THE DESIGN OF EVERYDAY THINGS

Donald A. Norman

CURRENCY
DOUBLEDAY

New York London Toronto Sydney Auckland

A CURRENCY BOOK
PUBLISHED BY DOUBLEDAY
a division of Bantam Doubleday Dell Publishing Group, Inc.
1540 Broadway, New York, New York 10036

This book was originally published in hardcover by Basic Books in 1988. This Doubleday/Currency edition reprinted by special arrangement with Basic Books Inc., Publishers, New York.

Library of Congress Cataloging-in-Publication Data
Norman, Donald A.
[Psychology of everyday things]
The design of everyday things / Donald A. Norman.
p. cm.
Reprint. Originally published: The psychology of everyday things. New York : Basic Books, c1988.
Includes bibliographical references.
1. Design, Industrial—Psychological aspects. 2. Human engineering. I. Title.
[TS171.4.N67 1990]
620.8'2—dc20 89-48989
CIP

ISBN 0-385-26774-6

Printed in the United States of America
First Doubleday/Currency Edition: 1990

20 19 18 17

CONTENTS

of visibility, appropriate clues, and feedback of one's actions. These principles constitute a form of psychology—the psychology of how people interact with things. A British designer once noted that the kinds of materials used in the construction of passenger shelters affected the way vandals responded. He suggested that there might be a psychology of materials.

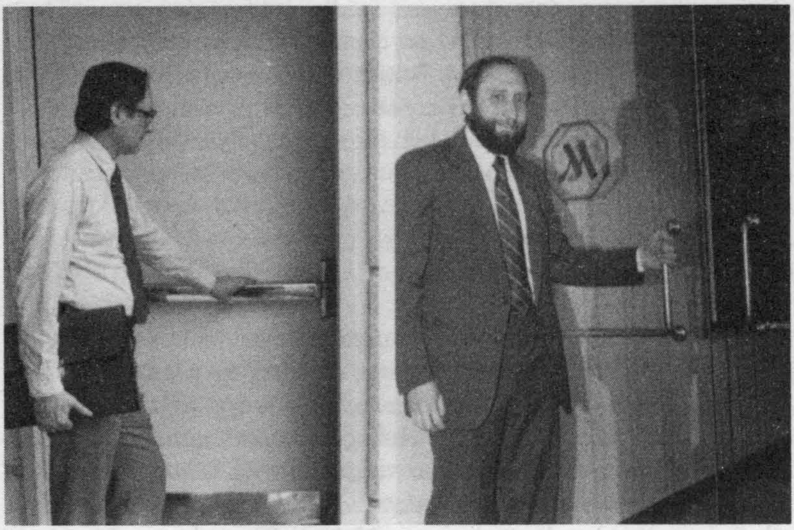
AFFORDANCES

"In one case, the reinforced glass used to panel shelters (for railroad passengers) erected by British Rail was smashed by vandals as fast as it was renewed. When the reinforced glass was replaced by plywood boarding, however, little further damage occurred, although no extra force would have been required to produce it. Thus British Rail managed to elevate the desire for defacement to those who could write, albeit in somewhat limited terms. Nobody has, as yet, considered whether there is a kind of psychology of materials. But on the evidence, there could well be!"[2]

There already exists the start of a psychology of materials and of things, the study of affordances of objects. When used in this sense, the term *affordance* refers to the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used (see figures 1.5 and 1.6). A chair affords ("is for") support and, therefore, affords sitting. A chair can also be carried. Glass is for seeing through, and for breaking. Wood is normally used for solidity, opacity, support, or carving. Flat, porous, smooth surfaces are for writing on. So wood is also for writing on. Hence the problem for British Rail: when the shelters had glass, vandals smashed it; when they had plywood, vandals wrote on and carved it. The planners were trapped by the affordances of their materials.[3]

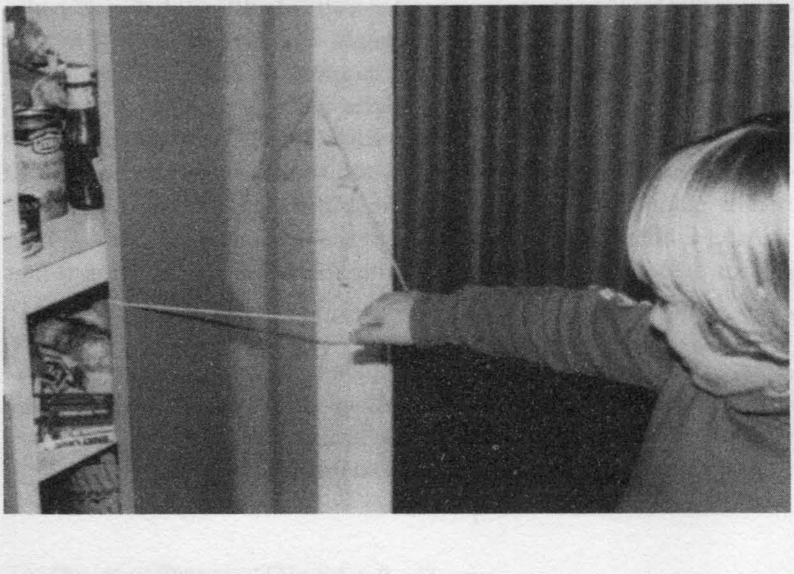
Affordances provide strong clues to the operations of things. Plates are for pushing. Knobs are for turning. Slots are for inserting things into. Balls are for throwing or bouncing. When affordances are taken advantage of, the user knows what to do just by looking: no picture, label, or instruction is required. Complex things may require explanation, but simple things should not. When simple things need pictures, labels, or instructions, the design has failed.

A psychology of causality is also at work as we use everyday things.



1.5 Affordances of Doors. Door hardware can signal whether to push or pull without signs. The flat horizontal bar of *A* (above left) affords no operations except pushing: it is excellent hardware for a door that must be pushed to be opened. The door in *B* (above right) has a different kind of bar on each side, one relatively small and vertical to signify a pull, the other relatively large and horizontal to signify a push. Both bars support the affordance of grasping: size and position specify whether the grasp is used to push or pull—though ambiguously.

1.6 When Affordances Fail. I had to tie a string around my cabinet door to afford pulling.



Something that happens right after an action appears to be caused by that action. Touch a computer terminal just when it fails, and you are apt to believe that you caused the failure, even though the failure and your action were related only by coincidence. Such false causality is the basis for much superstition. Many of the peculiar behaviors of people using computer systems or complex household appliances result from such false coincidences. When an action has no apparent result, you may conclude that the action was ineffective. So you repeat it. In earlier days, when computer word processors did not always show the results of their operations, people would sometimes attempt to change their manuscript, but the lack of visible effect from each action would make them think that their commands had not been executed, so they would repeat the commands, sometimes over and over, to their later astonishment and regret. It is a poor design that allows either kind of false causality to occur.

