The Design of Everyday Things

Donald A. Norman

Even the smartest among us can feel inept as we fail to figure out which light switch or oven burner to turn on, or whether to push, pull, or slide a door. The fault lies in product design that ignores the needs of users and the principles of cognitive psychology. A best-seller in the United States, this bible on the cognitive aspects of design contains examples of both good and bad design and simple rules that designers can use to improve the usability of objects as diverse as cars, computers, doors, and telephones.

Donald A. Norman is an executive at Hewlett-Packard. Formerly, he was Vice President and Apple Fellow at Apple Computer, where he headed the Apple Research Laboratories, and Professor of Cognitive Science at the University of California at San Diego. He is the author of many books, including *Things That Make Us Smart* and *The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex, and Information Appliances Are the Solution* (MIT Press, 1998).

"Provocative."
Time

"Norman . . . makes a strong case for the needlessness of badly conceived and badly designed everyday objects. . . . [T]his book may herald the beginning of a change in user habits and expectations, a change that manufacturers would be obliged to respond to. Button pushers of the world, unite."
Los Angeles Times

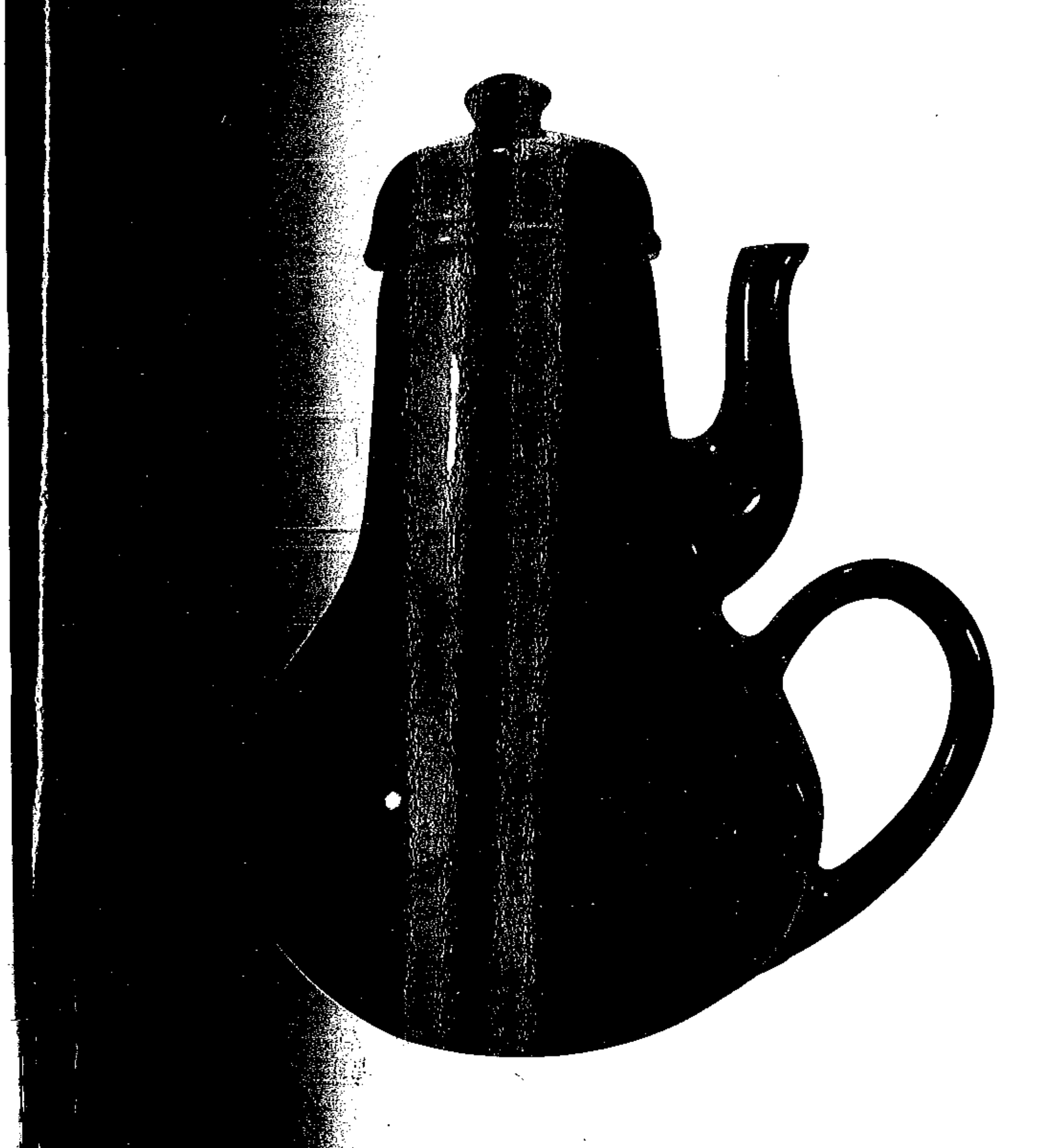
"[A] thoughtful exploration of Man vs. Machine."
People

"We are all victimized by the natural perversity of inanimate objects. Here is a book at last that strikes back both at the objects and at the designers, manufacturers, and assorted human beings who originate and maintain this perversity. It will do your heart good and may even point the way to correcting matters."
Isaac Asimov

The MIT Press Massachusetts Institute of Technology Cambridge, Massachusetts 02142
http://mitpress.mit.edu

NORDP 0-262-64037-6
90000
9 780262 640374

The Design of Everyday Things



Donald A. Norman

THE DESIGN OF EVERYDAY THINGS

Donald A. Norman

The MIT Press
London, England

First MIT Press edition, 1998

This book was originally published in hardcover by Basic Books in 1988 and in paperback by Doubleday/Currency in 1990.

Printed and bound in the United States of America.

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, 1988.
Includes bibliographical references and index.
1. Design, Industrial—Psychological aspects. 2. Human engineering. I. Title.
[TS171.4.N67 1990]
620.8'2—dc20 89-48989
CIP

ISBN 0-262-64037-6 (pb.: alk. paper)

CONTENTS

CHAPTER ONE

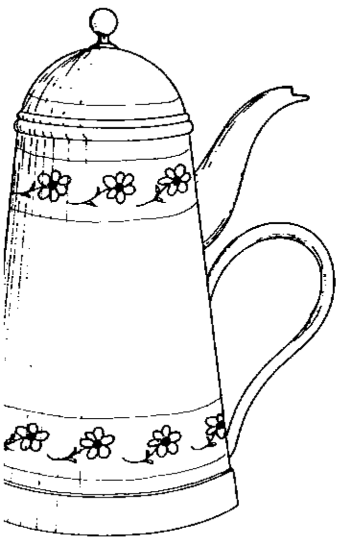
THE PSYCHOPATHOLOGY OF EVERYDAY THINGS

"Kenneth Olsen, the engineer who founded and still runs Digital Equipment Corp., confessed at the annual meeting that he can't figure out how to heat a cup of coffee in the company's microwave oven."[1]

You Would Need an Engineering Degree to Figure This Out

"You would need an engineering degree from MIT to work this," someone once told me, shaking his head in puzzlement over his brand new digital watch. Well, I have an engineering degree from MIT. (Kenneth Olsen has two of them, and he can't figure out a microwave oven.) Give me a few hours and I can figure out the watch. But why should it take hours? I have talked with many people who can't use all the features of their washing machines or cameras, who can't figure out how to work a sewing machine or a video cassette recorder, who habitually turn on the wrong stove burner.

Why do we put up with the frustrations of everyday objects, with objects that we can't figure out how to use, with those neat plastic-wrapped packages that seem impossible to open, with doors that trap people, with washing machines and dryers that have become too con-



1.1 **Carelman's Coffeepot for Masochists.** The French artist Jacques Carelman in his series of books *Catalogue d'objets introuvables (Catalog of unfindable objects)* provides delightful examples of everyday things that are deliberately unworkable, outrageous, or otherwise ill-formed. Jacques Carelman: "Coffeepot for Masochists." 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.

sing to use, with audio-stereo-television-video-cassette-recorders at claim in their advertisements to do everything, but that make it nost impossible to do anything?

The human mind is exquisitely tailored to make sense of the world. ive it the slightest clue and off it goes, providing explanation, rationization, understanding. Consider the objects—books, radios, kitchen pliances, office machines, and light switches—that make up our evyday lives. Well-designed objects are easy to interpret and underand. They contain visible clues to their operation. Poorly designed jects can be difficult and frustrating to use. They provide no clues— sometimes false clues. They trap the user and thwart the normal ocess of interpretation and understanding. Alas, poor design edominates. The result is a world filled with frustration, with objects at cannot be understood, with devices that lead to error. This book an attempt to change things.

The Frustrations of Everyday Life

I were placed in the cockpit of a modern jet airliner, my inability to erform gracefully and smoothly would neither surprise nor bother me. ut I shouldn't have trouble with doors and switches, water faucets nd stoves. "Doors?" I can hear the reader saying, "you have trouble

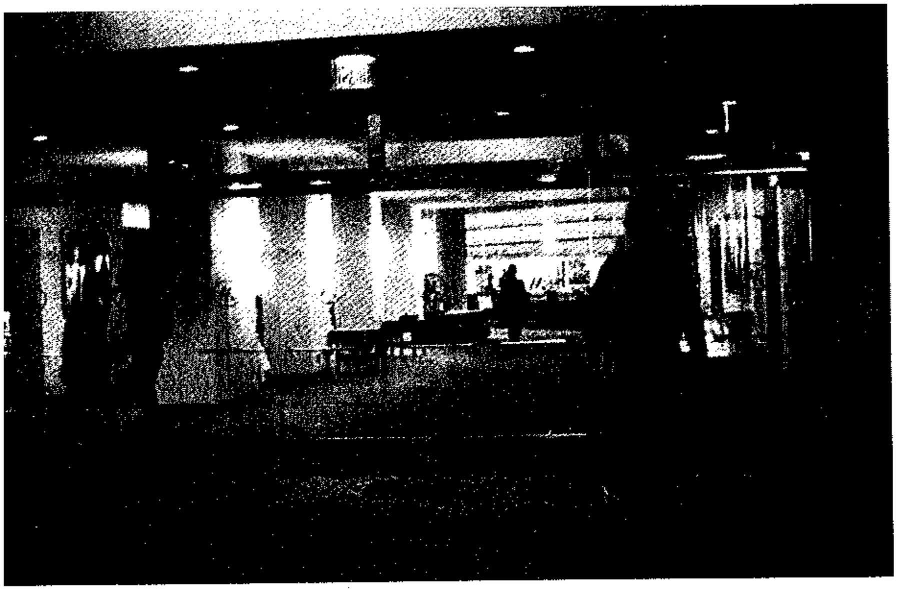
opening doors?" Yes. I push doors that are meant to be pulled, pull doors that should be pushed, and walk into doors that should be slid. Moreover, I see others having the same troubles—unnecessary troubles. There are psychological principles that can be followed to make these things understandable and usable.

Consider the door. There is not much you can do to a door: you can open it or shut it. Suppose you are in an office building, walking down a corridor. You come to a door. In which direction does it open? Should you pull or push, on the left or the right? Maybe the door slides. If so, in which direction? I have seen doors that slide up into the ceiling. A door poses only two essential questions: In which direction does it move? On which side should one work it? The answers should be given by the design, without any need for words or symbols, certainly without any need for trial and error.

A friend told me of the time he got trapped in the doorway of a post office in a European city. The entrance was an imposing row of perhaps six glass swinging doors, followed immediately by a second, identical row. That's a standard design: it helps reduce the airflow and thus maintain the indoor temperature of the building.

My friend pushed on the side of one of the leftmost pair of outer doors. It swung inward, and he entered the building. Then, before he could get to the next row of doors, he was distracted and turned around for an instant. He didn't realize it at the time, but he had moved slightly to the right. So when he came to the next door and pushed it, nothing happened. "Hmm," he thought, "must be locked." So he pushed the side of the adjacent door. Nothing. Puzzled, my friend decided to go outside again. He turned around and pushed against the side of a door. Nothing. He pushed the adjacent door. Nothing. The door he had just entered no longer worked. He turned around once more and tried the inside doors again. Nothing. Concern, then mild panic. He was trapped! Just then, a group of people on the other side of the entranceway (to my friend's right) passed easily through both sets of doors. My friend hurried over to follow their path.

How could such a thing happen? A swinging door has two sides. One contains the supporting pillar and the hinge, the other is unsupported. To open the door, you must push on the unsupported edge. If you push on the hinge side, nothing happens. In this case, the designer aimed for beauty, not utility. No distracting lines, no visible pillars, no visible hinges. So how can the ordinary user know which side to push



1.2 A Row of Swinging Glass Doors in a Boston Hotel. A similar problem to the doors from that European post office. On which side of the door should you push? When I asked people who had just used the doors, most couldn't say. Yet only a few of the people I watched had trouble with the doors. The designers had incorporated a subtle clue into the design. Note that the horizontal bars are not centered: they are a bit closer together on the sides you should push on. The design almost works—but not entirely, for not everyone used the doors right on the first try.

on? While distracted, my friend had moved toward the (invisible) supporting pillar, so he was pushing the doors on the hinged side. No wonder nothing happened. Pretty doors. Elegant. Probably won a design prize.

The door story illustrates one of the most important principles of design: *visibility.* The correct parts must be visible, and they must convey the correct message. With doors that push, the designer must provide signals that naturally indicate where to push. These need not destroy the aesthetics. Put a vertical plate on the side to be pushed, nothing on the other. Or make the supporting pillars visible. The vertical plate and supporting pillars are *natural* signals, *naturally* interpreted, without any need to be conscious of them. I call the use of natural signals *natural design* and elaborate on the approach throughout this book.

Visibility problems come in many forms. My friend, trapped between the glass doors, suffered from a lack of clues that would indicate what part of a door should be operated. Other problems concern the *mappings* between what you want to do and what appears to be possible, another topic that will be expanded upon throughout the book. Consider one type of slide projector. This projector has a single button to control whether the slide tray moves forward or backward. One button to do two things? What is the mapping? How can you figure out how to control the slides? You can't. Nothing is visible to give the slightest hint. Here is what happened to me in one of the many unfamiliar places I've lectured in during my travels as a professor:

The Leitz slide projector illustrated in figure 1.3 has shown up several times in my travels. The first time, it led to a rather dramatic incident. A conscientious student was in charge of showing my slides. I started my talk and showed the first slide. When I finished with the first slide and asked for the next, the student carefully pushed the control button and watched in dismay as the tray backed up, slid out of the projector and plopped off the table onto the floor, spilling its entire contents. We had to delay the lecture fifteen minutes while I struggled to reorganize the slides. It wasn't the student's fault. It was the fault of the elegant projector. With only one button to control the slide advance, how could one switch from forward to reverse? Neither of us could figure out how to make the control work.

All during the lecture the slides would sometimes go forward, sometimes backward. Afterward, we found the local technician, who explained it to us. A brief push of the button and the slide would go

Taste (7) für Diawechsel am Gerät

Diawechsel vorwärts = kurz drücken,

Diawechsel rückwärts = länger drücken.

Button (7) for changing the slides

Slide change forward = short press,

Slide change backward = longer press.

1.3 Leitz Pravodit Slide Projector. I finally tracked down the instruction manual for that projector. A photograph of the projector has its parts numbered. The button for changing slides is number 7. The button itself has no labels. Who could discover this operation without the aid of the manual? Here is the entire text related to the button, in the original German and in my English translation:

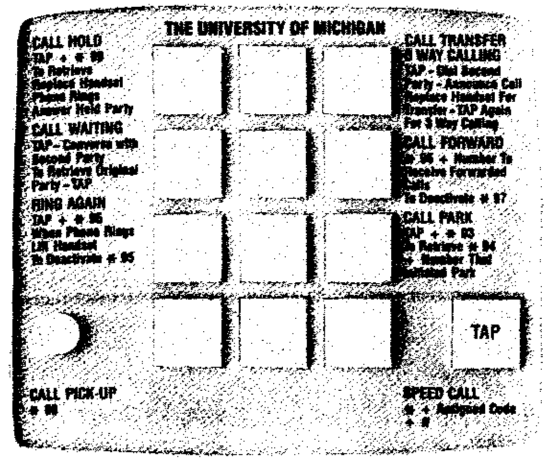
forward, a long push and it would reverse. (Pity the conscientious student who kept pushing it hard—and long—to make sure that the switch was making contact.) What an elegant design. Why, it managed to do two functions with only one button! But how was a first-time user of the projector to know this?

