement is. Sometimes the stove comes with clever little diagrams to indicate which control works which burner. Sometimes there is a short label. But the proper natural mapping requires no diagrams, no labels, and no instructions. There is a simple design principle lurking here:

If a design depends upon labels, it may be faulty. Labels are important and often necessary, but the appropriate use of natural mappings can minimize the need for them. Wherever labels seem necessary, consider another design.

The shame about stove design is that it isn't hard to do right. Textbooks of ergonomics, human factors, psychology, and industrial engineering all show various sensible solutions. And some stove manufacturers do use good designs. Oddly, some of the very best and the very worst are manufactured by the same companies and are illustrated side by side in the same catalogs.

Why do designers insist on frustrating users? Why do users still purchase stoves that cause so much trouble? Why not revolt and refuse to buy them unless the controls have an intelligent relationship to the burners? I bought a bad one myself.

Usability is not often thought of as a criterion during the purchasing process. Moreover, unless you actually test a number of units in a realistic environment doing typical tasks, you are not likely to notice the ease or difficulty of use. If you just look at something, it appears straightforward enough, and the array of wonderful features seems to be a virtue. You may not realize that you won't be able to figure out how to use those features. I urge you to test products before you buy them. Pretending to cook a meal, or setting the channels on a video set, or attempting to program a VCR will do. Do it right there in the store. Do not be afraid to make mistakes or ask stupid questions. Remember, any problems you have are probably the design's fault, not yours.

A major problem is that often the purchaser is not the user. Appliances may be in a home when people move in. In the office, the purchasing department orders equipment based upon such factors as price, personal relationships with the supplier, and perhaps reliability: usability is seldom considered. Finally, even when the purchaser is the end user, it is sometimes necessary to trade one desirable feature for an undesirable one. In the case of my family's stove, we did not like the arrangement of controls, but we bought the stove anyway: we traded off layout of the burner controls for another feature that was more important to us and available only from one manufacturer. (I return to these issues in chapter 6.)

The Tradeoff between Knowledge in the World and in the Head

Knowledge (or information) in the world and in the head are both essential in our daily functioning. But to some extent we can choose to lean more heavily on one or the other. That choice requires a trade-

<table>
<thead>
<tr>
<th>PROPERTY</th>
<th>KNOWLEDGE IN THE WORLD</th>
<th>KNOWLEDGE IN THE HEAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrievalability</td>
<td>Retrievable whenever visible or audible.</td>
<td>Not readily retrievable. Requires memory search or reminding.</td>
</tr>
<tr>
<td>Learning</td>
<td>Learning not required. Interpretation substitutes for learning. How easy it is to interpret information in the world depends upon how well it exploits natural mappings and constraints.</td>
<td>Requires learning, which can be considerable. Learning is made easier if there is meaning of structure to the material (or if there is a good mental model).</td>
</tr>
<tr>
<td>Efficiency of use</td>
<td>Tends to be slowed up by the need to find and interpret the external information.</td>
<td>Can be very efficient.</td>
</tr>
<tr>
<td>Ease of use at first encounter</td>
<td>High.</td>
<td>Low.</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>Can be unaesthetic and inelegant, especially if there is a need to maintain a lot of information. This can lead to clutter. In the end, aesthetic appeal depends upon the skill of the designer.</td>
<td>Nothing need be visible, which gives more freedom to the designer, which in turn can lead to better aesthetics.</td>
</tr>
</tbody>
</table>
off—gaining the advantages of knowledge in the world means losing
the advantages of knowledge in the head (figure 3.6).

Knowledge in the world acts as its own reminder. It can help us
recover structures that we otherwise would forget. Knowledge in the
head is efficient: no search and interpretation of the environment is
required. In order to use knowledge in the head we have to get it there,
which might require considerable amounts of learning. Knowledge in
the world is easier to learn, but often more difficult to use. And it relies
heavily upon the continued physical presence of the information;
change the environment and the information is changed. Performance
relies upon the physical presence of the task environment.

Reminders provide a good example of the relative tradeoffs between
the roles of internal versus external knowledge. Knowledge in the
world is accessible. It is self-reminding. It is always there, waiting to
be seen, waiting to be used. That is why we structure our offices and
our places of work so carefully. We put piles of papers where they can
be seen, or if we like a clean desk, we put them in standardized loca-
tions and teach ourselves (knowledge in the head) to look in these
standard places routinely. We use clocks and calendars and notes.
Knowledge in the mind is ephemeral: here now, gone later. We can’t
count on something being present in mind at any particular time, unless
it is triggered by some external event or unless we deliberately keep it
in mind through constant repetition (which then prevents us from
having other conscious thoughts). Out of sight, out of mind.\footnote{17}

\textbf{KNOWING WHAT TO DO}

\begin{quote}
"Q. I read a news item about a new videotape-
only player and rejoiced when the writer took a
healthy swipe at the incomprehensible instruc-
tions that accompany VCRs. I can’t even set the
time of day on mine!"

"There are many consumers out here like me—thwarted by an un-
fathomable machine and baffled by senseless instruc-
tions."

"Is there anyone, anywhere who will translate OR give a short
course in VCR at play school level?"
\end{quote}

Video cassette recorders—VCRs—can be frightening to people who
are unfamiliar with them. Indeed, the number of options, buttons,
controls, displays, and possible courses of action is formidable. But at
least when we have trouble operating a VCR we have something to
blame: the machine’s bewildering appearance and the lack of clues to
suggest what can be done and how to do it. Even more frustrating,
however, is that we often have trouble working devices that we expect
to be simple.

The difficulty of dealing with novel situations is directly related to
the number of possibilities. The user looks at the situation and tries to
discover which parts can be operated and what operations can be done.