TWENTY THOUSAND EVERYDAY THINGS

There are an amazing number of everyday things, perhaps twenty thousand of them. Are there really that many? Start by looking about you. There are light fixtures, bulbs, and sockets; wall plates and screws; clocks, watches, and watchbands. There are writing devices (I count twelve in front of me, each different in function, color, or style). There are clothes, with different functions, openings, and flaps. Notice the variety of materials and pieces. Notice the variety of fasteners—buttons, zippers, snaps, laces. Look at all the furniture and food utensils: all those details, each serving some function for manufacturability, usage, or appearance. Consider the work area: paper clips, scissors, pads of paper, magazines, books, bookmarks. In the room I'm working in, I counted more than a hundred specialized objects before I tired. Each is simple, but each requires its own method of operation, each has to be learned, each does its own specialized task, and each has to be designed separately. Furthermore, many of the objects are made of many parts. A desk stapler has sixteen parts, a household iron fifteen, the simple bathtub-shower combination twenty-three. You can't believe these simple objects have so many parts? Here are the eleven basic parts to a sink: drain, flange (around the drain), pop-up stopper, basin, soap dish, overflow vent, spout, lift rod, fittings, hot-water handle, and cold-water handle. We can count even more if we start taking the faucets, fittings, and lift rods apart.

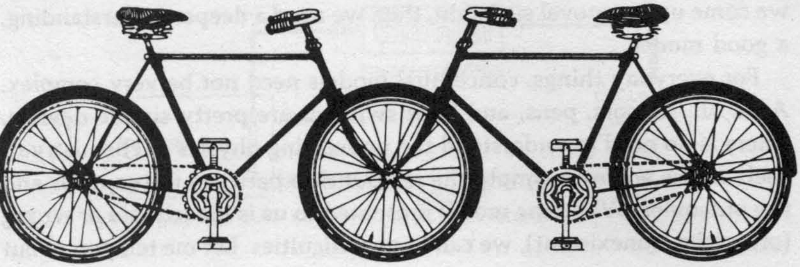
The book *What's What: A Visual Glossary of the Physical World* has more than fifteen hundred drawings and pictures and illustrates twenty-three thousand items or parts of items.[4] Irving Biederman, a psychologist who studies visual perception, estimates that there are probably "30,000 readily discriminable objects for the adult."[5] Whatever the exact number, it is clear that the difficulties of everyday life are amplified by the sheer profusion of items. Suppose that each everyday thing takes only one minute to learn; learning 20,000 of them occupies 20,000 minutes—333 hours or about 8 forty-hour work weeks. Furthermore, we often encounter new objects unexpectedly, when we are really concerned with something else. We are confused and distracted, and what ought to be a simple, effortless, everyday thing interferes with the important task of the moment.

How do people cope? Part of the answer lies in the way the mind works—in the psychology of human thought and cognition. Part lies in the information available from the appearance of the objects—the psychology of everyday things. And part comes from the ability of the designer to make the operation clear, to project a good image of the operation, and to take advantage of other things people might be expected to know. Here is where the designer's knowledge of the psychology of people coupled with knowledge of how things work becomes crucial.

CONCEPTUAL MODELS

Consider the rather strange bicycle illustrated in figure 1.7. You know it won't work because you form a *conceptual model* of the device and mentally simulate its operation. You can do the simulation because the parts are visible and the implications clear.

Other clues to how things work come from their visible structure—in particular from *affordances, constraints,* and *mappings.* Consider a pair of scissors: even if you have never seen or used them before, you can see that the number of possible actions is limited. The holes are clearly there to put something into, and the only logical things that will fit are fingers. The holes are affordances: they allow the the fingers to be inserted. The sizes of the holes provide *constraints* to limit the possible fingers: the big hole suggests several fingers, the small hole only one. The mapping between holes and fingers—the set of possible operations—is suggested and constrained by the holes. Moreover, the operation is not sensitive to finger placement: if you use the wrong fingers,



1.7 Carelman's Tandem "Convergent Bicycle (Model for Fiancés)." Jacques Carelman: "Convergent Bicycle" Copyright © 1969–76–80 by Jacques Carelman and A. D. A. G. P. Paris. From Jacques Carelman, *Catalog of Unfindable Objects,* Balland, éditeur, Paris-France. Used by permission of the artist.

the scissors still work. You can figure out the scissors because their operating parts are visible and the implications clear. The conceptual model is made obvious, and there is effective use of affordances and constraints.

As a counterexample, consider the digital watch, one with two to four push buttons on the front or side. What are those push buttons for? How would you set the time? There is no way to tell—no evident relationship between the operating controls and the functions, no constraints, no apparent mappings. With the scissors, moving the handle makes the blades move. The watch and the Leitz slide projector provide no visible relationship between the buttons and the possible actions, no discernible relationship between the actions and the end result.

eg

Principles of Design for Understandability and Usability

We have now encountered the fundamental principles of designing for people: (1) provide a good conceptual model and (2) make things visible.

PROVIDE A GOOD CONCEPTUAL MODEL

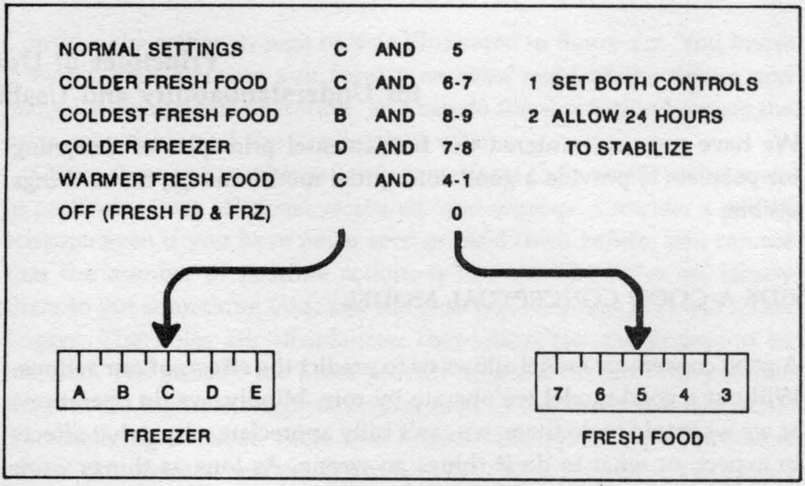
A good conceptual model allows us to predict the effects of our actions. Without a good model we operate by rote, blindly; we do operations as we were told to do them; we can't fully appreciate why, what effects to expect, or what to do if things go wrong. As long as things work properly, we can manage. When things go wrong, however, or when we come upon a novel situation, then we need a deeper understanding, a good model.