As another example, consider the beautiful Amphithéâtre Louis-Laird in the Paris Sorbonne, which is filled with magnificent paintings of great figures in French intellectual history. (The mural on the ceiling shows lots of naked women floating about a man who is valiantly trying to read a book. The painting is right side up only for the lecturer—it is upside down for all the people in the audience.) The room is a delight to lecture in, at least until you ask for the projection screen to be lowered. "Ah," says the professor in charge, who gestures to the technician, who runs out of the room, up a short flight of stairs, and out of sight behind a solid wall. The screen comes down and stops. "No, no," shouts the professor, "a little bit more." The screen comes down again, this time too much. "No, no, no!" the professor jumps up and down and gestures wildly. It's a lovely room, with lovely paintings. But why can't the person who is trying to lower or raise the screen see what he is doing?

New telephone systems have proven to be another excellent example of incomprehensible design. No matter where I travel, I can count upon finding a particularly bad example.

When I visited Basic Books, the publishers of this book, I noticed a new telephone system. I asked people how they liked it. The question unleashed a torrent of abuse. "It doesn't have a hold function," one woman complained bitterly—the same complaint people at my university made about their rather different system. In older days, business phones always had a button labeled "hold." You could push the button and hang up the phone without losing the call on your line. Then you could talk to a colleague, or pick up another telephone call, or even pick up the call at another phone with the same telephone number. A light on the hold button indicated when the function was in use. It was an invaluable tool for business. Why didn't the new phones at Basic Books or in my university have a hold function, if it is so essential? Well, they did, even the very instrument the woman was complaining about. But there was no easy way to discover the fact, nor to learn how to use it.

I was visiting the University of Michigan and I asked about the new



1.4 **Plate Mounted Over the Dial of the Telephones at the University of Michigan.** These inadequate instructions are all that most users see. (The button labeled "TAP" at the lower right is used to transfer or pick up calls—it is pressed whenever the instruction plate says "TAP." The light on the lower left comes on whenever the telephone rings.)

system there. "Yech!" was the response, "and it doesn't even have a hold function!" Here we go again. What is going on? The answer is simple: first, look at the instructions for hold. At the University of Michigan the phone company provided a little plate that fits over the keypad and reminds users of the functions and how to use them. I carefully unhooked one of the plates from the telephone and made a photocopy (figure 1.4). Can you understand how to use it? I can't. There is a "call hold" operation, but it doesn't make sense to me, not for the application that I just described.

The telephone hold situation illustrates a number of different problems. One of them is simply poor instructions, especially a failure to relate the new functions to the similarly named functions that people already know about. Second, and more serious, is the lack of visibility of the operation of the system. The new telephones, for all their added sophistication, lack both the hold button and the flashing light of the old ones. The hold is signified by an arbitrary action: dialing an arbitrary sequence of digits (*8, or *99, or what have you: it varies from one phone system to another). Third, there is no visible outcome of the operation.

Devices in the home have developed some related problems: functions and more functions, controls and more controls. I do not think that simple home appliances—stoves, washing machines, audio and television sets—should look like Hollywood's idea of a spaceship control room. They already do, much to the consternation of the consumer who, often as not, has lost (or cannot understand) the instruction manual, so—faced with the bewildering array of controls and displays—simply memorizes one or two fixed settings to approximate what is desired. The whole purpose of the design is lost.

In England I visited a home with a fancy new Italian washer-drier combination, with super-duper multi-symbol controls, all to do everything you ever wanted to do with the washing and drying of clothes. The husband (an engineering psychologist) said he refused to go near it. The wife (a physician) said she had simply memorized one setting and tried to ignore the rest.

Someone went to a lot of trouble to create that design. I read the instruction manual. That machine took into account everything about today's wide variety of synthetic and natural fabrics. The designers worked hard; they really cared. But obviously they had never thought of trying it out, or of watching anyone use it.

If the design was so bad, if the controls were so unusable, why did the couple purchase it? If people keep buying poorly designed products, manufacturers and designers will think they are doing the right thing and continue as usual.

The user needs help. Just the right things have to be visible: to indicate what parts operate and how, to indicate how the user is to interact with the device. Visibility indicates the mapping between intended actions and actual operations. Visibility indicates crucial distinctions—so that you can tell salt and pepper shakers apart, for example. And visibility of the effects of the operations tells you if the lights have turned on properly, if the projection screen has lowered to the correct height, or if the refrigerator temperature is adjusted correctly. It is lack of visibility that makes so many computer-controlled devices so difficult to operate. And it is an excess of visibility that makes the gadget-ridden, feature-laden modern audio set or video cassette recorder (VCR) so intimidating.

The Psychology of Everyday Things

This book is about the psychology of everyday things. POET emphasizes the understanding of everyday things, things with knobs and dials, controls and switches, lights and meters. The instances we have just examined demonstrate several principles, including the importance

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.

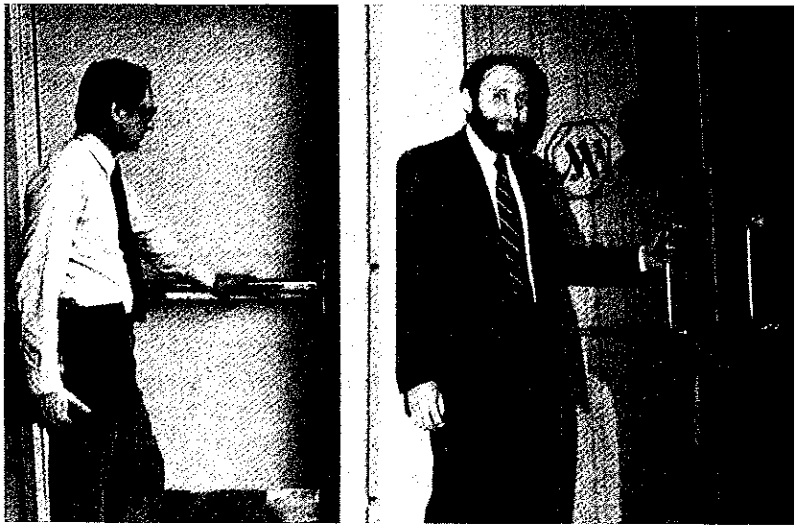
AFFORDANCES

"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]

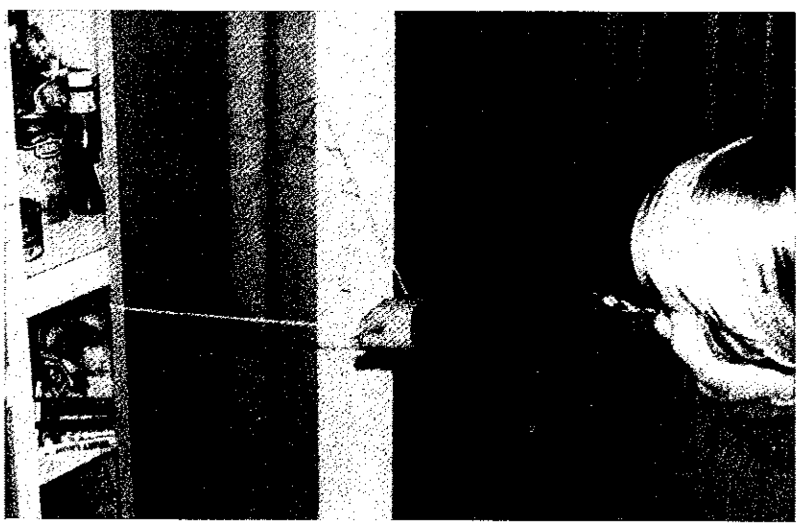
Affordances provide strong clues to the operations of things. Plates are for pushing. Knobs are for turning. Slots are for inserting things into. Balls are for throwing or bouncing. When affordances are taken advantage of, the user knows what to do just by looking: no picture, label, or instruction is 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.

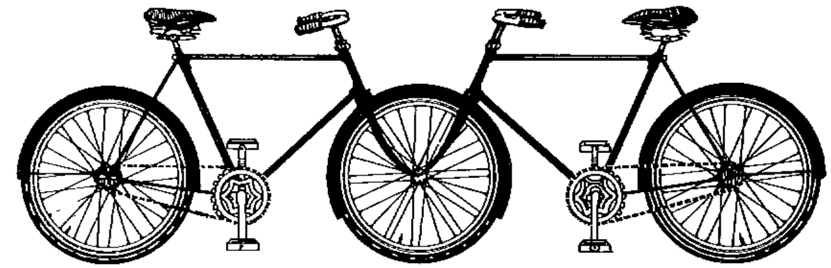
The book *What's What: A Visual Glossary of the Physical World* has more than fifteen hundred drawings and pictures and illustrates twenty-three thousand items or parts of items.[4] Irving Biederman, a psychologist who studies visual perception, estimates that there are probably "30,000 readily discriminable 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.

ONCEPTUAL 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.

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

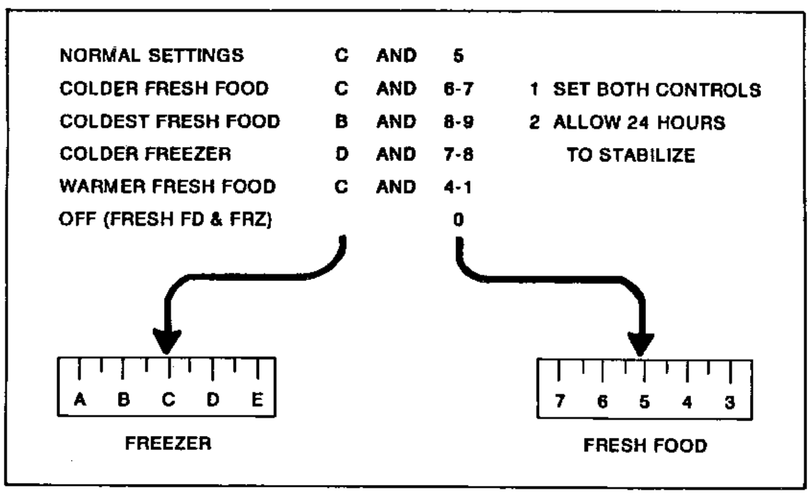
A good conceptual model allows us to predict the effects of our actions. Without a good model we operate by rote, blindly; we do operations as we were told to do them; we can't fully appreciate why, what effects to expect, or what to do if things go wrong. As long as things work properly, we can manage.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

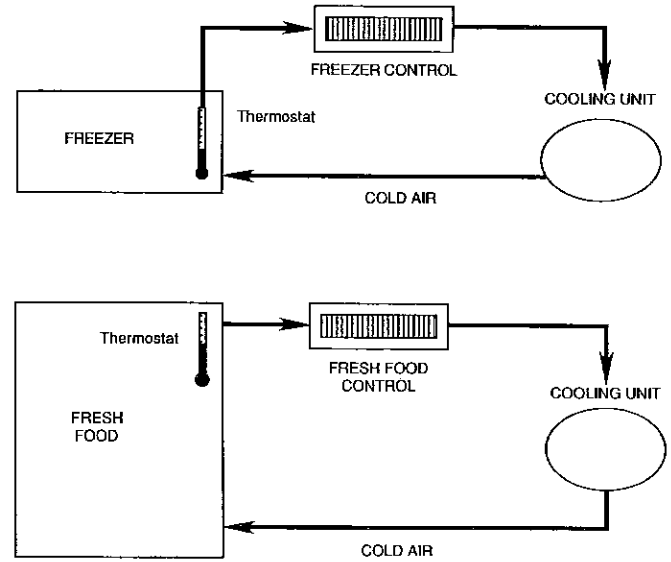
For everyday things, conceptual models need not be very complex. After all, scissors, pens, and light switches are pretty simple devices. There is no need to understand the underlying physics or chemistry of each device we own, simply the relationship between the controls and the outcomes. When the model 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?

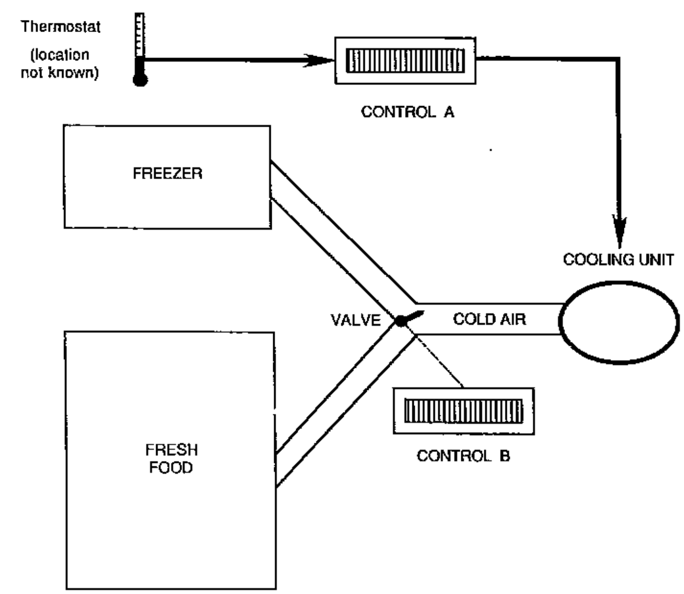
You try it. Figure 1.8 shows the instruction plate from inside the refrigerator. Now, suppose the freezer is too cold, the fresh food section just right. You want to make the freezer warmer, keeping the fresh food constant. Go on, read the instructions, figure them out.

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.

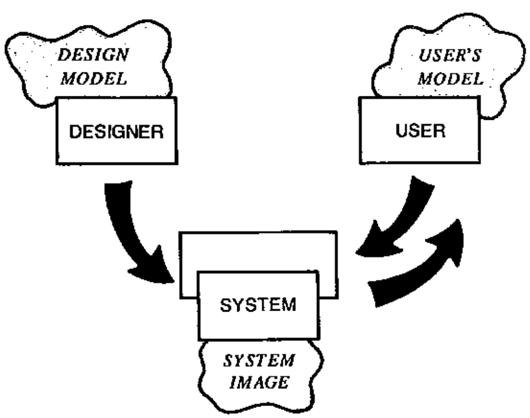


Oh, perhaps I'd better warn you. The two controls are not independent. The freezer control affects the fresh food temperature, and the fresh food control affects the freezer. And don't forget to wait twenty-four hours to check on whether you made the right adjustment, if you can remember what you did.

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?

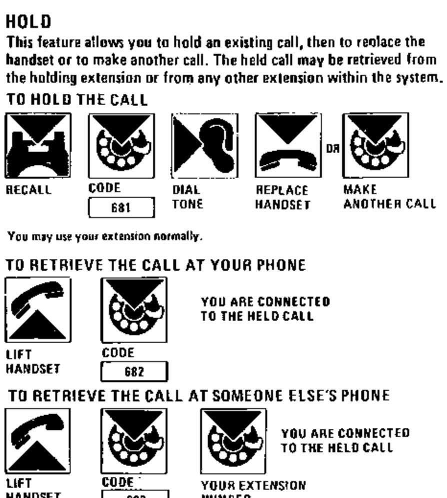
The modern telephone did not happen by accident: it was carefully designed. Someone—more likely a team of people—invented a list of features thought desirable, invented what seemed to them to be plausible ways of controlling the features, and then put it all together. My university, focusing on cost 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.

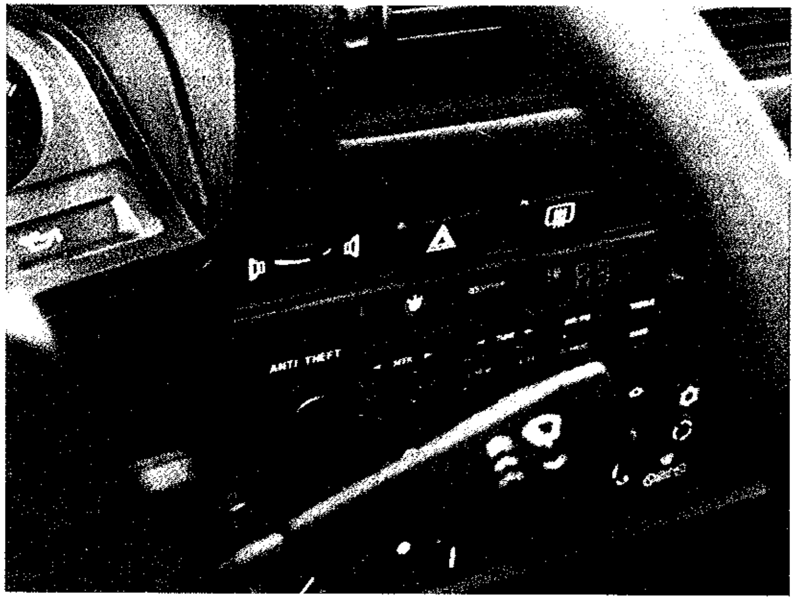
"How can you ever learn all those controls?" I asked the driver (with the aid of a German-speaking colleague). The driver was clearly puzzled by the question.

"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 3½-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 countdown 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.