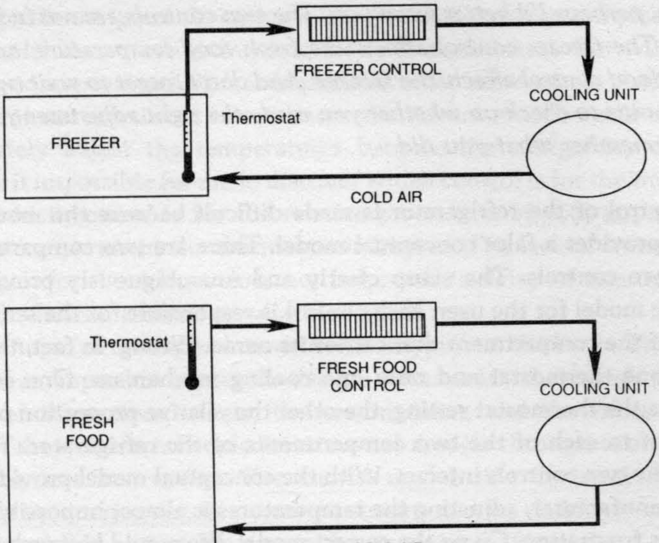
For everyday things, conceptual models need not be very complex. After all, scissors, pens, and light switches are pretty simple devices. There is no need to understand the underlying physics or chemistry of each device we own, simply the relationship between the controls and the outcomes. When the model presented to us is inadequate or wrong (or, worse, nonexistent), we can have difficulties. Let me tell you about my refrigerator.

My house has an ordinary, two-compartment refrigerator—nothing very fancy about it. The problem is that I can't set the temperature properly. There are only two things to do: adjust the temperature of the freezer compartment and adjust the temperature of the fresh food compartment. And there are two controls, one labeled "freezer," the other "fresh food." What's the problem?

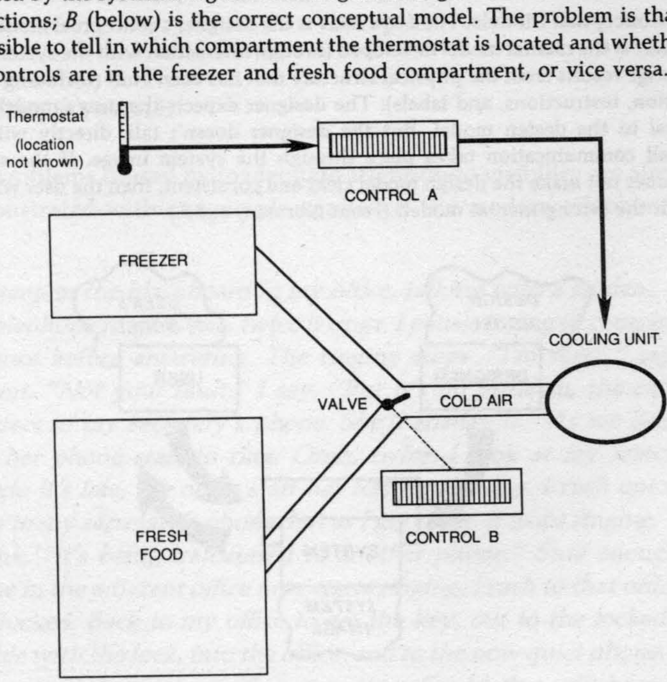
You try it. Figure 1.8 shows the instruction plate from inside the refrigerator. Now, suppose the freezer is too cold, the fresh food section just right. You want to make the freezer warmer, keeping the fresh food constant. Go on, read the instructions, figure them out.

1.8 My Refrigerator. Two compartments—fresh food and freezer—and two controls (in the fresh food unit). The illustration shows the controls and instructions. Your task: Suppose the freezer is too cold, the fresh food section just right. How would you adjust the controls so as to make the freezer warmer and keep the fresh food the same? (From Norman, 1986.)





1.9 Two Conceptual Models for My Refrigerator. The model *A* (above) is provided by the system image of the refrigerator as gleaned from the controls and instructions; *B* (below) is the correct conceptual model. The problem is that it is impossible to tell in which compartment the thermostat is located and whether the two controls are in the freezer and fresh food compartment, or vice versa.

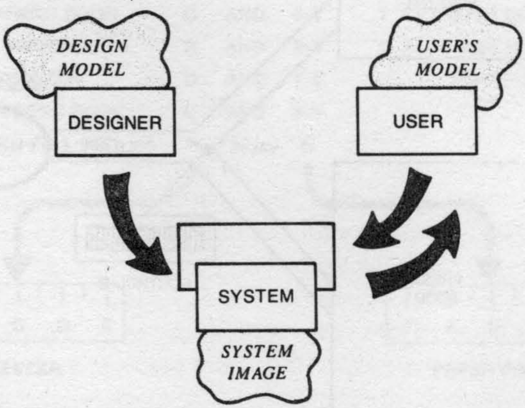


Oh, perhaps I'd better warn you. The two controls are not independent. The freezer control affects the fresh food temperature, and the fresh food control affects the freezer. And don't forget to wait twenty-four hours to check on whether you made the right adjustment, if you can remember what you did.

Control of the refrigerator is made difficult because the manufacturer provides a false conceptual model. There are two compartments and two controls. The setup clearly and unambiguously provides a simple model for the user: each control is responsible for the temperature of the compartment that carries its name. Wrong. In fact, there is only one thermostat and only one cooling mechanism. One control adjusts the thermostat setting, the other the relative proportion of cold air sent to each of the two compartments of the refrigerator. This is why the two controls interact. With the conceptual model provided by the manufacturer, adjusting the temperatures is almost impossible and always frustrating. Given the correct model, life would be much easier (figure 1.9).

Why did the manufacturer present the wrong conceptual model?

1.10 Conceptual Models. The *design model* is the designer's conceptual model. The *user's model* is the mental model developed through interaction with the system. The *system image* results from the physical structure that has been built (including documentation, instructions, and labels). The designer expects the user's model to be identical to the design model. But the designer doesn't talk directly with the user—all communication takes place through the system image. If the system image does not make the design model clear and consistent, then the user will end up with the wrong mental model. (From Norman, 1986.)



Perhaps the designers thought the correct model was too complex, that the model they were giving was easier to understand. But with the wrong conceptual model, it is impossible to set the controls. And even though I am convinced I now know the correct model, I still cannot accurately adjust the temperatures because the refrigerator design makes it impossible for me to discover which control is for the thermostat, which control is for the relative proportion of cold air, and in which compartment the thermostat is located. The lack of immediate feedback for the actions does not help: with a delay of twenty-four hours, who can remember what was tried?

The topic of conceptual models will reappear in the book. They are part of an important concept in design: *mental models,* the models people have of themselves, others, the environment, and the things with which they interact. People form mental models through experience, training, and instruction. The mental model of a device is formed largely by interpreting its perceived actions and its visible structure. I call the visible part of the device the *system image* (figure 1.10). When the system image is incoherent or inappropriate, as in the case of the refrigerator, then the user cannot easily use the device. If it is incomplete or contradictory, there will be trouble.

MAKE THINGS VISIBLE

The problems caused by inadequate attention to visibility are all neatly demonstrated with one simple appliance: the modern telephone.

I stand at the blackboard in my office, talking with a student, when my telephone rings. Once, twice it rings. I pause, trying to complete my sentence before answering. The ringing stops. "I'm sorry," says the student. "Not your fault," I say. "But it's no problem, the call now transfers to my secretary's phone. She'll answer it." As we listen we hear her phone start to ring. Once, twice. I look at my watch. Six o'clock: it's late, the office staff has left for the day. I rush out of my office to my secretary's phone, but as I get there, it stops ringing. "Ah," I think, "it's being transferred to another phone." Sure enough, the phone in the adjacent office now starts ringing. I rush to that office, but it is locked. Back to my office to get the key, out to the locked door, fumble with the lock, into the office, and to the now quiet phone. I hear a telephone down the hall start to ring. Could that still be my call, *making its way mysteriously, with a predetermined lurching path, through the phones of the building? Or is it just another telephone call coincidentally arriving at this time?*

In fact, I could have retrieved the call from my office, had I acted quickly enough. The manual states: "Within your pre-programmed pick-up group, dial 14 to connect to incoming call. Otherwise, to answer any ringing extension, dial ringing extension number, listen for busy tone. Dial 8 to connect to incoming call." Huh? What do those instructions mean? What is a "pre-programmed pick-up group," and why do I even want to know? What is the extension number of the ringing phone? Can I remember all those instructions when I need them? No.

Telephone chase is the new game in the modern office, as the automatic features of telephones go awry—features designed without proper thought, and certainly without testing them with their intended users. There are several other games, too. One game is announced by the plea, "How do I answer this call?" The question is properly whined in front of a ringing, flashing telephone, receiver in hand. Then there is the paradoxical game entitled "This telephone doesn't have a hold function." The accusation is directed at a telephone that actually *does* have a hold function. And, finally, there is "What do you mean I called you, you called me!"

Many of the modern telephone systems have a new feature that automatically keeps trying to dial a number for you. This feature resides under names such as automatic redialing or automatic callback. I am supposed to use this feature whenever I call someone who doesn't answer or whose line is busy. When the person next hangs up the phone, my phone will dial it again. Several automatic callbacks can be active at a time. Here's how it works. I place a phone call. There's no answer, so I activate the automatic callback feature. Several hours later my telephone rings. I pick it up and say "Hello," only to hear a ringing sound and then someone else saying "Hello."

"Hello," I answer, "who is this?"

"Who is this?" I hear in reply, "you called me."

"No," I say, "you called me, my phone just rang."

Slowly I realize that perhaps this is my delayed call. Now, let me see, who was I trying to call several hours ago? Did I have several callbacks in place? Why was I making the call?

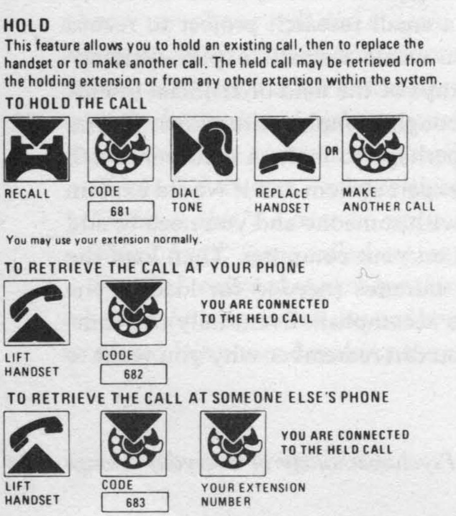
The modern telephone did not happen by accident: it was carefully designed. Someone—more likely a team of people—invented a list of features thought desirable, invented what seemed to them to be plausible ways of controlling the features, and then put it all together. My university, focusing on cost and perhaps dazzled by the features, bought the system, spending millions of dollars on a telephone installation that has proved vastly unpopular and even unworkable. Why did the university buy the system? The purchase took several years of committee work and studies and presentations by competing telephone companies, and piles of documentation and specification. I myself took part, looking at the interaction between the telephone system and the computer networks, ensuring that the two would be compatible and reasonable in price. To my knowledge, nobody ever thought of trying out the telephones in advance. Nobody suggested installing them in a sample office to see whether users' needs would be met or whether users could understand how to operate the phone. The result: disaster. The main culprit—lack of visibility—was coupled with a secondary culprit—a poor conceptual model. Any money saved on the installation and purchase is quickly disappearing in training costs, missed calls, and frustration. Yet from what I have seen, the competing phone systems would not have been any better.

I recently spent six months at the Applied Psychology Unit in Cambridge, England. Just before I arrived the British Telecom Company had installed a new telephone system. It had lots and lots of features. The telephone instrument itself was unremarkable (figure 1.11). It was the standard twelve-button, push-button phone, except that it had an extra key labeled "R" off on the side. (I never did find out what that key did.)

The telephone system was a standing joke. Nobody could use all the features. One person even started a small research project to record people's confusions. Another person wrote a small "expert systems" computer program, one of the new toys of the field of artificial intelligence; the program can reason through complex situations. If you wanted to use the phone system, perhaps to make a conference call among three people, you asked the expert system and it would explain how to do it. So, you're on the line with someone and you need to add a third person to the call. First turn on your computer. Then load the expert system. After three or four minutes (needed for loading the program), type in what you want to accomplish. Eventually the computer will tell you what to do—if you can remember why you want to

1.11 British Telecom Telephone. This was in my office at the Applied Psychology Unit in Cambridge, England. It certainly looks simple, doesn't it?

1.12 Two Ways to Use Hold on Modern Telephones. Illustration *A* (below left) is the instruction manual page for British Telecom. The procedure seems especially complicated, with three 3-digit codes to be learned: 681, 682, and 683. Illustration *B* (below right) shows the equivalent instructions for the Ericsson Single Line Analog Telephone installed at the University of California, San Diego. I find the second set of instructions easier to understand, but one must still dial an arbitrary digit: 8 in this case.



HOLD

This feature allows you to hold an existing call, then to replace the handset or to make another call. The held call may be retrieved from the holding extension or from any other extension within the system.

TO HOLD THE CALL

RECALL — CODE 681 — DIAL TONE — REPLACE HANDSET — OR — MAKE ANOTHER CALL

You may use your extension normally.

TO RETRIEVE THE CALL AT YOUR PHONE

LIFT HANDSET — CODE 682 — YOU ARE CONNECTED TO THE HELD CALL

TO RETRIEVE THE CALL AT SOMEONE ELSE'S PHONE

LIFT HANDSET — CODE 683 — YOUR EXTENSION NUMBER — YOU ARE CONNECTED TO THE HELD CALL

CALL HOLD/CALL PARK

With party on line

- Press **R** key
- Listen for recall dial tone (three beeps and dial tone)
- Hang up handset

TO RETRIEVE FROM SAME PHONE

- Lift handset; you are connected to the call

TO RETRIEVE FROM ANOTHER PHONE

- Lift handset
- Dial extension where call was parked; listen for busy tone
- Dial **8**; you are connected to the call

NOTE: Call will remain parked for 3 minutes before re-ringing do it, and if the person on the other end of the line is still around. But, as it happens, using the expert system is a lot easier than reading and understanding the manual provided with the telephone (figure 1.12).