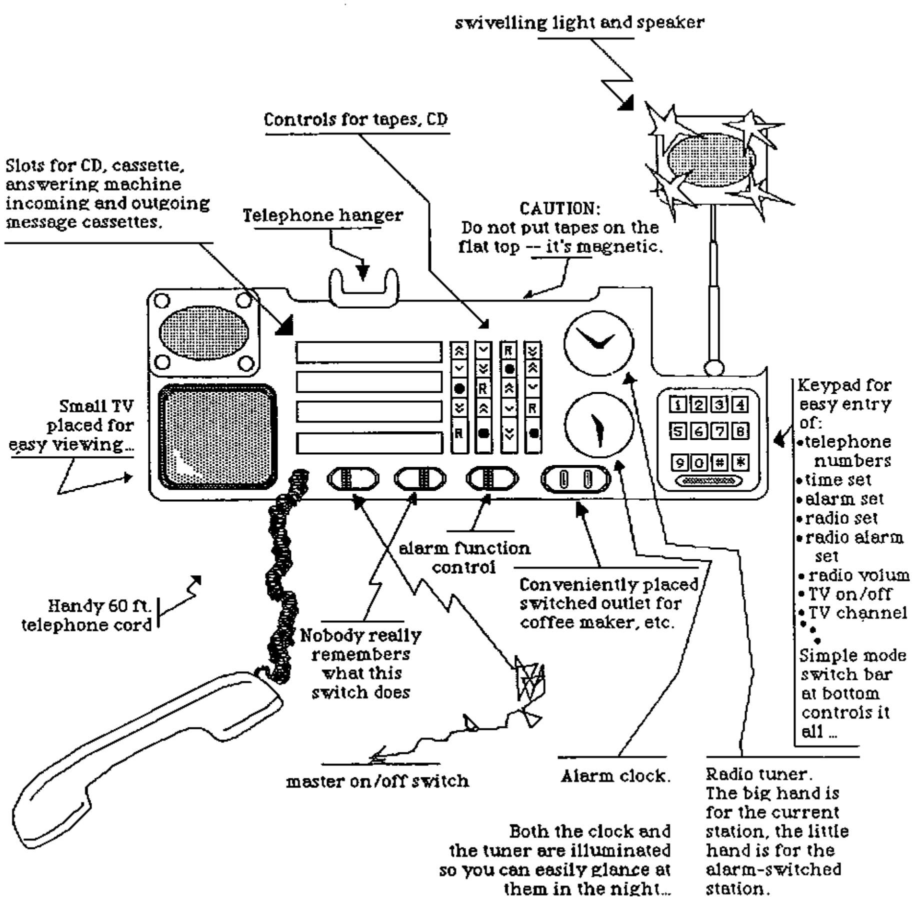
Your job is (A) to recommend what to build, then (B) to design the control panel, and finally (C) to certify that it is actually both what customers want and easy to use.

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.

CHAPTER TWO

THE PSYCHOLOGY OF EVERYDAY ACTIONS

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."

Falsely Blaming Yourself

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.[1] 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 naive 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

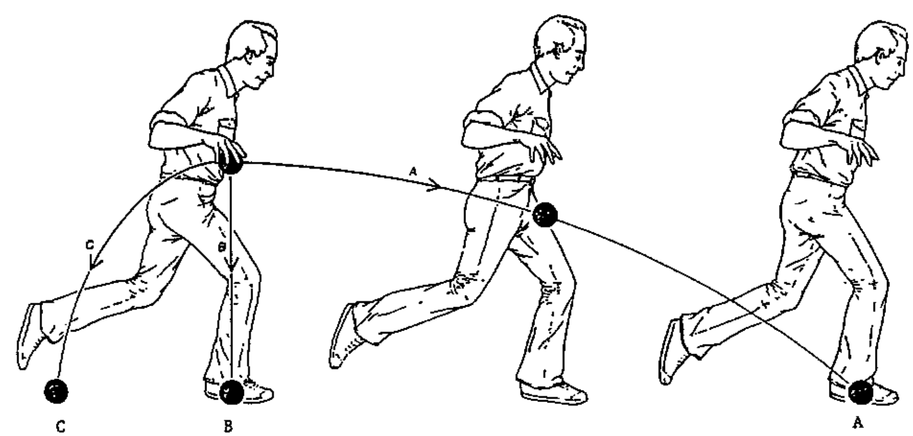
For example, Aristotle thought that moving objects kept moving only if something kept pushing them. Today's physicist says nonsense: a moving object continues to move unless some force is exerted to stop it. This is Newton's first law of motion, and it contributed to the development of moderr physics. 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?[2]

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 McCloskey. Copyright © 1983 by *Scientific American,* Inc. All rights reserved.)



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 unsettable 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 operators 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 flightcrew was logical, and was what most pilots probably would have done if confronted by the same situation."[6]

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

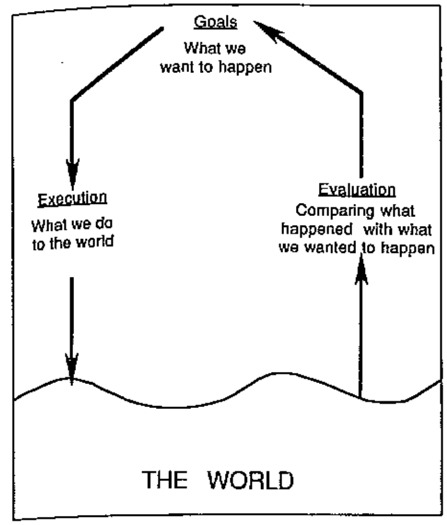
I am in Italy, at a conference. I watch the next speaker attempt to thread a film onto a projector that he has never used before. He puts the reel into place, then takes it off and reverses it. Another person comes to help. Jointly they thread the film through the projector and hold the free end, discussing 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.*

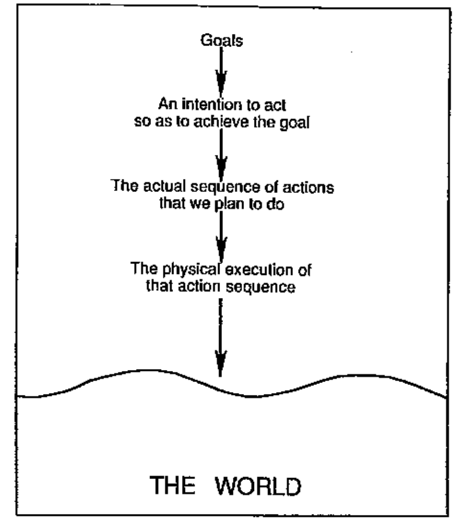
What makes something—like threading the projector—difficult to do? To answer this question, the central one of this book, we need to know what happens when someone does something. We need to examine the structure of an action.

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).

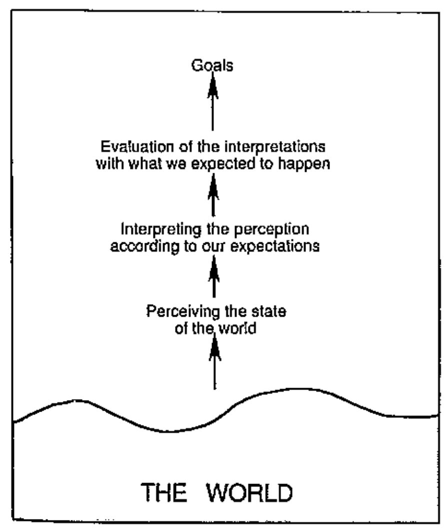
Real tasks are not quite so simple. The original goal may be imprecisely specified—perhaps "get something to eat," "get to work," "get dressed," "watch television." Goals do not state precisely what to do—where and how to move, what to pick up. To lead to actions goals must be transformed into specific 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.

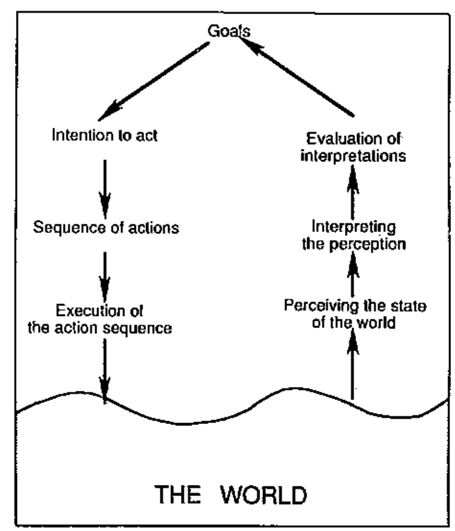
Suppose I am sitting in my armchair, reading a book. It is dusk, and the light has gotten dimmer and dimmer. I decide I need more light (that is the goal: get more light). My goal has to be translated into the intention that states the appropriate action in the world: push the switch button on the lamp. 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





2.2 The Action Cycle (above left). Human action has two aspects, execution and evaluation. Execution involves doing something. Evaluation is the comparison of what happened in the world with what we wanted to happen (our goal). **2.3 Stages of Execution** (above right). Start at the top with the *goal,* the state that is to be achieved. The goal is translated into an *intention* to do some action. The intention must be translated into a set of internal commands, an *action sequence* that can be performed to satisfy the intention. The action sequence is still a mental event: nothing happens until it is *executed,* performed upon the world. **2.4 Stages of Evaluation** (below left). Evaluation starts with our *perception* of the world. This perception must then be *interpreted* according to our expectations and then compared *(evaluated)* with respect to both our intentions (from figure 2.3) and our goals. **2.5 Seven Stages of Action** (below right).
The stages of execution from figure 2.3 (intentions, action sequence, and execution) are coupled with the stages of evaluation from figure 2.4 (perception, interpretation, and evaluation), with goals common to both stages.





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.[7]

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.*

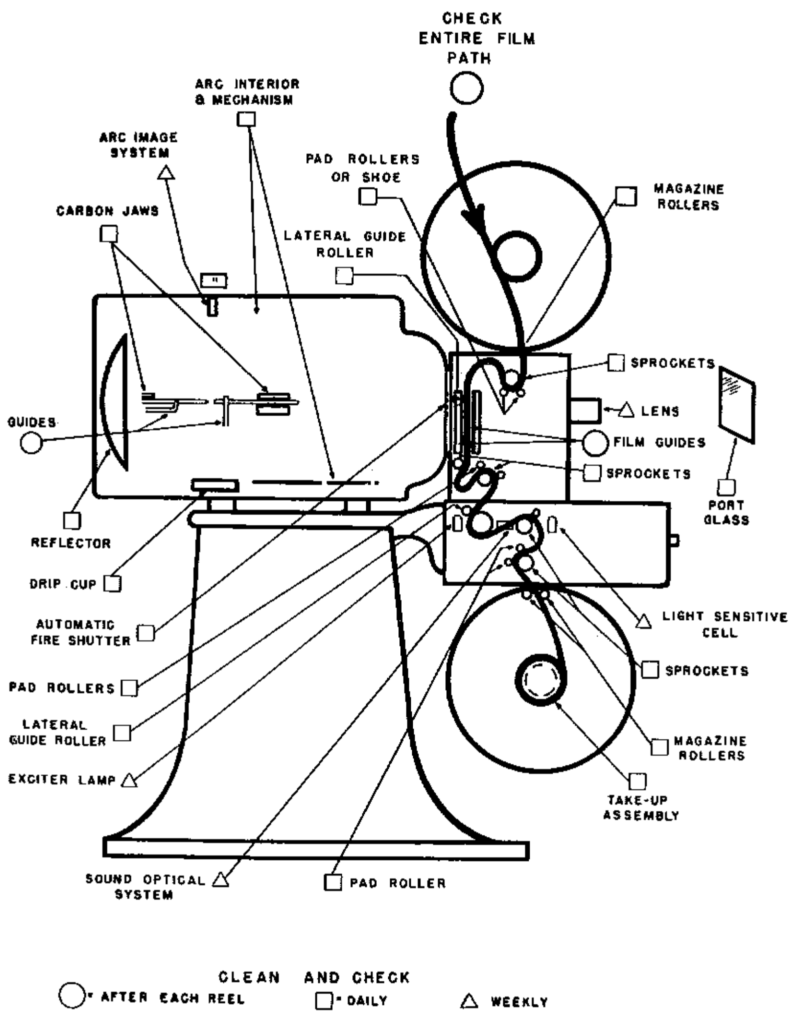
The seven-stage process of action can be started at any point. People do not always behave as full, logical, reasoning organisms, starting with high-level goals and working to achieve them. Our goals are often ill-formed and vague. We may respond to the events of the world (in what is called data-driven 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 everyday tasks, the difficulty resides entirely in deriving the relationships between the mental intentions and interpretations and the physical actions and states. There are several *gulfs* 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 com-

2.6 Threading the Movie Projector. The dark line at the right shows the path of the film. This picture doesn't tell the whole story, for the several loops of film have to be threaded just right, neither too loose nor too taut. (From *Projectionist's manual,* Department of the Army and the Air Force, May 1966.)



ponents and states of the environment. And these gulfs present major problems for users.[8]

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 gulfs 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 Gulfs 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.

How Easily Can One:

Determine The Function of the Device?

Tell What Actions Are Possible?	Tell if System is in Desired State?
Determine Mapping from Intention to Physical Movement?	Determine Mapping from System State to Interpretation?
Perform the Action?	Tell What State the System is In?

2.7 Using the Seven Stages to Ask Design Questions

- *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.

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]

CHAPTER FOUR

KNOWING WHAT TO DO

"Q. I read a news item about a new videotape-only player and rejoiced when the writer took a healthy swipe at the incomprehensible instructions 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 unfathomable machine and baffled by senseless instructions.

"Is there anyone, anywhere who will translate OR give a short course in VCR at play school level?"[1]

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.

Problems occur whenever there is more than one possibility. If there is only one part that can be operated and only one possible action to do, there will be no difficulty. Of course, if the designer has been too clever, hiding all the visible clues, the user may believe there are no alternatives and not even know how to begin.

When we encounter a novel object, how can we tell what to do with it? Either we have dealt with something similar in the past and transfer old knowledge to the new object, or we obtain instruction. In these cases, the information we need is in the head. Another approach is to use information in the world, particularly if the design of the new object has presented us with information that can be interpreted.

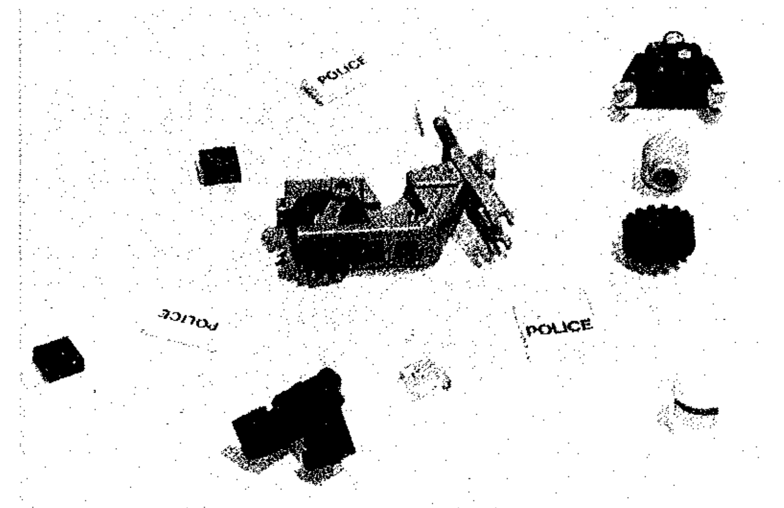
How can design signal the appropriate actions? To answer the question we build upon the principles discussed in chapter 3. One important set of signals comes through the natural constraints of objects, physical constraints that limit what can be done. Another set of signals comes from the affordances of objects, which convey messages about their possible uses, actions, and functions. A flat plate affords pushing, an empty container affords filling, and so on. Affordances can signal how an object can be moved, what it will support, and whether anything will fit into its crevices, over it, or under it. Where do we grab it, which parts move, and which parts are fixed? Affordances suggest the range of possibilities, constraints limit the number of alternatives. The thoughtful use of affordances and constraints together in design lets a user determine readily the proper course of action, even in a novel situation.

A Classification of Everyday Constraints

To understand the operation of constraints better, I did some simple experiments. I asked people to put things together from the parts given them; they had never seen the finished structure, and they were not even told what they should be constructing.[2] Let me illustrate with one of the examples: building a motorcycle from a Lego set (a children's construction toy).

The Lego motorcycle (figure 4.1) is a simple toy constructed of thirteen parts, some rather specialized. Of the thirteen parts, only two are alike—rectangles with the word *police* on them. One other piece is a blank rectangle of the same size. Three other pieces match in size and shape but are different colors. So there are two sets of three pieces in

4.1 **Lego Motorcycle.** The toy is shown assembled and in pieces. The thirteen parts are so cleverly constructed that even an adult can put them together. The design exploits constraints to specify just which pieces fit where. Physical constraints limit alternative placements. Semantic and cultural constraints provide the necessary clues for further decisions. For example, semantic constraints stop the user from putting the head backward on the body and cultural constraints dictate the placement of the three lights (the small rectangles, which are red, blue, and yellow).



which any of the three pieces are interchangeable, except for the semantic or cultural interpretation of the resulting construction. It turns out that the appropriate role for every single piece of the motorcycle is unambiguously determined by a set of physical, semantic, and cultural constraints. This means that people could construct the motorcycle without any instructions or assistance, although they had never seen it assembled. In this case, construction is entirely natural, if the builder knows about motorcycles and about the cultural assumptions that serve to constrain the placement of parts.

Affordances of the pieces were important in determining just how they fit together. The cylinders and holes characteristic of Lego suggested the major construction rule. The sizes and shapes of the parts suggested their operation. Physical constraints limited what parts would fit together. Other types of constraints also operated; all in all there were four different classes of constraints—physical, semantic, cultural, and logical. These classes are apparently universal, appearing in a wide variety of situations, and sufficient.

PHYSICAL CONSTRAINTS

Physical limitations constrain possible operations. Thus, a large peg cannot fit into a small hole. The motorcycle windshield would fit in only one place, with only one orientation. The value of physical constraints is that they rely upon properties of the physical world for their operation; no special training is necessary. With the proper use of physical constraints there should be only a limited number of possible actions—or, at least, desired actions can be made obvious, usually by being especially salient.

Physical constraints are made more effective and useful if they are easy to see and interpret, for then the set of actions is restricted before anything has been done. Otherwise, the physical constraint prevents the wrong action from succeeding only after it has been tried. The Lego windshield was sometimes tried in the wrong orientation first; the design could have made the correct position more visible. The everyday door key can be inserted into a vertical slot only if the key is held vertically. But this still leaves two possible orientations. A well-designed key will either work in both orientations or provide a clear physical signal for the correct one. Good automobile door keys are made so that orientation doesn't matter. A poorly designed car key can

be yet another of those minor frustrations of everyday life—not so minor, perhaps, when you're standing outside the car in a storm with both arms full of packages.

SEMANTIC CONSTRAINTS

Semantic constraints rely upon the meaning of the situation to control the set of possible actions. In the case of the motorcycle, there is only one meaningful location for the rider, who must sit facing forward. The purpose of the windshield is to protect the rider's face, so it must be in front of the rider. Semantic constraints rely upon our knowledge of the situation and of the world. Such knowledge can be a powerful and important clue.

CULTURAL CONSTRAINTS

Some constraints rely upon accepted cultural conventions, even if they do not affect the physical or semantic operation of the device. One cultural convention is that signs are meant to be read; for the motorcycle, the pieces with the word *police* on them have to be placed right side up. Cultural constraints determine the locations of the three lights, which are otherwise physically interchangeable. Red is the culturally defined standard for a stop light, which is placed in the rear. White or yellow (in Europe) is the standard color for headlights, which go in front. And a police vehicle often has a blue flashing light on top.

Each culture has a set of allowable actions for social situations. Thus, we know how to behave in a restaurant, even one we have never been to before. This is how we manage to cope when our host leaves us alone in that strange room, at that strange party, with those strange people. And this is why we sometimes feel frustrated, so incapable of action, when we are confronted with a restaurant or group of people from an unfamiliar culture, where our normally accepted behavior is clearly inappropriate and frowned upon. Cultural issues are at the root of of many of the problems we have with new machines: there are as yet no accepted conventions or customs for dealing with them.

Those of us who study these things believe that guidelines for cultural behavior are represented in the mind by means of schemas, knowledge structures that contain the general rules and information necessary for interpreting situations and for guiding behavior. In some stereotypical situations (for example, in a restaurant), the schemas may be very specialized. Cognitive scientists Roger Schank and Bob Abelson have proposed that in these cases we follow "scripts" that can guide the sequence of behavior. The sociologist Ervin Goffman calls the social constraints on acceptable behavior frames, and he shows how they govern behavior even when a person is in a novel situation or novel culture. Danger awaits those who deliberately violate the frames for a culture.[3]

Next time you are in an elevator, stand facing the rear. Look at the strangers in the elevator and smile. Or scowl. Or say hello. Or say, "Are you feeling well? You don't look well." Walk up to random passersby and give them some money. Say something like, "You make me feel good, so here is some money." In a bus or streetcar, give your seat to the next athletic-looking teenager you see. The act is especially effective if you are elderly, or pregnant, or disabled.

LOGICAL CONSTRAINTS

In the case of the motorcycle, logic dictated that all the pieces should be used, with no gaps in the final product. The three lights of the Lego motorcycle presented a special problem for many people. They could use the cultural constraint to figure out that the red was the stop light and should go in the rear, that the yellow was the headlight and should go in the front, but what about the blue? Many people had no cultural or semantic information that would help them place the blue light. For them, logic provided the answer: only one piece left, only one possible place to go. The blue light was logically constrained.

Natural mappings work by providing logical constraints. There are no physical or cultural principles here; rather there is a logical relationship between the spatial or functional layout of components and the things that they affect or are affected by. If two switches control two lights, the left switch should work the left light, the right switch the right light. If the lights are mounted one way and the switches another, the natural mapping is destroyed. If two indicators reflect the state of two different parts of a system, the location and operation of the indicators should have a natural relationship to the spatial or functional layout of the system. Alas, natural mappings are not often exploited.

Applying Affordances and Constraints to Everyday Objects

The characteristics of affordances and constraints can be applied to the design of everyday objects, much simplifying our encounters with them. Doors and switches present interesting examples, for poor design causes unnecessary problems for their users. Yet the common problems have simple solutions, which properly exploit affordances and natural constraints.

THE PROBLEM WITH DOORS

In chapter 1 we encountered the sad story of my friend who was trapped between sets of glass doors at a post office, trapped because there were no clues to the doors' operation. When we approach a door, we have to find both the side that opens and the part to be manipulated; in other words, we need to figure out what to do and where to do it. We expect to find some visible signal for the correct operation: a plate, an extension, a hollow, an indentation—something that allows the hand to touch, grasp, turn, or fit into. This tells us where to act. The next step is to figure out how: we must determine what operations are permitted, in part using the affordances, in part guided by constraints.

Doors come in amazing variety. Some open only if a button is pushed, and some don't appear to open at all, having neither buttons, nor hardware, nor any other sign of their operation. The door might be operated with a foot pedal. Or maybe it is voice operated, and we must speak the magic phrase. *("Open Simsim!")* In addition, some doors have signs on them: pull, push, slide, lift, ring bell, insert card, type password, smile, rotate, bow, dance, or, perhaps, just ask. Somehow, when a device as simple as a door has to come with an instruction manual—even a one-word manual—then it is a failure, poorly designed.

Appearances deceive. I have seen people trip and fall when they attempted to push open a door that worked automatically, the door opening inward just as they attempted to push against it. On most subway trains, the doors open automatically at each station. Not so in Paris. I watched someone on the Paris Métro try to get off the train and fail. When the train came to his station, he got up and stood patiently *in front of the door, waiting for it to open. It never opened. The train simply started up again and went on to the next station. In the Métro, you have to open the doors yourself by pushing a button, or depressing a lever, or sliding them (depending upon which kind of car you happen to be on).*

Consider the hardware for an unlocked door. It need not have any moving parts: it can be a fixed knob, plate, handle, or groove. Not only will the proper hardware operate the door smoothly, but it will also indicate just how the door is to be operated: it will exhibit the proper affordances. Suppose the door opens by being pushed. The easiest way to indicate this is to have a plate at the spot where the pushing should be done. A plate, if large enough for the hand, clearly and unambiguously marks the proper action. Moreover, the plate constrains the possible actions: there is little else that one can do with a plate except push. Unfortunately, even this simple clue is misused. Doors that should be pulled or slid sometimes have plates (figure 4.2). Doors that should be pushed sometimes have both plates and knobs or a handle and no plate.

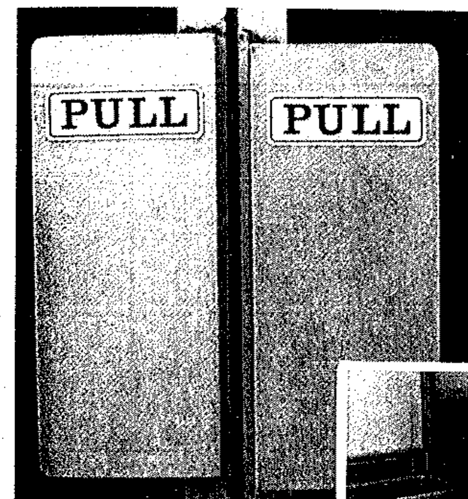
The violation of the simple use of constraints on doors can have serious implications. Look at the door in figure 4.3 *A*: this fire exit door has a push bar, a good example of an unambiguous signal to push, and a good design (required by law in the United States) because it forces proper behavior when panicked people press against a door as they attempt to flee a fire. But look again. On which side should you push? There is no way of knowing. Add some paint to the part that is to be pushed, or fasten a plate over it (figure 4.3 *B*): these provide strong cultural signals to guide the action properly. Push bars offer strong physical constraints, simplifying the task of knowing what to do. The use of cultural constraints simplifies the task of figuring out where to do it.

Some hardware cries out to be pulled. Although anything that can be pulled can also be pushed, the proper design will use cultural constraints so that the signal to pull will dominate. But even this can be messed up. I have seen doors with a mixture of signals, one implying push, the other pull. I have watched people passing through the door of figure 4.3 (*A*). And they had trouble, even people who worked in the building and who therefore used the door several times every day.

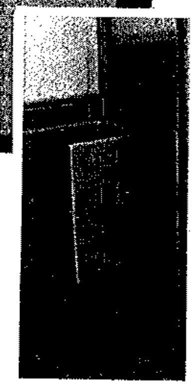
Sliding doors seem to present special difficulties. In fact, there are several good ways to signal the operation of a sliding door unambiguously. For example, a vertical slit in the door can be used in only one

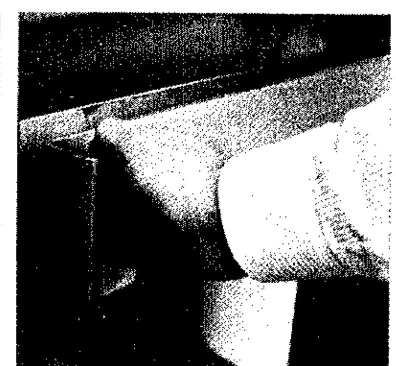
4.2 **The Design of Doors.** The doors at the left show two excellent examples of design: different handles, side by side on the same automobile, each neatly signaling its proper operation. The vertical placement of the lever on the handle to the left causes the hand to be held in a vertical plane, signifying a slide. The horizontal placement of the lever on the door handle to the right, coupled with the overhang and indentation that neatly afford entrance by the hand, signifies a pull. Two different types of doors, adjacent to one another, and yet there is no confusion between them.

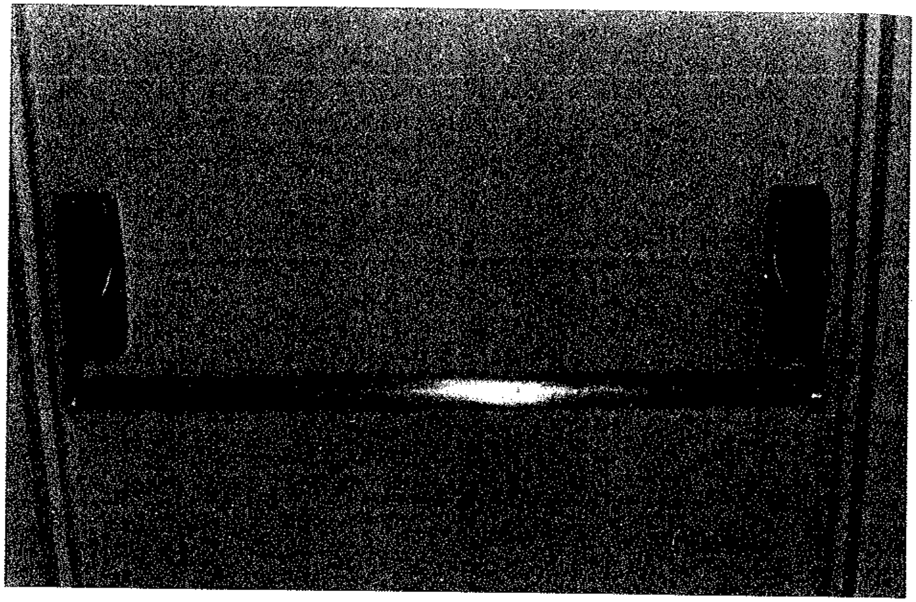
The handle depicted at the left shows inappropriate signals. This form of handle clearly marks grasp, twist, or pull—except that this particular door slides: a classic case of inappropriate design.



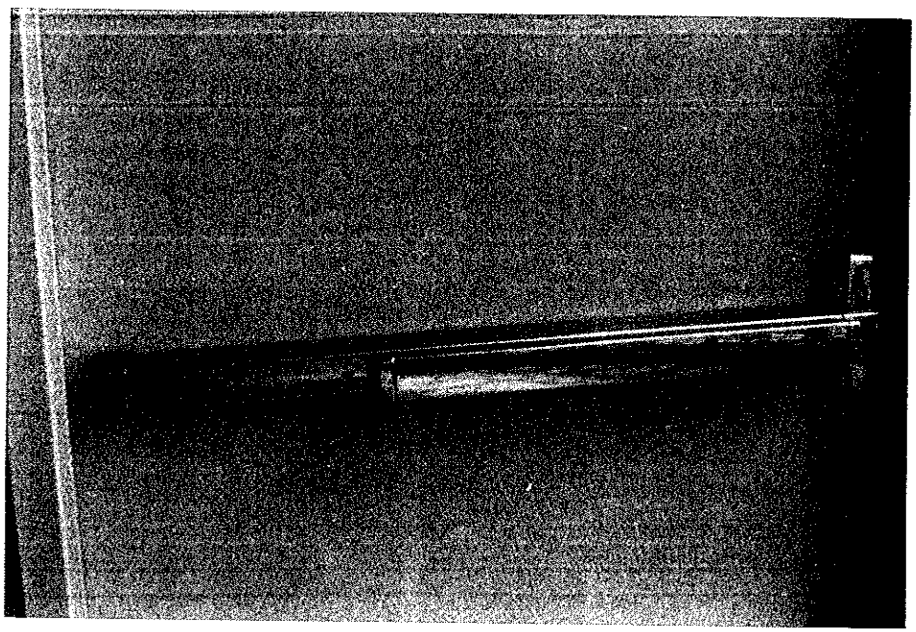
At left and below are photographs of hardware for doors that open by being pulled. The large plates at the left are a signal to push, but in fact the door is supposed to be pulled: no wonder the door needs the signs. The simple U-shaped brackets below is a much better design, but they are ambiguous enough that a sign still seems to be needed. Contrast with the two handles at the top, neither of which needs a sign yet is always operated properly. If a door handle needs a sign, then its design is probably faulty.







4.3 Doors in Two Commercial Buildings. Pushing the bar opens the door, but on which side do you push? Bar *A* (above) hides the signal, making it impossible to know on which side to push. A frustrating door. Bar *B* (below) has a flat plate mounted on the side that is to be pushed; this is a naturally interpreted signal. A nice design, no frustration for the user.



way: the fingers are inserted and the door slid. The location of the slit specifies not only where to exert the force but also in which direction. The critical signal is any depression in the door large enough for the fingers to fit into, but without an overhang. Similarly, any projection will also work, as long as it neither has an overhang nor is appropriate for being grasped with the hand. On a properly designed door, the fingers can exert pressure along the sides of the depression or projection—needed for sliding—but they can't pull or twist. I have seen elegant sliding doors, aesthetically pleasing, yet with clear signals to the user—in a conference room in Italy, on a door on a Métro train in Paris, on some Scandinavian furniture. Yet more often, it seems, sliding doors are built with the wrong signals, with clumsy hardware in positions that jam the fingers. Sliding doors somehow challenge the designer to get them wrong.

Some doors have appropriate hardware, well placed. The outside door handles of most modern automobiles are excellent examples of design. The handles are often recessed receptacles that simultaneously indicate the place and mode of action: the receptacle cannot be used except by inserting the fingers and pulling. Horizontal slits guide the hand into a pulling position; vertical slits signal a sliding motion. Strangely enough, the inside door handles for automobiles tell a different story. Here, the designer has faced a different kind of problem, and the appropriate solution has not yet been found. As a result, although the outside door handles of cars are often excellent, the inside ones are often difficult to find, hard to figure out how to operate, and difficult to use.