Why is that telephone system so hard to understand? Nothing in it is conceptually difficult. Each of the operations is actually quite simple. A few digits to dial, that's all. The telephone doesn't even look complicated. There are only fifteen controls: the usual twelve buttons—ten labeled 0 through 9, #, and *—plus the handset itself, the handset button, and the mysterious "R" button. All except the "R" are the everyday parts of a normal modern telephone. Why was the system so difficult?

A designer who works for a telephone company told me the following story:

"I was involved in designing the faceplate of some of those new multifunction phones, some of which have buttons labeled "R." *The* "R" *button is kind of a vestigial feature. It is very hard to remove features of a newly designed product that had existed in an earlier version. It's kind of like physical evolution. If a feature is in the genome, and if that feature is not associated with any negativity (i.e., no customers gripe about it), then the feature hangs on for generations.*

"It is interesting that things like the "R" *button are largely determined through examples. Somebody asks, 'What is the* "R" *button used for?' and the answer is to give an example: 'You can push* "R" *to access loudspeaker paging.' If nobody can think of an example, the feature is dropped. Designers are pretty bright people, however. They can come up with a plausible-sounding example for almost anything. Hence, you get features, many many features, and these features hang on for a long time. The end result is complex interfaces for essentially simple things."*[6]

As I pondered this problem, I decided it would make sense to compare the phone system with something that was of equal or greater complexity but easier to use. So let us temporarily leave the difficult telephone system and take a look at my automobile. I bought a car in Europe. When I picked up the new car at the factory, a man from the company sat in the car with me and went over each control, explaining its function. When he had gone through the controls once, I said fine, thanked him, and drove away. That was all the instruction it took. There are 112 controls inside the car. This isn't quite as bad as it sounds. Twenty-five of them are on the radio. Another 7 are the temperature control system, and 11 work the windows and sunroof. The trip computer has 14 buttons, each matched with a specific function. So four devices—the radio, temperature controls, windows, and trip computer—have together 57 controls, or just over 50 percent of the ones available.

Why is the automobile, with all its varied functions and numerous controls, so much easier to learn and to use than the telephone system, with its much smaller set of functions and controls? What is good about the design of the car? Things are visible. There are good mappings, natural relationships, between the controls and the things controlled. Single controls often have single functions. There is good feedback. The system is understandable. In general, the relationships among the user's intentions, the required actions, and the results are sensible, nonarbitrary, and meaningful.

What is bad about the design of the telephone? There is no visible structure. Mappings are arbitrary: there is no rhyme or reason to the relationship between the actions the user must perform and the results to be accomplished. The controls have multiple functions. There isn't good feedback, so the user is never sure whether the desired result has been obtained. The system, in general, is not understandable; its capabilities aren't apparent. In general, the relationships among the user's intentions, the required actions, and the results are completely arbitrary.

Whenever the number of possible actions exceeds the number of controls, there is apt to be difficulty. The telephone system has twenty-four functions, yet only fifteen controls—none of them labeled for specific action. In contrast, the trip computer for the car performs seventeen functions with fourteen controls. With minor exceptions, there is one control for each function. In fact, the controls with more than one function are indeed harder to remember and use. When the number of controls equals the number of functions, each control can be specialized, each can be labeled. The possible functions are visible, for each corresponds with a control. If the user forgets the functions, the controls serve as reminders. When, as on the telephone, there are more functions than controls, labeling becomes difficult or impossible. There is nothing to remind the user. Functions are invisible, hidden from sight. No wonder the operation becomes mysterious and difficult. The controls for the car are visible and, through their location and mode of operation, bear an intelligent relationship to their action. Visibility acts as a good reminder of what can be done and allows the control to specify how the action is to be performed. The good relationship between the placement of the control and what it does makes it easy to find the appropriate control for a task. As a result, there is little to remember.

THE PRINCIPLE OF MAPPING

Mapping is a technical term meaning the relationship between two things, in this case between the controls and their movements and the results in the world. Consider the mapping relationships involved in steering a car. To turn the car to the right, one turns the steering wheel clockwise (so that its top moves to the right). The user must identify two mappings here: one of the 112 controls affects the steering, and the steering wheel must be turned in one of two directions. Both are somewhat arbitrary. But the wheel and the clockwise direction are natural choices: visible, closely related to the desired outcome, and providing immediate feedback. The mapping is easily learned and always remembered.

Natural mapping, by which I mean taking advantage of physical analogies and cultural standards, leads to immediate understanding. For example, a designer can use spatial analogy: to move an object up, move the control up. To control an array of lights, arrange the controls in the same pattern as the lights. Some natural mappings are cultural or biological, as in the universal standard that a rising level represents more, a diminishing level, less. Similarly, a louder sound can mean a greater amount. Amount and loudness (and weight, line length, and brightness) are additive dimensions: add more to show incremental increases. Note that the logically plausible relationship between musical pitch and amount does not work: Would a higher pitch mean less or more of something? Pitch (and taste, color, and location) are substitutive dimensions: substitute one value for another to make a change. There is no natural concept of more or less in the comparison of different pitches, or hues, or taste qualities. Other natural mappings follow from the principles of perception and allow for the natural grouping or patterning of controls and feedback (see figure 1.13).

Mapping problems are abundant, one of the fundamental causes of difficulties. Consider the telephone. Suppose you wish to activate the callback on "no reply" function. To initiate this feature on one tele-

1.13 Seat Adjustment Control from a Mercedes-Benz Automobile. This is an excellent example of natural mapping. The control is in the shape of the seat itself: the mapping is straightforward. To move the front edge of the seat higher, lift up on the front part of the button. To make the seat back recline, move the button back. Mercedes-Benz automobiles are obviously not everyday things for most people, but the principle doesn't require great expense or wealth. The same principle could be applied to much more common objects.

phone system, press and release the "recall" button (the button on the handset), then dial 60, then dial the number you called.

There are several problems here. First, the description of the function is relatively complex—yet incomplete: What if two people set up callback at the same time? What if the person does not come back until a week later? What if you have meanwhile set up three or four other functions? What if you want to cancel it? Second, the action to be performed is arbitrary. (Dial 60. Why 60? Why not 73 or 27? How does one remember an arbitrary number?) Third, the sequence ends with what appears to be a redundant, unnecessary action: dialing the number of the person to be called. If the phone system is smart enough to do all these other things, why can't it remember the number that was just attempted; why must it be told all over again? And finally, consider the lack of feedback. How do I know I did the right action? Maybe I disconnected the phone. Maybe I set up some other special feature. There is no visible or audible way to know immediately.

A device is easy to use when there is visibility to the set of possible actions, where the controls and displays exploit natural mappings. The principles are simple but rarely incorporated into design. Good design takes care, planning, thought. It takes conscious attention to the needs of the user. And sometimes the designer gets it right:

Once, when I was at a conference at Gmunden, Austria, a group of us went off to see the sights. I sat directly behind the driver of the brand new, sleek, high-technology German tour bus. I gazed in wonder at the hundreds of controls scattered all over the front of the bus.