Unfortunately, the worst door hardware is found where we spend most of our time: at home and in the office. In many cases, the choice of hardware appears haphazard, used for convenience (or profitability). Architects and interior designers seem to prefer designs that are visually elegant and win prizes. This often means that a door and its hardware are designed to merge with the interior: the door may barely be visible, the hardware merges with door, and the operation is completely obscure. From my experience, the worst offenders are cabinet doors. It is sometimes not even possible to determine where the doors are, let alone whether and from where they are slid, lifted, pushed, or pulled. The focus on aesthetics may blind the designer (and the purchaser) to the lack of usability.

A particularly frustrating design is that of the door that opens outward by being pushed inward. The push releases the catch and ener-

gizes a spring, so that when the hand is taken away the door springs open. It's a very clever design, but most puzzling to the first-time user. A plate would be the appropriate signal, but designers sometimes do not wish to mar the smooth surface of the door. I have such a latch in the glass door of the cabinet in which I store phonograph records. You can see through the door, and it is obvious that there is no room for the door to open inward; to push on the door seems contradictory. New and infrequent users of this door usually reject pushing and open it instead by pulling, which often requires them to use fingernails, knife blades, or more ingenious methods to pry it open.

THE PROBLEM WITH SWITCHES

At any lecture I give, my first demonstration needs no preparation. I can count on the light switches of the room or auditorium to be unmanageable. "Lights please," someone will say. Then fumble, fumble, fumble. Who knows where the switches are and which lights they control? The lights seem to work smoothly only when a technician is hired to sit in a control room somewhere, turning them on and off.

The switch problems in an auditorium are annoying, but similar problems in airplanes and nuclear power plants are dangerous. The controls all look the same. How do the operators avoid the occasional mistake, confusion, or accidental bumping against the wrong control? Or misaim? They don't. Fortunately, airplanes and power plants are pretty robust. A few errors every hour are not important—usually.

One type of popular small airplane has identical-looking switches for flaps and landing gear right next to one another. You might be surprised to learn how many pilots, while on the ground, have decided to raise the flaps and instead raised the wheels. This very expensive error happened frequently enough that the National Transportation Safety Board wrote a report about it. The analysts politely pointed out that the proper design principles to avoid these errors have been known for thirty years. Why were those design errors still being made?

Basic switches and controls should be relatively simple to design well. But there are two fundamental difficulties. The first is the grouping problem, how to determine which switch goes with which function.

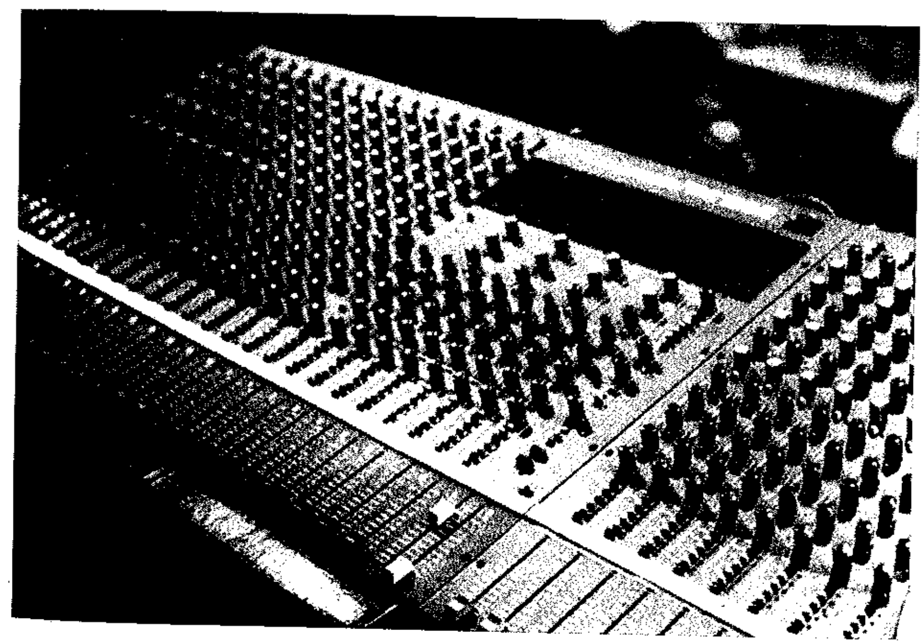
The second is the mapping problem. For example, when there are many lights and an array of switches, how can you determine which switch controls which light?

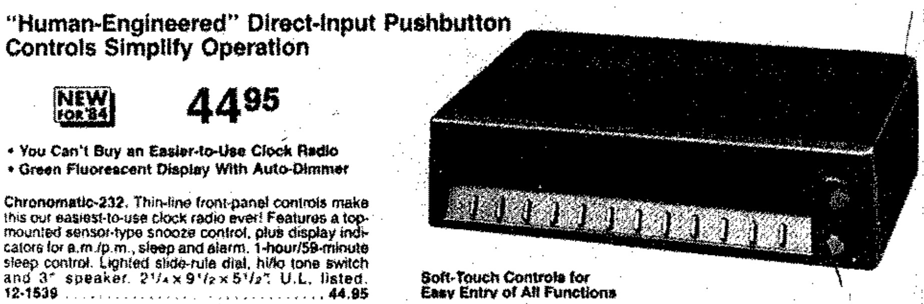
The switch problem becomes serious only where there are many of them. It isn't a problem in situations with one switch, and it is only a minor problem where there are two switches. But the difficulties mount rapidly with more than two switches at the same location. Multiple switches are more likely to occur in offices, auditoriums, and industrial locations than in homes (figure 4.4).

WHICH SWITCH CONTROLS WHICH FUNCTION?

Switches for unrelated functions are often placed together, usually with no distinguishing marks to help the user know which switch controls which function. Designers love rows of identical-looking switches. The switches look good, are easy to mount, are inexpensive to build, and please the aesthetic sensibilities of the viewer. But they

4.4 Typical Audio Mixing Control. This picture was taken in an auditorium in England. Fortunately, errors on panels like these are seldom serious, often not even noted.





4.5 A Clock Radio, "Human Engineered" to Simplify Operation. Note the row of identical-looking switches. (Copyright Tandy Corporation. Used with permission.)

make it easy to err. With identical switches all in a row, it is difficult to distinguish the switch for the coffee maker from the switch to the central power for the computer. Or the set-the-time switch from the turn-off-the-radio switch (figure 4.5). Or the landing gear switch from the flap control switch.

Consider my car radio: twenty-five controls, many apparently arbitrary. All tiny (so that they will fit the limited space available). Imagine trying to use the radio while driving at high speed, at night. Or in winter when wearing gloves, so that the attempt to push one button succeeds in pushing two, or the attempt to turn the loudness control also adjusts the tone control. You should be able to use things in the dark. A car radio should be usable with a minimum of visual cues. But the radio designers probably designed it in the laboratory, with little or no thought about the car, or the driver. For all I know the design won a prize for its visual aesthetics.

It should go without saying that controls that cause trouble should not be located where they can be operated by accident, especially in the dark, or when the person is trying to use the device without looking. It should go without saying, but in fact, it is necessary to say it.

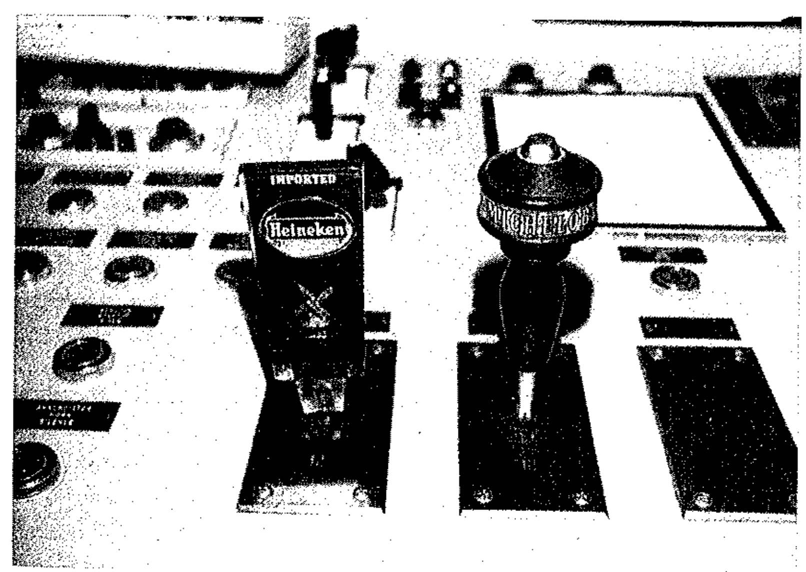
There is a simple, well-known solution to the grouping problem: set the switches for one set of functions apart from the switches that control other functions. Another solution is to use different types of switches. The solutions can be combined. To solve the problem with the airplane flap and landing gear switches, separate the switches and don't line them up in a row. Also use shape coding: a tire-shaped switch

can control the landing gear, and the flap switch can be a long, thin rectangle—the shape of a flap. Putting controls in different locations makes it less likely that a misaimed hand will throw the wrong switch. And using shape coding means that a potential error may be caught and that the correct switch can be found by feel alone (figure 4.6). That's how to solve this first problem, now let us turn to the other one.

HOW ARE THE SWITCHES ARRANGED?

With the lights in a room, you know that all the switches control lights. But which switch controls which light? Room lights are usually organized in a two-dimensional structure and they are usually horizontal (that is, they are on the ceiling or, if they are lamps, they are placed along the floor or on tables). But switches are usually arranged in a one-dimensional row mounted on the wall, a vertical surface. How can a one-dimensional row of switches map onto a two-dimensional array of lights? And with the switches being mounted on the wall and the

4.6 Make the Controls Look and Feel Different. The control-room operators in a nuclear power plant tried to overcome the problem of similar-looking knobs by placing beer-keg handles over them. This is good design, even if after the fact; the operators should be rewarded. (From Seminara, Gonzales, & Parsons, 1977. Photograph courtesy of Joseph L. Seminara.)



lights being on the ceiling, you have to do a mental rotation of the switches to get them to conform to the lights. The mapping problem is unsolvable with the current structure of switches.

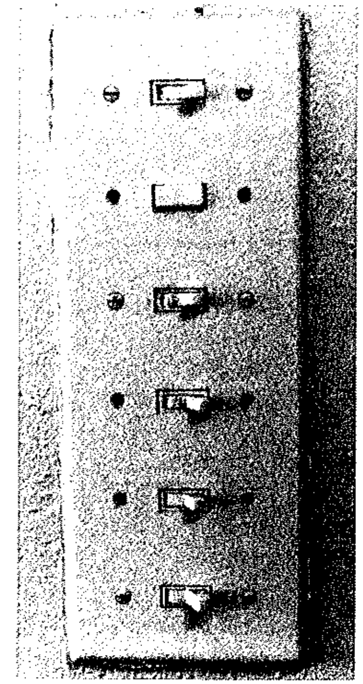
Electricians usually try to lay out the switches in the same order as the lights they control, but the mismatch in the spatial arrangement of the lights and the switches makes it difficult, if not impossible, to produce a full natural mapping. Electricians have to use standard components, and the designers and manufacturers of those standard components worried only about fitting the proper number of switches into them safely. Nobody thought about how the lights were to be arranged or how the switches ought to be laid out.

My house was designed by two brash young architects, award winning, who, among other things, liked neat rows of light switches. We got a horizontal row of four identical switches in the front hall, a vertical column of six identical switches in the living room. "You will get used to it," the architects assured us when we complained. We never did. Finally we had to change the switches, making each one different. Even so we made lots of mistakes.

In my psychology laboratory, the lights and their switches were located in many different places, yet most people wanted to control the lights upon entering the area. The area is large, with three major hallways and approximately fifteen rooms. Moreover, this floor of the building has no windows, so it is dark unless the lights are turned on.

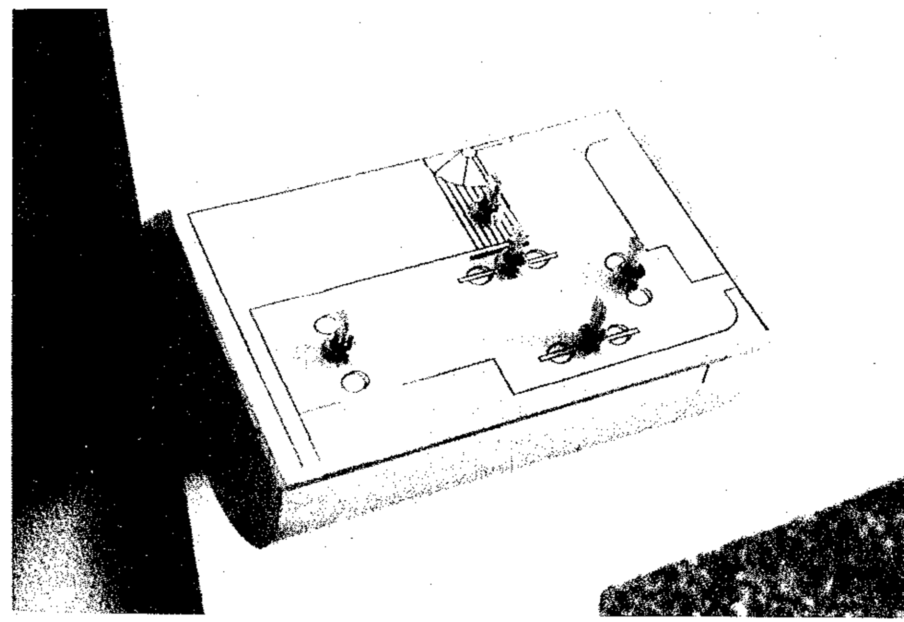
If light switches are placed on the wall, there is no way they can exactly correspond in position to the placement of the lights. Why place the switches flat against the wall? Why not redo things? Why not place the switches horizontally, in exact analogy to the things being controlled, with a two-dimensional layout so that the switches can be placed on a floorplan of the building in exact correspondence to the areas that they control? Match the layout of the lights with the layout of the switches: the principle of natural mapping. In my laboratory, as in my home, the solution was to construct a simple switchplate that mirrored the physical arrangement of the area, with small light switches placed in relevant locations.[4] Figure 4.7 shows the situation at my home, and figure 4.8 shows what we did at the laboratory.

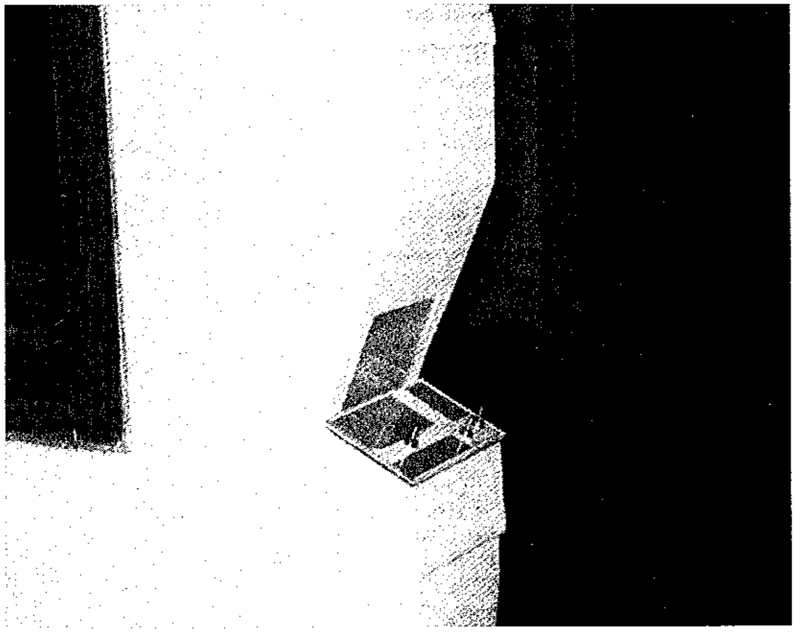
How well do the new switch arrangements work? Quite well, I am happy to report. One laboratory user sent me the following note:



4.7 The vertical array of six switches at the right is what our architects provided to control the lights in our odd-shaped living room. We could never remember which switch did what.

The photograph below shows our solution: switches arranged to match the room layout. (One more switch, for a projection screen, will be mounted on the vertical plate just above the light switches. The switch panel was constructed for the author by David Wargo.)





4.8 The original layout of switches in my laboratory had the light switches scattered. We put all the switches in one convenient location, arranged on a floor plan of the laboratory. (The switch panel was constructed by David Wargo.)

"You know, I actually kind of like those new switches now—they seem easy to use, and it's nice to have all the switches in one location when you first walk in. You can just sort of swipe at them on your way past and light up the area you want—very quick. So while I was worried they wouldn't be advantageous for the experienced user, I was wrong."

Can the new switches be used everywhere? Probably not. But there is no reason they couldn't be widely adopted. There are a series of technical problems still to be addressed: builders and electricians need standardized components. How about making up standard light switch boxes, made to be mounted *on* the wall (instead of *in* the wall as they are today), where the switches are mounted on the top of the box, on the horizontal surface. And on the top, make up a matrix of supports so that there can be free, relatively unrestricted placement of the switches in whatever pattern best suits the room. Use smaller switches

if necessary. Maybe get rid of those standardized light plates. The matrix design would require drilling holes differently for each room, but if the switches were designed to fit into standard sized circular or rectangular holes, the holes could be drilled or punched quite easily.

My suggestion requires that the switch box stick out from the wall, whereas today's boxes are mounted so that the switches are flush with the wall. Some might consider my solution ugly. Well, then, indent the boxes, placing them *in* the wall. After all, if there is room inside the wall for the existing switch boxes, there is also room for an indented horizontal surface. Or mount the switches on a little pedestal, or on a ledge.

Visibility and Feedback

So far we have concentrated upon constraints and mappings. But for knowing what to do there are other relevant principles, too, especially visibility and feedback:

1. *Visibility.* Make relevant parts visible.
2. *Feedback.* Give each action an immediate and obvious effect.

When we use a novel object, a number of questions guide our actions:

- Which parts move; which are fixed?
- Where should the object be grasped? What part is to be manipulated? What is to be held? Where is the hand to be inserted? If it is speech sensitive, where does one talk?
- What kind of movement is possible: pushing, pulling, turning, rotating, touching, stroking?
- What are the relevant physical characteristics of the movements? With how great a force must the object be manipulated? How far can it be expected to move? How can success be gauged?
- What parts of the object are supporting surfaces? How much size and weight will the object support?

The same kinds of questions arise whether we are trying to decide what to do or attempting to evaluate the results of an action. In examining the object, we have to decide which parts signify the state of the object and which are solely decorative, or nonfunctional, or part of the background or supports. What things change? What has changed over the previous state? Where should we be watching or listening to detect any changes? The important things to watch should be visible and clearly marked; the results of any action should be immediately apparent.

MAKING VISIBLE THE INVISIBLE

The principle of visibility is violated over and over again in everyday things. In numerous designs crucial parts are carefully hidden away. Handles on cabinets distract from some design aesthetics, and so they are deliberately made invisible or left out. The cracks that signify the existence of a door can also distract from the pure lines of the design, so these significant cues are also minimized or eliminated. The result can be a smooth expanse of gleaming material, with no sign of doors or drawers, let alone of how those doors and drawers might be operated. Electric switches are often hidden: many electric typewriters have the on/off switch hidden underneath; many computers and computer terminals have the on/off switch in the rear, difficult to find and awkward to use;[5] and the switches that control kitchen garbage disposal units are often hidden away, sometimes nearly impossible to find.

Many systems are vastly improved by the act of making visible what was invisible before. Consider the VCR.

> *"UMPTEEN-DAY-UMPTEEN-EVENT PROGRAMMING. Because time-shifting is so popular, manufacturers and retailers play up a VCR's ability to record automatically. The typical VCR can record four events (video jargon for programs) over a 4-day span. . . .*
>
> *"It's one thing to know that a VCR can record eight events in 14 days. It's quite another to make the machine behave. You have to go through a tedious series of steps to tell the VCR when to start recording, what channel to record, how long to run the tape, and so on.*
>
> *"Some VCR's are much easier to program than others. . . . Best of all, we think, is a feature called on-screen programming. Commands that appear on the TV screen help you enter the time, date, and channel of the program you want to tape."*[6]

As the quotation from *Consumer Reports* indicates, the act of setting up these units to do the recording is horribly complex and difficult. The same article later warns that if you are not careful in your selection,

"you could wind up with a VCR that brings out fear and loathing whenever you try to change the channel resets or set it up to record a program when you are away." It does not take much examination to discover the reason for the difficulties: there is no visual feedback. As a result, users (1) have trouble remembering their place in the lengthy sequence of required steps; (2) have trouble remembering what next needs to be done; and (3) cannot easily check the information just entered to see if it is what was intended, and then cannot easily change it, if they decide it is wrong.

The gulfs both in execution (the first two problems) and in evaluation (the last problem) are significant for these VCRs. Both can be bridged by the use of a display. Displays often cost money and take up room, which is why designers hesitate to use them, but in the case of a VCR, a display device is usually already available: the TV set. And, indeed, those VCRs that can be programmed through the use of an on-screen TV display are much easier to use. Visibility makes all the difference.

NOTHING SUCCEEDS LIKE A GOOD DISPLAY

Over and over again we find unwarranted complexity that could be avoided were the device to contain a good display. With the modern telephone (see chapter 1), a display that could prompt the user through the series of steps required for programming would make the difference between a valuable, usable system and a next-to-useless one. So, too, with any device of complexity, whether it be the washing machine, microwave oven, or office copying machine. Nothing succeeds like visual feedback, which in turn requires a good visual display.

WHAT CAN BE DONE?

New technologies, especially the inexpensive microprocessors available today (the heart of the computer) make possible the incorporation of powerful and intelligent systems even in simple, everyday things, from toys to kitchen appliances to office machines. But new capabilities must be accompanied by appropriate displays, also now relatively inexpensive. I asked the students in one of my classes to generate some possibilities for adding visibility to everyday devices. Here are some of them:

- *Display the song titles for compact discs.* Why not take advantage of the storage capacity of an audio compact disc (CD) and have it display not only the number of the song or track (as it now does) but also the title? Each title could be accompanied by other information, such as performers, composer, or playing time. Thus, in programming the CD, you could select by name rather than by number, and you would always know what you were hearing.

- *Display the names of television programs.* If each television station would also broadcast its station identification and the title of the current program, the viewer who tuned in during the middle of a show could easily find out what it was. The information could be sent in computer-readable format during the retrace interval (the time that the beam is off the screen).

- *Print the cooking information for foods on the food package in computer-readable form.* This is a scheme for bypassing the need to make things visible. The cooking of frozen foods often requires several different cooking times, waiting times, and heat settings. The programming is complex. If the cooking information were on the package in machine-readable form, one could put the food in the microwave oven, pass a scanner over the printed information, and let the oven program itself.

USING SOUND FOR VISIBILITY

Sometimes things can't be made visible. Enter sound: sound can provide information available in no other way. Sound can tell us that things are working properly or that they need maintenance or repair. It can even save us from accidents. Consider the information provided by:

- The click when the bolt on a door slides home
- The "zzz" sound when a zipper works properly
- The "tinny" sound when a door doesn't shut right
- The roaring sound when a car muffler gets a hole
- The rattle when things aren't secured
- The whistle of a tea kettle when the water boils
- The click when the toast pops up
- The increase in pitch when a vacuum cleaner gets clogged
- The indescribable change in sound when a complex piece of machinery starts to have problems

Many devices do use sound, but only for signals. Simple sounds, such as buzzers, bells, or tones. Computers use bleeping, whining, and clicking sounds. This use of sound is valuable and serves an important function, but it is very limited in power; it is as if the use of visual cues were limited to different colored, flashing lights. We could use sound for much more communciation than we do.

These days computers produce several sounds, and keypads, microwave ovens, and telephones beep and burp. These are not naturalistic sounds; they do not convey hidden information. When used properly, a beep can assure you that you've pressed a button, but the sound is as annoying as informative. Sounds should be generated so as to give information about the source. They should convey something about the actions that are taking place, actions that matter to the user but that would otherwise not be visible. The buzzes, clicks, and hums that you hear while a telephone call is being completed are one good example: take out those noises and you are less certain that the connection is being made.

Bill Gaver, who has been studying use of sound in my laboratory, points out that real, natural sound is as essential as visual information because sound tells us about things we can't see, and it does so while our eyes are occupied elsewhere. Natural sounds reflect the complex interaction of natural objects: the way one part moves against another; the material of which the parts are made—hollow or solid, metal or wood, soft or hard, rough or smooth. Sounds are generated when materials interact, and the sound tells us whether they are hitting, sliding, breaking, tearing, crumbling, or bouncing. Moreover, sounds differ according to the characteristics of the objects, according to their size, solidity, mass, tension, and material. And they differ with how fast things are going and how far away from us they are.

If they are to be useful, sounds must be generated intelligently, with an understanding of the natural relationship between the sounds and the information to be conveyed. Sounds on artificial devices should be as useful as sounds in the real world. Gaver has proposed that sound could play an important role in computer-based applications. Here, rich, naturalistic sounds could serve as auditory icons, caricatures of naturally occurring sounds that could provide information about the concepts being represented not easily conveyed in other ways.[7]

You have to be very careful with sound, however. It easily becomes cute rather than useful. It can annoy and distract as easily as it can aid. One of the virtues of sounds is that they can be detected even when attention is applied elsewhere. But this virtue is also a deficit, for sounds are often intrusive. Sounds are difficult to keep private unless the intensity is low or earphones are used. This means both that neighbors may be annoyed and that others can monitor your activities. The use of sound to convey information is a powerful and important idea, but still in its infancy.

Just as the presence of sound can serve a useful role in providing feedback about events, the absence of sound can lead to the same kinds of difficulties we have already encountered from a lack of feedback. The absence of sound can mean an absence of information, and if feedback from an action is expected to come from sound, silence can lead to problems.

I once stayed in the guest apartment of a technological institute in the Netherlands. The building was newly completed, with many interesting architectural features. The architect had gone to great lengths to keep the noise level low; the ventilation system could not be heard. In similar fashion, the ventilation for the room came and went through invisible slots in the ceiling (so I am told; I never did find them).

All was fine until I took a shower. The bathroom seemed to have no ventilation at all, so everything became wet, then eventually cold and clammy. There was a switch in the bathroom that I thought might be the control for an exhaust fan. When I pushed the switch, a light on it came on and stayed on. Further pushing had no effect.

I noticed that whenever I returned to the apartment after an absence, the light would be off. So each time I entered the apartment, I went into the bathroom and pushed the button. By listening closely, I could hear a slight "thump" in the distance the first time the button was depressed. I decided it was some kind of signal. Perhaps it was a call button, summoning the maid, or the janitor, or maybe even the fire department (though no one showed up). I did also consider that it might control a ventilation system, but I could hear no flow of air. I examined the inside of the entire bathroom with care, trying to find an air inlet. I even got a chair and a flashlight and examined the ceiling. Nothing.

At the end of my stay, the person driving me to the airport, explained that the button controlled the exhaust fan. The fan was on as long as the light was on, and it turned off, automatically, in about five minutes. The architect was very good at disguising the ventilation system and at keeping the noise level down.

Here is a case where the architect was too successful: the feedback was clearly lacking. The light was not enough—in fact, it was quite misleading. Noise would have been welcome. It would have signaled that there really was ventilation.

CHAPTER FIVE

TO ERR IS HUMAN

"LONDON—An inexperienced computer-operator pressed the wrong key on a terminal in early December, causing chaos at the London Stock Exchange. The error at stockbrokers Greenwell Montagu led to systems staff working through the night in an attempt to cure the problem."[1]

People make errors routinely. Hardly a minute of a normal conversation can go by without a stumble, a repetition, a phrase stopped midway through to be discarded or redone. Human language provides special mechanisms that make corrections so automatic that the participants hardly take notice; indeed, they may be surprised when errors are pointed out. Artificial devices do not have the same tolerance. Push the wrong button, and chaos may result.

Errors come in several forms. Two fundamental categories are slips and mistakes. Slips result from automatic behavior, when subconscious actions that are intended to satisfy our goals get waylaid en route. Mistakes result from conscious deliberations. The same processes that make us creative and insightful by allowing us to see relationships between apparently unrelated things, that let us leap to correct conclusions on the basis of partial or even faulty evidence, also lead to error.

Proper design calls for a forcing function here. There are several viable schemes. The cover over the game pack compartment could control an interlock, so that it automatically turned off the power whenever it was opened. Or the power switch could move a lever blocking the top of the game pack compartment, so that the packs could not be removed or inserted unless the lever were out of the way, turning off the power. There are other possibilities. My point is, of course, that the design should have included one; without the forcing function, failure to heed the warning is almost guaranteed.

A Design Philosophy

There are lots of ways for a designer to deal with errors.[17] The critical thing, however, is to approach the topic with the proper philosophy. The designer shouldn't think of a simple dichotomy between errors and correct behavior; rather, the entire interaction should be treated as a cooperative endeavor between person and machine, one in which misconceptions can arise on either side. This philosophy is much easier to implement on something like a computer which has the ability to make decisions on its own than on things like doors and power plants, which do not have such intelligence. But the philosophy of user-centered system design still holds. Think of the user's point of view. Assume that every possible mishap will happen, so protect against it. Make actions reversible. Try to make them less costly. All the required principles have been thoroughly discussed in this book.

- Put the required knowledge in the world. Don't require all the knowledge to be in the head. Yet do allow for more efficient operation when the user has learned the operations, has gotten the knowledge in the head.
- Use the power of natural and artificial constraints: physical, logical, semantic, and cultural. Use forcing functions and natural mappings.
- Narrow the gulfs of execution and evaluation. Make things visible, both for execution and evaluation. On the execution side, make the options readily available. On the evaluation side, make the results of each action apparent. Make it possible to determine the system state readily, easily, and accurately, and in a form consistent with the person's goals, intentions, and expectations.

CHAPTER SIX

THE DESIGN CHALLENGE

They began work at once, and by the next September the first [typewriter] machine was finished, and letters were written with it. It worked successfully so far as to write rapidly and correctly, but trial and experience showed it to be far short of an acceptable, practicable writing machine. . . .

One device after another was conceived and developed till twenty-five or thirty experimental instruments were made, each succeeding one a little different from and a little better than the one preceding. They were put into the hands of stenographers, practical persons who were presumed to know better than anyone else what would be needed and satisfactory. Of these, James O. Clephane, of Washington, D.C., was one. He tried the instruments as no one else had tried them; he destroyed them, one after another, as fast as they could be made and sent him, till the patience of Mr. Sholes [the inventor] was exhausted. But Mr. Densmore insisted that this was the very salvation of the enterprise; that it showed the weak spots and defects, and that the machine must be made so that anybody could use it, or all efforts might as well be abandoned; that such a test was a blessing and not a misfortune, for which the enterprise should be thankful.[1]

The Natural Evolution of Design

Much good design evolves: the design is tested, problem areas are discovered and modified, and then it is continually retested and remodified until time, energy, and resources run out. This natural design process is characteristic of products built by craftspeople, especially folk objects. With handmade objects such as rugs, pottery, hand tools, or furniture, each new object can be modified slightly from the previous one, eliminating small problems, making small improvements, or testing new ideas. Over time, this process results in functional, aesthetically pleasing objects.

Improvements can take place through natural evolution as long as each previous design is studied and the craftsperson is willing to be flexible. The bad features have to be identified. The folk artists change the bad features and keep the good ones unchanged. If a change makes matters worse, well, it just gets changed again on the next go-around. Eventually the bad features get modified into good ones, while the good ones are kept. The technical term for this process is "hill-climbing," analogous to climbing a hill in the dark. Move your foot in one direction. If it is downhill, try another direction. If the direction is uphill, take one step. Keep doing this until you have reached a point where all steps would be downhill; then you are at the top of the hill—or at least at a local peak.[2]

'ORCES THAT WORK AGAINST :VOLUTIONARY DESIGN

Natural design does not work in every situation: there must be enough time for the process to be carried out, and the item must be simple. Modern designers are subject to many forces that do not allow for the slow, careful crafting of an object over decades and generations. Most of today's items are too complex, with too many variables, for this slow sifting of improvements. But simple improvements ought to be possible. You would think that objects such as automobiles, appliances, or computers, which periodically come out in new models, could benefit from the experience of the previous model. Alas, the multiple forces of a competitive market seem not to allow this.

One negative force is the demands of time: new models are already into their design process before the old ones have even been released to customers. Moreover, mechanisms for collecting and feeding back the experiences of customers seldom exist. Another force is the pressure to be distinctive, to stand out, to make each design look different from what has gone before. It is the rare organization that is content to let a good product stand or to let natural evolution perfect it slowly. No, each year a "new, improved" model must come out, usually incorporating new features that do not use the old as a starting point. In far too many instances, the results spell disaster for the consumer.

There is yet another problem: the curse of individuality. Designers have to make an individual stamp, their mark, their signature. And if different companies manufacture the same type of item, each must do it differently to allow its product to be distinguished from others'. A mixed curse, individuality, for through the desire to be different come some of our best ideas and innovations. But in the world of sales, if a company were to make the perfect product, any other company would have to change it—which would make it worse—in order to promote its own innovation, to show that it was different. How can natural design work under these circumstances? It can't.