"How can you ever learn all those controls?" I asked the driver (with the aid of a German-speaking colleague). The driver was clearly puzzled by the question.

"What do you mean?" he replied. "Each control is just where it ought to be. There is no difficulty."

A good principle, that. Controls are where they ought to be. One function, one control. Harder to do, of course, than to say, but essentially this is the principle of natural mappings: the relationship between controls and actions should be apparent to the user. I return to this topic later in the book, for the problem of determining the "naturalness" of mappings is difficult, but crucial.

I've already described how my car's controls are generally easy to use. Actually, the car has lots of problems. The approach to usability used in the car seems to be to make sure that you can reach everything and see everything. That's good, but not nearly good enough.

Here is a simple example: the controls for the loudspeakers—a simple control that determines whether the sound comes out of the front speakers, the rear, or a combination (figure 1.14). Rotate the wheel from left to right or right to left. Simple, except how do you know which way to rotate the control? Which direction moves the sound to the rear, which to the front? If you want sound to come out of the front speaker, you should be able to move the control to the front. To get it out of the back, move the control to the back. Then the form of the motion would mimic the function and make a natural mapping. But the way the control is actually mounted in the car, forward and backward get translated into left and right. Which direction is which? There is no natural relationship. What's worse, the control isn't even labeled. Even the instruction manual does not say how to use it.

1.14 The Front/Rear Speaker Selector of an Automobile Radio. Rotating the knob with the pictures of the speaker at either side makes the sound come entirely out of the front speakers (when the knob is all the way over to one side), entirely out of the rear speakers (when the knob is all the way the other way), or equally out of both (when the knob is midway). Which way is front, which rear? You can't tell by looking. While you're at it, imagine trying to manipulate the radio controls while keeping your eyes on the road.

The control should be mounted so that it moves forward and backward. If that can't be done, rotate the control 90° on the panel so that it moves vertically. Moving something up to represent forward is not as natural as moving it forward, but at least it follows a standard convention.

In fact, we see that both the car and the telephone have easy functions and difficult ones. The car seems to have more of the easy ones, the telephone more of the difficult ones. Moreover, with the car, enough of the controls are easy that I can do almost everything I need to. Not so with the telephone: it is very difficult to use even a single one of the special features.

The easy things on both telephone and car have a lot in common, as do the difficult things. When things are visible, they tend to be easier than when they are not. In addition, there must be a close, *natural* relationship between the control and its function: a *natural mapping.*

THE PRINCIPLE OF FEEDBACK

Feedback—sending back to the user information about what action has actually been done, what result has been accomplished—is a well-known concept in the science of control and information theory. Imagine trying to talk to someone when you cannot even hear your own voice, or trying to draw a picture with a pencil that leaves no mark: there would be no feedback.

In the good old days of the telephone, before the American telephone system was divided among competing companies, before telephones were fancy and had so many features, telephones were designed with much more care and concern for the user. Designers at the Bell Telephone Laboratories worried a lot about feedback. The push buttons were designed to give an appropriate feel—tactile feedback. When a button was pushed, a tone was fed back into the earpiece so the user could tell that the button had been properly pushed. When the phone call was being connected, clicks, tones, and other noises gave the user feedback about the progress of the call. And the speaker's voice was always fed back to the earpiece in a carefully controlled amount, because the auditory feedback (called "sidetone") helped the person regulate how loudly to talk. All this has changed. We now have telephones that are much more powerful and often cheaper than those that existed just a few years ago—more function for less money. To be fair, these new designs are pushing hard on the paradox of technology: added functionality generally comes along at the price of added complexity. But that does not justify backward progress.

Why are the modern telephone systems so difficult to learn and to use? Basically, the problem is that the systems have more features and less feedback. Suppose all telephones had a small display screen, not unlike the ones on small, inexpensive calculators. The display could be used to present, upon the push of a button, a brief menu of all the features of the telephone, one by one. When the desired one was encountered, the user would push another button to indicate that it should be invoked. If further action was required, the display could tell the person what to do. The display could even be auditory, with speech instead of a visual display. Only two buttons need be added to the telephone: one to change the display, one to accept the option on display. Of course, the telephone would be slightly more expensive. The tradeoff is cost versus usability.[7]

Pity the Poor Designer

Designing well is not easy. The manufacturer wants something that can be produced economically. The store wants something that will be attractive to its customers. The purchaser has several demands. In the store, the purchaser focuses on price and appearance, and perhaps on prestige value. At home, the same person will pay more attention to functionality and usability. The repair service cares about maintainability: how easy is the device to take apart, diagnose, and service? The needs of those concerned are different and often conflict. Nonetheless, the designer may be able to satisfy everyone.

A simple example of good design is the $3\frac{1}{2}$-inch magnetic diskette for computers, a small circle of "floppy" magnetic material encased in hard plastic. Earlier types of floppy disks did not have this plastic case, which protects the magnetic material from abuse and damage. A sliding metal cover protects the delicate magnetic surface when the diskette is not in use and automatically opens when the diskette is inserted into the computer. The diskette has a square shape: there are apparently eight possible ways to insert it into the machine, only one of which is correct. What happens if I do it wrong? I try inserting the disk sideways. Ah, the designer thought of that. A little study shows that the case really isn't square: it's rectangular, so you can't insert a longer side. I try backward. The diskette goes in only part of the way. Small protrusions, indentations, and cutouts prevent the diskette from being inserted backward or upside down: of the eight ways one might try to insert the diskette, only one is correct, and only that one will fit. An excellent design.

Take another example of good design. My felt-tipped marking pen has ribs along only one of its sides; otherwise all sides look identical. Careful examination shows that the tip of the marker is angled and makes the best line if the marker is held with the ribbed side up, a natural result if the forefinger rests upon the ribs. No harm results if I hold the marker another way, but the marker writes less well. The ribs are a subtle design cue—functional, yet visibly and aesthetically unobtrusive.

The world is permeated with small examples of good design, with the amazing details that make important differences in our lives. Each detail was added by some person, a designer, carefully thinking through the uses of the device, the ways that people abuse things, the kinds of errors that can get made, and the functions that people wish to have performed.

Then why is it that so many good design ideas don't find their way into products in the marketplace? Or something good shows up for a short time, only to fall into oblivion? I once spoke with a designer about the frustrations of trying to get the best product out:

It usually takes five or six attempts to get a product right. This may be acceptable in an established product, but consider what it means in a new one. Suppose a company wants to make a product that will perhaps make a real difference. The problem is that if the product is truly revolutionary, it is unlikely that anyone will quite know how to design it right the first time; it will take several tries. But if a product is introduced into the marketplace and fails, well that is it. Perhaps it could be introduced a second time, or maybe even a third time, but after that it is dead: everyone believes it to be a failure.

I asked him to explain. "You mean," I said, "that it takes five or six tries to get an idea right?"

"Yes," he said, "at least that."

"But," I replied, "you also said that if a newly introduced product doesn't catch on in the first two or three times, then it is dead?"

"Yup," he said.

"Then new products are almost guaranteed to fail, no matter how good the idea."

"Now you understand," said the designer. "Consider the use of voice messages on complex devices such as cameras, soft-drink machines, and copiers. A failure. No longer even tried. Too bad. It really is a good idea, for it can be very useful when the hands or eyes are busy elsewhere. But those first few attempts were very badly done and the public scoffed—properly. Now, nobody dares try it again, even in those places where it is needed."

The Paradox of Technology

Technology offers the potential to make life easier and more enjoyable; each new technology provides increased benefits. At the same time, added complexities arise to increase our difficulty and frustration. The development of a technology tends to follow a U-shaped curve of complexity: starting high; dropping to a low, comfortable level; then climbing again. New kinds of devices are complex and difficult to use. As technicians become more competent and an industry matures, devices become simpler, more reliable, and more powerful. But then, after the industry has stabilized, newcomers figure out how to add increased power and capability, but always at the expense of added complexity and sometimes decreased reliability. We can see the curve of complexity in the history of the watch, radio, telephone, and television set. Take the radio. In the early days, radios were quite complex. To tune in a station required several adjustments, including one for the antenna, one for the radio frequency, one for intermediate frequencies, and controls for both sensitivity and loudness. Later radios were simpler and had controls only to turn it on, tune the station, and adjust the loudness. But the latest radios are again very complex, perhaps even more so than early ones. Now the radio is called a tuner, and it is littered with numerous controls, switches, slide bars, lights, displays, and meters. The modern sets are technologically superior, offering higher quality sound, better reception, and enhanced capability. But what good is the technology if it is too complex to use?

The design problem posed by technological advances is enormous. Consider the watch. A few decades ago, watches were simple. All you had to do was set the time and keep them wound. The standard control was the stem: a knob at the side of the watch. Turning the knob wound the spring that worked the watch. Pulling the knob out and turning it made the hands move. The operations were easy to learn and easy to do. There was a reasonable relation between the turning of the knob and the resulting turning of the hands. The design even took into account human error: the normal position of the stem was for winding the spring, so that an accidental turn would not reset the time.

In the modern digital watch the spring is gone, replaced by a motor run by long-lasting batteries. All that remains is the task of setting the watch. The stem is still a sensible solution, for you can go fast or slow, forward or backward, until the exact desired time is reached. But the stem is more complex (and therefore more expensive) than simple push-button switches. If the only change in the transition from the spring-wound analog watch to the battery-run digital watch were in how the time was set, there would be little difficulty. The problem is that new technology has allowed us to add more functions to the watch: the watch can give the day of the week, the month, and the year; it can act as a stop watch (which itself has several functions), a count-down timer, and an alarm clock (or two); it has the ability to show the time for different time zones; it can act as a counter and even as a calculator. But the added functions cause problems: How do you design a watch that has so many functions while trying to limit the size, cost, and complexity of the device? How many buttons does it take to make the watch workable and learnable, yet not too expensive? There are no easy answers. Whenever the number of functions and required operations exceeds the number of controls, the design becomes arbitrary, unnatural, and complicated. The same technology that simplifies life by providing more functions in each device also complicates life by making the device harder to learn, harder to use. This is the paradox of technology.

The paradox of technology should never be used as an excuse for poor design. It is true that as the number of options and capabilities of any device increases, so too must the number and complexity of the controls. But the principles of good design can make complexity manageable.

In one of my courses I gave as homework the assignment to design a multiple-function clock radio:

You have been employed by a manufacturing company to design their new product. The company is considering combining the following into one item:

- *AM-FM radio*
- *Cassette player*
- *CD player*
- *Telephone*
- *Telephone answering machine*
- *Clock*
- *Alarm clock (the alarm can turn on a tone, radio, cassette, or CD)*
- *Desk or bed lamp*

The company is trying to decide whether to include a small (two-inch screen) TV set and a switched electric outlet that can turn on a coffee maker or toaster.

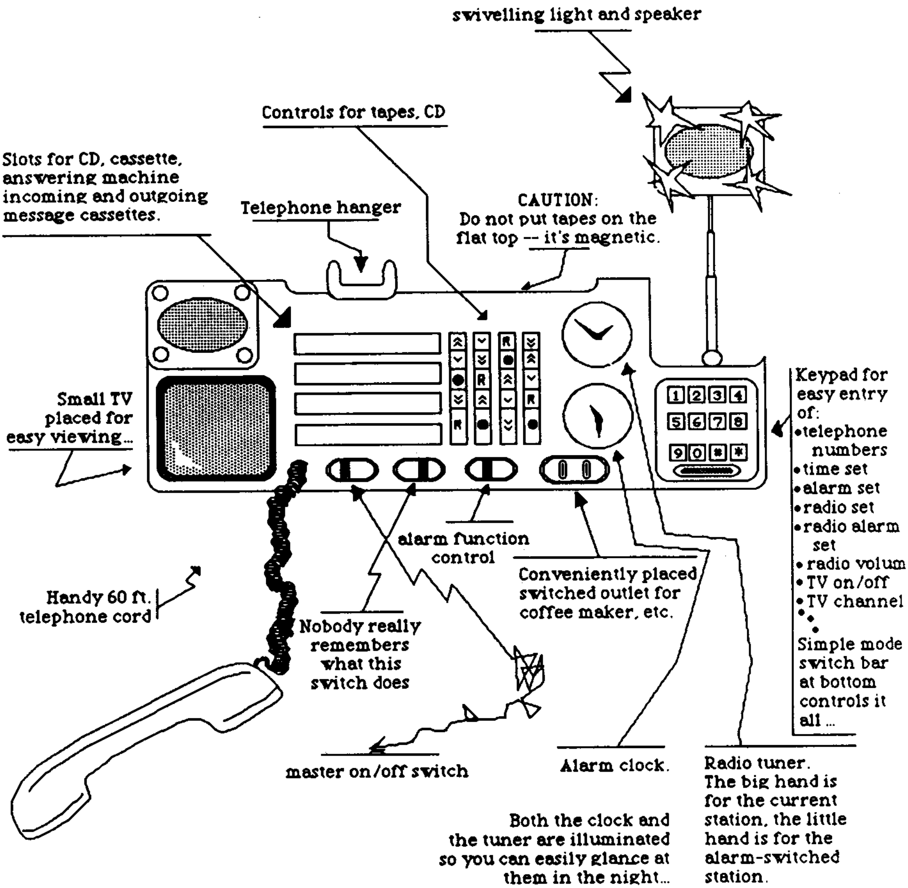
Your job is (A) to recommend what to build, then (B) to design the control panel, and finally (C) to certify that it is actually both what customers want and easy to use.

State what you would do for the three parts of your job: A, B, and C. Explain how you would go about validating and justifying your recommendations.