Consider the telephone. The early telephone evolved slowly, over several generations. It once was a most awkward device, with handset and microphone, one held with each hand. You had to turn a crank to generate a signal that would ring the bell at the other end of the line. Voice transmission was poor. Over the years improvements were slowly made in size and shape, reliability, and features that simplified its use. The instrument was heavy and robust: drop it on the floor, and not only did it still work but you seldom lost the telephone connection. The layout of the dial or the push buttons resulted from careful experimentation in the laboratory. The size and spacing of the keys were carefully selected to work for a wide variety of the population, including the very young and the very old. The sounds of the telephone were also carefully designed to produce feedback. Push a button and you heard a tone in the earphone. Speak into the microphone, and a carefully determined percentage of your own voice was fed back into the earphone, the better to help you regulate how loudly you were talking. The clicks, buzzes, and other noises you heard while a connection was being established provided useful indications of progress.

All these minor aspects of the telephone were arrived at slowly, over years of development protected by the monopoly status of most national telephone systems. In today's wildly competitive market, there is a fierce desire to bring out a product that appeals to a wide body of people and that is distinctive and different—the market demands speed and novelty. Many of the most useful refinements are being lost. Push buttons are apt to be arranged haphazardly, with the keys oversize or tiny. The sounds have been taken away. Many telephones don't even give feedback when the buttons are pushed. All the folklore of design has been lost with the brash new engineers who can't wait to add yet the latest electronic gimmickry to the telephone, whether needed or not.

One simple detail can make the point: the ridge of plastic next to the switch hook—the button under the receiver that, when depressed, hangs up the call. Ever knock the telephone off the table and onto the floor while you were talking? Wasn't it nice when you didn't get disconnected, frustrating when you did? The monopolistic Bell System designers explicitly recognized this problem and designed with it in mind. They made the telephone heavy and sturdy enough to withstand the fall. And they protected the critical button with a shield that prevents the switch hook from hitting the ground. Look carefully at

6.1 Design Subtleties. In the older Bell System instrument, the prongs that held the receiver also prevented the switch hook from being accidentally depressed. More recent telephones often lack such niceties.

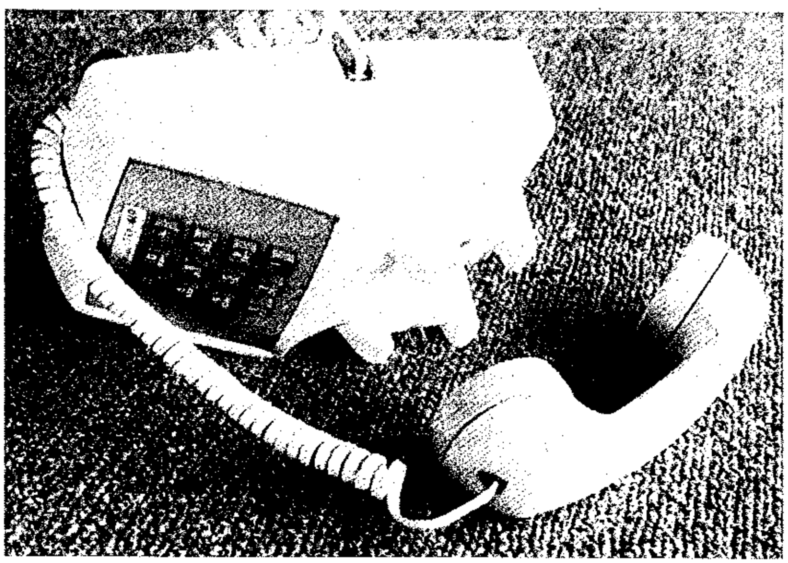


figure 6.1: see that on the one telephone the buttons cannot reach the ground and so are not depressed. A small feature, but an important one. Economic pressures have made the newer telephones lighter, less expensive, and less sturdy—throwaway phones, they are often called. And the protective shield? Often as not, there is none—in this case not because of cost, but because the new designers probably never thought of it, probably never realized its function. The result? This scenario, repeated in office over office.

Mark is sitting at his desk when the phone rings. "Hello," he answers. "Yeah, I can help you—let me get out the manual." He reaches, pulling the telephone with him. Bang! Crash! The phone falls on the floor, hanging itself up. "Damn," mutters Mark, "I don't even know who that was."

THE TYPEWRITER: A CASE HISTORY IN THE EVOLUTION OF DESIGN

> *"Among all the mechanical inventions for which the age is noted, none, perhaps, has more rapidly come into general use than the typewriter. . . . The time is coming when it will almost, or quite as much, supersede the steel pen as that has the good, gray goose quill."*[3]

The history of the typewriter is the story of dedicated inventors in many countries, each striving to develop a machine for rapid writing. They tried many versions in their struggle to get the one that fit all the constraints—that worked, could be manufactured at reasonable cost, and could be used.

Consider the typewriter keyboard, with its arbitrary, diagonally sloping arrangement of keys and its even more arbitrary arrangement of letters on the keys. The current standard keyboard was designed by Charles Latham Sholes in the 1870s. The design is called the "qwerty" keyboard (because in the American version the top row of letters begins with "qwerty"), or sometimes the Sholes keyboard. The Sholes typewriter was not the first, but it was the most successful of the early versions; it eventually became the Remington typewriter, the model upon which most manual typewriters were constructed. Why such a weird keyboard?

The design of the keyboard has a long and peculiar history. Early typewriters experimented with a wide variety of layouts, using three basic themes. One was circular, with the letters laid out alphabetically; the operator would find the proper spot and depress a lever, lift a rod, or do whatever other mechanical operation the device required. Another popular layout was like the piano keyboard, with the letters laid out in a long row; some of the early keyboards, including an early version by Sholes, even had black and white keys. Both the circular layout and the piano keyboard proved awkward. In the end, a third arrangement was adopted by all: a rectangular arrangement of keys, still in alphabetical order. The levers manipulated by the keys were large and ungainly, and the size, spacing, and arrangement of the keys were dictated by these mechanical considerations, not by the characteristics of the human hand.

Why did the alphabetical ordering change? To overcome a mechanical problem. When the typist went too quickly the typebars would collide, jamming the mechanism. The solution was to change the locations of the keys: letters such as *i* and *e* that were often typed in succession were placed on opposite sides of the machine so that their bars would not collide.[4] Other typewriting technologies did not follow the qwerty arrangement. Typesetting machines (such as the Linotype machine) use a completely different layout; the Linotype keyboard is called "shrdlu," after the pattern of keys it follows, and is modeled after the relative frequency of letters in English. This was how hand printers arranged the letters that they would remove from bins and insert manually into the printing forms. Ah, yes, the natural evolution of design.

Not all early keyboards had a backspace, and the "tabulation" key ("tab" on modern keyboards) was a revolutionary breakthrough. The first typewriters could print only upper case letters. The addition of lower case letters was, at first, accomplished by adding a new key for each lower case letter, so in effect there were two separate keyboards. Some early typewriters organized the keys for upper case differently than for lower case. Imagine how difficult it would be to learn that keyboard! It took years to develop the shift key so that both upper and lower case letters could share the same key. This was a nontrivial invention, combining mechanical ingenuity with a dual-faced typebar.

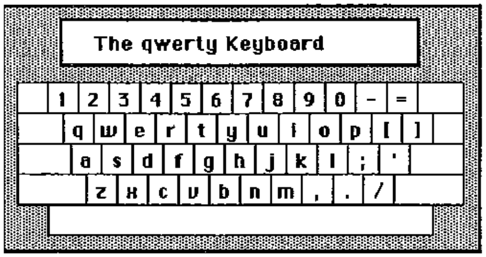
In the end, the keyboard was designed through an evolutionary process, but the main driving forces were mechanical. Modern keyboards do not have the same problems; jamming isn't a possibility with

electronic keyboards and computers. Even the style of typing has changed. In the early years, people kept their eyes on the keyboard and typed with one or two fingers of each hand. Then one courageous person, Frank McGurrin of Salt Lake City, memorized the key locations and learned to type with all his fingers, without looking at the keyboard. His skills were not recognized at first; it took a national contest held in Cincinnati, Ohio, in 1877 to prove that this method was indeed superior.[5] In the end, the qwerty keyboard was adopted throughout the world with but minor variations. We are committed to it, even though it was designed to satisfy constraints that no longer apply, was based on a style of typing no longer used, and is difficult to learn.

Tinkering with keyboard design is a popular pastime (figure 6.2). Some schemes keep the existing mechanical layout of the keys, but arrange the assignments of letters more efficiently. Others improve the physical layout as well, arranging the keys to accommodate the mirror-image symmetry of the hands and the varied spacing and agility of the fingers. Still others reduce the number of keys dramatically by having patterns of keys—chords—represent the letters, permitting one-handed or faster two-handed typing. But none of these innovations takes hold because the qwerty keyboard, while deficient, is good enough. Although its antijamming arrangement no longer has mechanical justification, it does put many common letter pairs on opposing hands; one hand can be getting ready to type its letter while the other is finishing, so typing is speeded up.

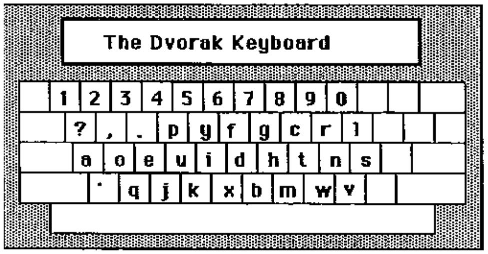
What about alphabetical keyboards (figure 6.3)? Wouldn't they at least these be easier to learn? Nope.[6] Because the letters have to be laid out in rows, just knowing the alphabet isn't enough. You also have to know where the rows break. Even if you could learn that, it would still be easier to scan the keyboard than to compute where a key might be. Then you are better off if common letters are located where you are apt to find them by scanning—a property that the qwerty keyboard provides. If you don't know any keyboard, there is little difference in typing speed among a qwerty keyboard, an alphabetic keyboard, and even a random arrangement of keys. If you know even a little of the qwerty, that is enough to make it better than the others. And for expert typists, the alphabetical arrangements are always *slower* than qwerty.

There is a better way—the Dvorak keyboard—painstakingly developed by (and named after) one of the founders of industrial engineering. It is easier to learn and allows for about 10 percent faster typing, but that is simply not enough of an improvement to merit a revolution

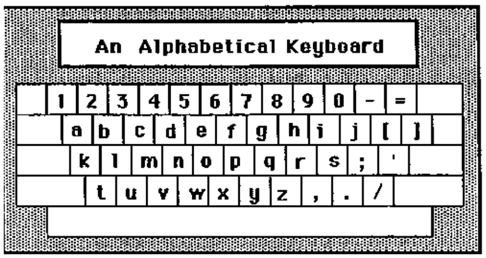


6.2 Typewriter Keyboards.

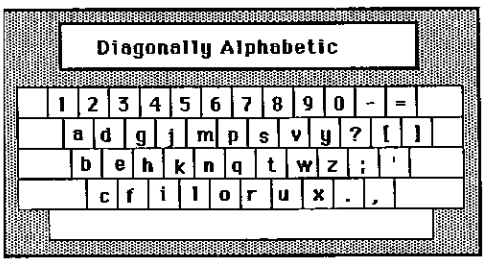
The standard American layout of keys —the Sholes or qwerty keyboard.



The American Simplified Keyboard (often called ASK), a simplified version of the original Dvorak keyboard; on the original, the numerals and punctuation keys are arranged differently.



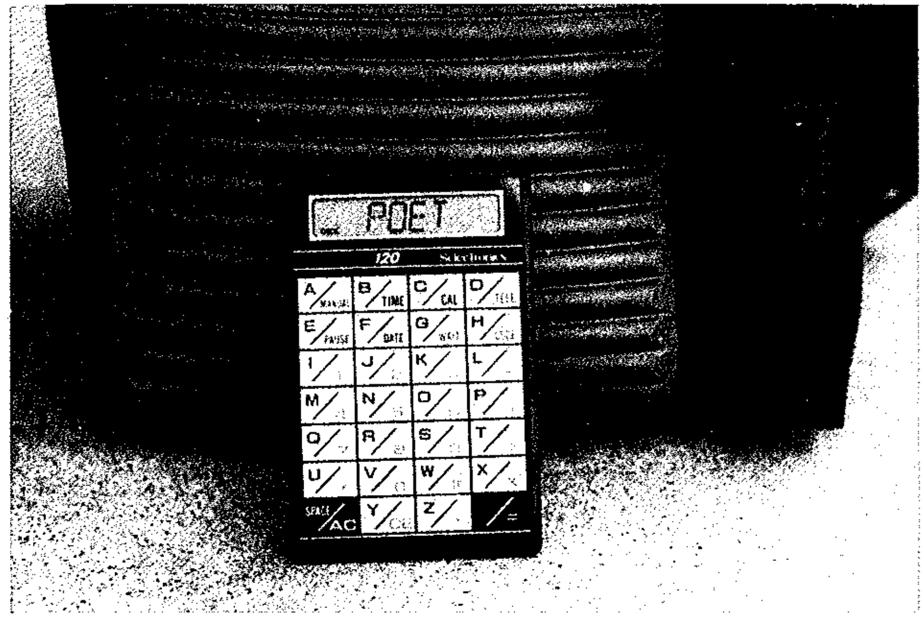
Most alphabetically organized keyboards arrange the alphabet along horizontal rows, as shown (and in the keyboards of figure 6.3).



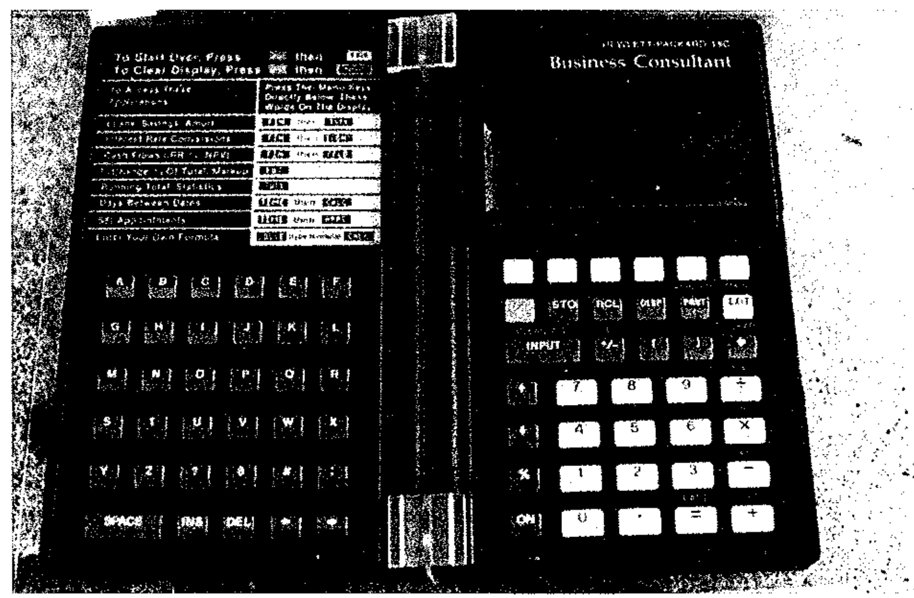
This alphabetical arrangement is superior, however: with its diagonal arrangement, letters increase systematically up the alphabet from left to right without major breaks.

The keyboard at left has randomly arranged letters.

Beginners succeed about the same on all these keyboards: alphabetical works barely better than random. For experts, ASK is best, followed by qwerty: alphabetical keyboards are quite inferior. Moral: Don't bother with alphabetical keyboards.



6.3 Products with Alphabetical Keyboards. Even though several experiments show that these are of no use to novices and detrimental to experts, every year designers plunge ahead and foist yet another alphabetical keyboard on us. Even if you manage to learn one, you will not have learned to use all the different ones.



in the keyboard. Millions of people would have to learn a new style of typing. Millions of typewriters would have to be changed. The severe constraints of existing practice prevent change, even where the change would be an improvement.[7]

Couldn't we at least do better with two hands at once? Yes, we could. Court stenographers can outtype anyone else. They use chord keyboards, typing syllables directly onto the page—syllables, not letters. Chord keyboards have very few keys—as few as five or six, but usually ten to fifteen. Many chord keyboards allow you to type single letters or whole words with one depression of the hand on several keys. If you use all ten fingers at the same time, then there are 1,023 possible combinations. That is enough for all the letters and numbers, lower case and upper case, plus a lot of words—if only you can learn the patterns. Chord keyboards have a horrible disadvantage: they are very hard to learn and very hard to retain; all the knowledge has to be in the head. Walk up to any regular keyboard and you can use it right away. Just search for the letter you want and push the key. With a chord keyboard, you have to press several keys simultaneously. There is no way to label the keys properly and no way to know what to do just by looking. Some chord keyboards are incredibly clever and remarkably easy to learn, considering. I tried to learn one of the easier ones. Thirty minutes' practice, and I knew the alphabet. But if I didn't use the keyboard for a week, I forgot the chords. The gain did not seem worth the effort. What about one-handed chord keyboards? Wouldn't it be worth a lot of time and effort to be able to type with one hand? Perhaps, if you are flying a jet aircraft with one hand and need to enter data into your computer with the other. But not for the rest of us.[8]

All this brings up an important lesson in design. Once a satisfactory product has been achieved, further change may be counterproductive, especially if the product is successful. You have to know when to stop.