Draw a rough sketch of a control panel for the items in the indented list, with a brief *justification and analysis of the factors that went into the choice of design.*

There are several things I looked for in the answer. (Figure 1.15 is an unacceptable solution.) First, how well did the answer address the

1.15 Possible Solution to My Homework Assignment. Completely unacceptable. (Thanks to Bill Gaver for devising and drawing this sample.)



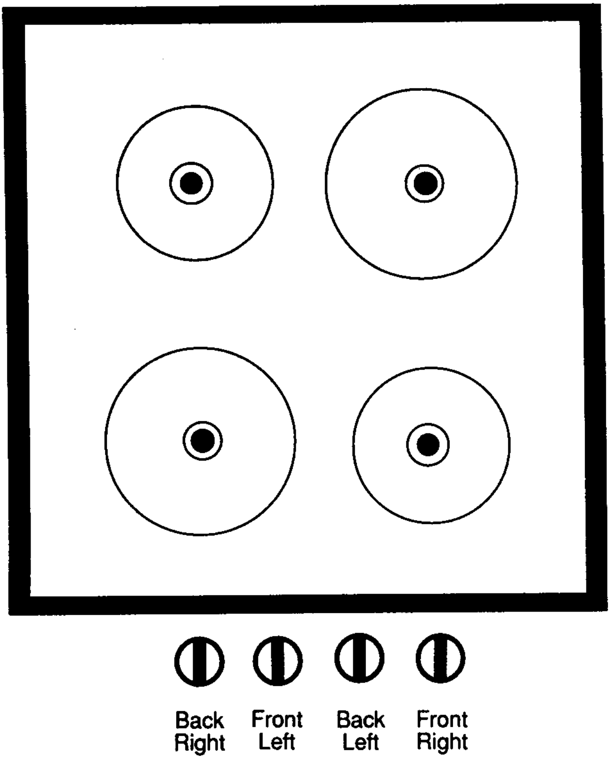
real needs of the user? I expected my students to visit the homes of potential users to see how their current devices were being used and to determine how the combined multipurpose device would be used. Next, I evaluated whether all the controls were usable and understandable, allowing all the desired functions to be operated with minimum confusion or error. Clock radios are often used in the dark, with the user in bed and reaching overhead to grope for the desired control. Therefore the unit had to be usable in the dark by feel only. It was not supposed to be possible to make a serious mistake by accidentally hitting the wrong control. (Alas, many existing clock radios do not tolerate serious errors—for example, the user may reset the time by hitting the wrong button accidentally.) Finally, the design was expected to take into account real issues in cost, manufacturability, and aesthetics. The finished design had to pass muster with users. The point of the exercise was for the student to realize the paradox of technology: added complexity and difficulty cannot be avoided when functions are added, but with clever design, they can be minimized.

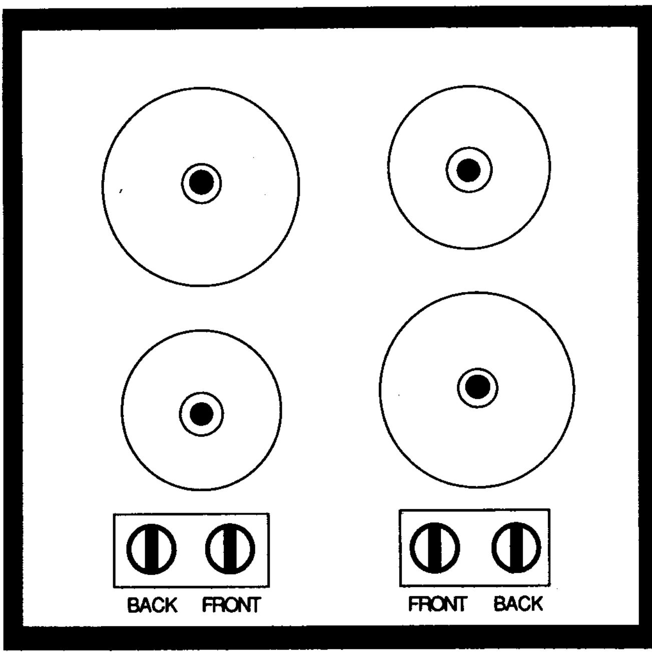
NATURAL MAPPINGS

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 (4×3) 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

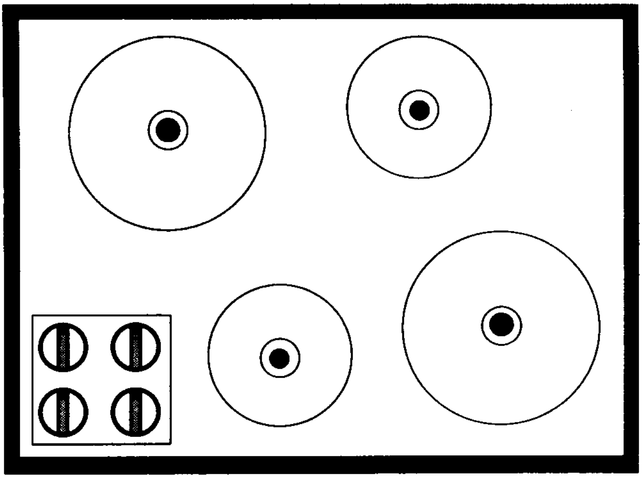

Back Right
Front Left
Back Left
Front Right

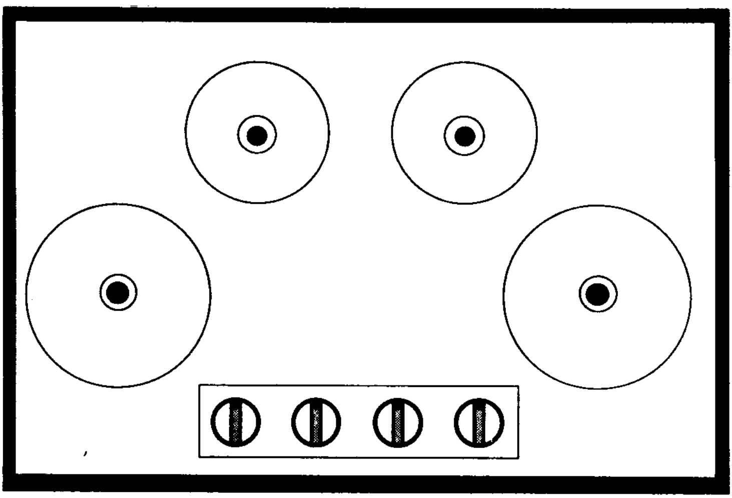

BACK
FRONT
FRONT
BACK

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-

3.6 Tradeoffs

PROPERTY	KNOWLEDGE IN THE WORLD	KNOWLEDGE IN THE HEAD
Retrievability	Retrievable whenever visible or audible.	Not readily retrievable. Requires memory search or reminding.
Learning	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.	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).
Efficiency of use	Tends to be slowed up by the need to find and interpret the external information.	Can be very efficient.
Ease of use at first encounter	High.	Low.
Aesthetics	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.	Nothing need be visible, which gives more freedom to the designer, which in turn can lead to better aesthetics.

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 locations 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.[17]