You can observe the design iterations and experiments with the computer keyboard. The layout of the basic keyboard is now standardized through international agreement. But computer keyboards need extra keys, and these are not standardized. Some keyboards have an extra key between the shift key and the "z" key. The return key takes on different shapes and locations. The special keys of the computer keyboard—for example, control, escape, break, delete (not to be confused with backspace), and the "arrow" or cursor control keys—vary in location with the phases of the year, varying even among the products of a single manufacturer. Much confusion and strong emotions result.

Note, too, that the computer allows for flexible letter arrangements. It is a simple matter on some computers to switch the interpretation of the keys from qwerty to Dvorak: one command and the change is done. But unless the Dvorak fan also pries off and rearranges the keycaps, the Dvorak fan has to ignore the labels on the keys and rely on memory. Someday key labeling will be done by electronic displays on each key, so changing the labels will also become trivial. So computer technology may liberate users from forced standardization. Everyone could select the keyboard of personal choice.

Why Designers Go Astray

"[Frank Lloyd] Wright evidently wasn't very sympathetic about complaints. When Herbert F. Johnson, the late president of S. C. Johnson, Inc., in Racine, Wis., called Wright to say that his roof was leaking all over a dinner guest, the architect is said to have responded, 'Tell him to move his chair.'"[9]

If everyday design were ruled by aesthetics, life might be more pleasing to the eye but less comfortable; if ruled by usability, it might be more comfortable but uglier. If cost or ease of manufacture dominated, products might not be attractive, functional, or durable. Clearly, each consideration has its place. Trouble occurs when one dominates all the others.

Designers go astray for several reasons. First, the reward structure of the design community tends to put aesthetics first. Design collections feature prize-winning clocks that are unreadable, alarms that cannot easily be set, can openers that mystify. Second, designers are not typical users. They become so expert in using the object they have designed that they cannot believe that anyone else might have problems; only interaction and testing with actual users throughout the design process can forestall that. Third, designers must please their clients, and the clients may not be the users.

PUTTING AESTHETICS FIRST

"It probably won a prize" is a disparaging phrase in this book. Why? Because prizes tend to be given for some aspects of a design, to the neglect of all others—usually including usability. Consider the following example, in which a usable, livable design was penalized by the design profession. The assignment was to design the Seattle offices of the Federal Aviation Administration (FAA). The most noteworthy feature of the design process was that those who would work in the building had a major say in the planning. One of the members of the design team, Robert Sommer, describes the process as follows:

"Architect Sam Sloan coordinated a project in which employees . . . were able to select their own office furniture and plan office layout. This represented a major departure from prevailing practices in the federal services where such matters were decided by those in authority. Since both the Seattle and Los Angeles branches of the FAA were scheduled to move into new buildings at about the same time, the client for the project, the General Services Administration, agreed with architect Sloan's proposal to involve employees in the design process in Seattle, while leaving the Los Angeles office as a control condition where traditional methods of space planning would be followed."[10]

So there really were two designs: one in Seattle, with heavy participation by the users, and one in Los Angeles, designed in the conventional manner by architects. Which design do the users prefer? Why the Seattle one, of course. Which one got the award? Why the Los Angeles one, of course. Here is Sommer's description of the outcome:

"Several months following the move into the new buildings, surveys by the research team were made in Los Angeles and Seattle. The Seattle workers were more satisfied with their building and work areas than were the Los Angeles employees. . . . It is noteworthy that the Los Angeles building has been given repeated awards by the American Institute of Architects while the Seattle building received no recognition. One member of the AIA jury justified his denial of an award to the Seattle building on the basis of its 'residential quality' and 'lack of discipline and control of the interiors,' which was what the employees liked the most about it. This reflects the well-documented differences in preferences between architects and occupants. . . . The director of the Seattle office admitted that many visitors were surprised that this is a federal facility. Employees in both locations rated their satisfaction with their job performance before and after the move into the new building. There was no change in the Los Angeles office and a 7 percent improvement in rated job performance in the Seattle office."[11]

Aesthetics, not surprisingly, comes first at museums and design centers. I have spent much time in the science museum of my own city, San Diego, watching visitors try out the displays. The visitors try hard, and although they seem to enjoy themselves, it is quite clear that they usually miss the point of the display. The signs are highly decorative; but they are often poorly lit, difficult to read, and have lots of gushing language with little explanation. Certainly the visitors are not enlightened about science (which is supposed to be the point of the exhibit). Occasionally I help out when I see bewildered faces by explaining the scientific principles being demonstrated by the exhibit (after all, many of the exhibits in this sort of museum are really psychology demonstrations, many of which I explain in my own introductory classes). I am often rewarded with smiles and nods of understanding. I took one of my graduate classes there to observe and comment; we all agreed about the inadequacy of the signs, and, moreover, we had useful suggestions. We met with a museum official and tried to explain what was happening. He didn't understand. His problems were the cost and durability of the exhibits. "Are the visitors learning anything?" we asked. He still didn't understand. Attendance at the museum was high. It looked attractive. It had probably won a prize. Why were we wasting his time?

Many museums and design centers make prime examples of pretty displays and signs coupled with illegible and uninformative labels. Mostly, I suspect, it's because these buildings are judged as places of art, where the exhibits are meant to be admired, not to be learned from. I made several trips to the Design Centre in London to collect material for this book. I hoped it would have a good library and bookshop (it did) and good exhibits, demonstrating the proper principles for combining aesthetics, economics, usability, and manufacturability. I found the London Design Centre itself to be an exercise in poor design. Take the cafeteria: just about impossible to use. Behind the counter, the four workers continually get in each other's way. The layout of the back-counter facilities seems without structure or function. Food is carefully heated for the customer, but it gets cold by the time the customer gets through the line. The cafeteria has tiny round tables, which are also too high. There are elegant round stools to sit on. The set up is impossible to use if you are elderly or young or have your hands full of packages. Of course, the design may have been a deliberate attempt to discourage use of the cafeteria. Consider this scenario.

The cafeteria is well designed, with spacious tables and comfortable chairs. But it then becomes too popular, interfering with the true pur-

pose of the Design Centre, which is to encourage good design among British manufacturers. The popularity of the Centre and its cafeteria to tourists is unexpected. The Design Centre decides to discourage people from using the cafeteria. They take out the original tables and chairs and replace them with dysfunctional, uncomfortable ones, all in the name of good design—the goal in this case being to discourage people from using the cafeteria and lingering. Actually, restaurants often install uncomfortable chairs for just this reason. Fast-food places often have no chairs or tables. So my complaints provide evidence that the design criteria were met, that the design was successful.[12]

In London I visited the Boilerworks, a part of the Victoria and Albert Museum, to look at a special exhibit called "natural design." The exhibit itself was one of the best examples of unnatural design I have ever witnessed. Pretty, tasteful signs near each display. Dramatically striking layout of the objects. But you couldn't tell which sign went with which exhibit, or what the text meant. Alas, this seems typical of museums.

A major part of the design process ought to be the study of just how the objects being designed are to be used. In the case of the cafeteria at the London Design Centre, the designers should imagine a crowd of people in line, imagine where the line will start and end, and study what effect the line will have on the rest of the museum. Study the work patterns of the cafeteria employees: consider them responding to customer requests. Where will they have to move? What objects will they have to reach? If there are several employees, will they get in each other's way? And then consider the customers. Grandparents with heavy coats, umbrellas, packages, and perhaps three small children—how will they pay for their purchases? Is there a place for them to put down their packages so they can open their wallets or purses and get out their money? Can this be done in a way that minimizes the disruption for the next people in line and improves the speed and efficiency of the cashier? And finally, consider the customers at the tables. Struggling to get up on a high stool to eat off a tiny table. And don't just imagine: go out and look at the current design, or at other cafeterias. Interview potential customers, interview the cafeteria employees.

In the case of science museums, studies have to be made on people who are the same as the intended audience. The designers and employees already know too much: they can no longer put themselves into the role of the viewer.

Let me be positive for a change: there are science museums and exhibits that work well. The science museums in Boston and in Toronto, the Monterey Aquarium, the Exploratorium in San Francisco. There are probably many others that I do not know about. Consider the Exploratorium. It is dark and grungy on the outside, located in a remodeled, left-over building. Very little is devoted to sleekness or aesthetics. The emphasis is on using and understanding the exhibits. The staff is interested in explaining things.

It is possible to do things right. Just don't let the focus on cost, or durability, or aesthetics destroy the major point of the museum: to be used, to be understood. The problem of focus, I call this.

DESIGNERS ARE NOT TYPICAL USERS

Designers often think of themselves as typical users. After all, they are people too, and they are often users of their own designs. Why don't they notice, why don't they have the same problems as the rest of us? The designers I have spoken with are thoughtful, concerned people. They do want to do things properly. Why, then, are so many failing?

All of us develop an everyday psychology—professionals call it "folk psychology" or, sometimes, "naive psychology"—and it can be as erroneous and misleading as the naive physics that we examined in chapter 2. Worse, actually. As human beings, we have access to our conscious thoughts and beliefs but not to our subconscious ones. Conscious thoughts are often rationalizations of behavior, explanations after the fact. We tend to project our own rationalizations and beliefs onto the actions and beliefs of others. But the professional should be able to realize that human belief and behavior are complex and that the individual is in no position to discover all the relevant factors. There is no substitute for interaction with and study of actual users of a proposed design.

> "Steve Wozniak, the whiz-kid co-founder of Apple Computer offered the first public glimpse of CORE, his latest brainchild. . . .
>
> "CORE, which stands for controller of remote electronics, is a single device that allows consumers to fully operate their home equipment by remote control as long as the equipment is all in one room. . . .

"CORE comes with a 40-page user manual. But Wozniak says users of his gizmo . . . won't be daunted because, initially, most will be 'techies.' "[13]

There is a big difference between the expertise required to be a designer and that required to be a user. In their work, designers often become expert with the *device* they are designing. Users are often expert at the *task* they are trying to perform with the device.[14]

Steve Wozniak designs a device to help people like himself, people who complain that their house is cluttered with too many remote control devices for their electronic components. So he produces a single controller that replaces the many. But the task is complex, the instruction manual thick. Not a problem, we are told, the initial users will be "techies." Just like Wozniak, presumably. But how accurate is that characterization? Do we even know that the technically ambitious, the "techies," will really be able to understand and use the device? The only way to find out is to test the designs on users—people as similar to the eventual purchaser of the product as possible. Furthermore, the designer's interaction with potential users must take place from the very beginning of the design process, for it soon becomes too late to make fundamental changes.

Professional designers are usually aware of the pitfalls. But most design is not done by professional designers, it is done by engineers, programmers, and managers. One designer described the issues to me this way:

"People, generally engineers or managers, tend to feel that they are humans, therefore they can design something for other humans just as well as the trained interface expert. It's really interesting to watch engineers and computer scientists go about designing a product. They argue and argue about how to do things, generally with a sincere desire to do the right thing for the user. But when it comes to assessing the tradeoffs between the user interface and internal resources in a product, they almost always tend to simplify their own lives. They will have to do the work, they try to make the internal machine architecture as simple as possible. Internal design elegance sometimes maps to user interface elegance, but not always. Design teams really need vocal advocates for the people who will ultimately use the interface."[15]

Designers have become so proficient with the product that they can no longer perceive or understand the areas that are apt to cause difficulties. Even when designers become users, their deep understanding and close contact with the device they are designing means that they operate it almost entirely from knowledge in the head. The user, especially the first-time or infrequent user, must rely almost entirely on knowledge in the world. That is a big difference, fundamental to the design.

Innocence lost is not easily regained. The designer simply cannot predict the problems people will have, the misinterpretations that will arise, and the errors that will get made. And if the designer cannot anticipate errors, then the design cannot minimize their occurrence or their ramifications.

THE DESIGNER'S CLIENTS MAY NOT BE USERS

Designers must please their clients, who are often not the end users. Consider major household appliances such as stoves, refrigerators, dishwashers, and clothes washers and dryers; and faucets and thermostats for heating and air conditioning systems. They are often purchased by housing developers or landlords. In business, purchasing departments make decisions for large companies and owners or managers make decisions in small companies. In all these cases, the purchaser is probably interested primarily in price, perhaps in size or appearance, almost certainly not in usability. And once devices are purchased and installed, the purchaser has no further interest in them. The manufacturer is primarily concerned about these decision makers, its immediate customers, not the eventual users.

In some situations cost must be put first, especially in government or industry. In my university, copying machines are purchased by the Printing and Duplicating Center, then dispersed to the various departments. The copiers are purchased after a formal "request for proposals" has gone out to manufacturers and dealers of machines. The selection is almost always based solely on price, plus a consideration of the cost of maintenance. Usability? Not considered. The state of California requires by law that universities purchase things on a price basis; there are no legal requirements regarding understandability or usability of the product. That is one reason we get unusable copying machines and telephone systems. If users complained strongly enough, usability could become a requirement in the purchasing specifications, and that demand could trickle back to the designers. But without this feedback, designers must often design the cheapest possible products because those are what sell.

Designers face a tough task. They answer to their clients, and it may be hard to find out who the actual users are. Sometimes they are even prohibited from contacting the users for fear they will incidentally reveal company plans for new products or mislead users into believing that new products are about to be developed. The design process is a captive of corporate bureaucracy, with each stage in the process adding its own assessment and dictating the changes it believes essential for its concerns. The design is almost certainly altered as it leaves the designers and proceeds through manufacturing and marketing. All participants are well intentioned, and their particular concerns are legitimate. The factors should all be considered simultaneously, however, and not subject to the accidents of time sequence or the realities of corporate rank and clout. One designer wrote me this about his problems:

"Most designers live in a world where the gulf of evaluation is infinite. True, we often know the product too well to envision how people will use it, yet we are separated from the end users by multiple layers of corporate bureaucracy, marketing, customer services, etc. These people believe they know what customers want and feedback from the real world is limited by filters they impose. If you accept the problem definition (product requirements) from these outside sources without personal investigation you will design an inferior product regardless of your best intentions. If this initial hurdle is overcome you are only halfway home. The best design ideas are often ruined by the development-manufacturing process that takes place when they leave the design studio. What this really points out is that the process by which we design is flawed, probably more so than our conception of how to create quality designs."[16]

The Complexity of the Design Process

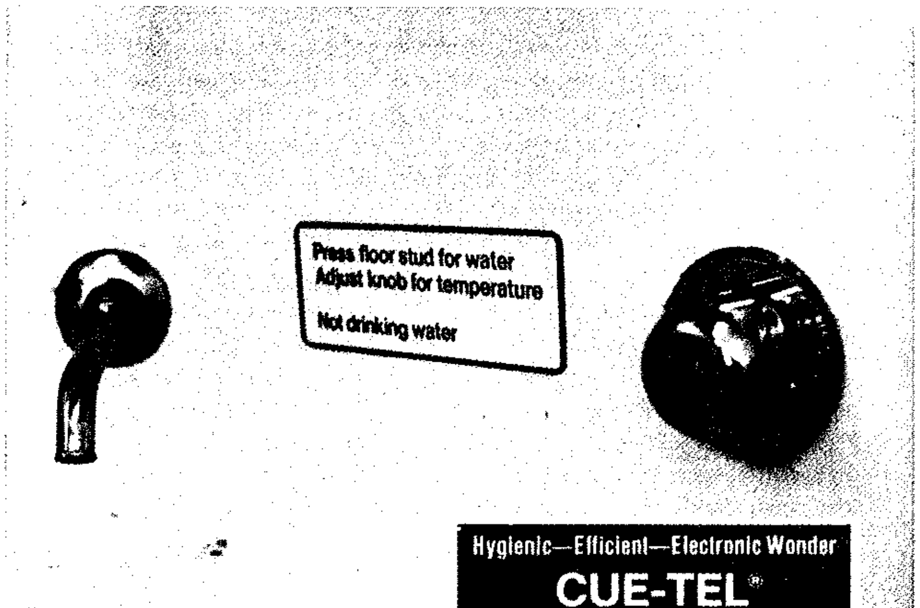
"Design is the successive application of constraints until only a unique product is left."[17]

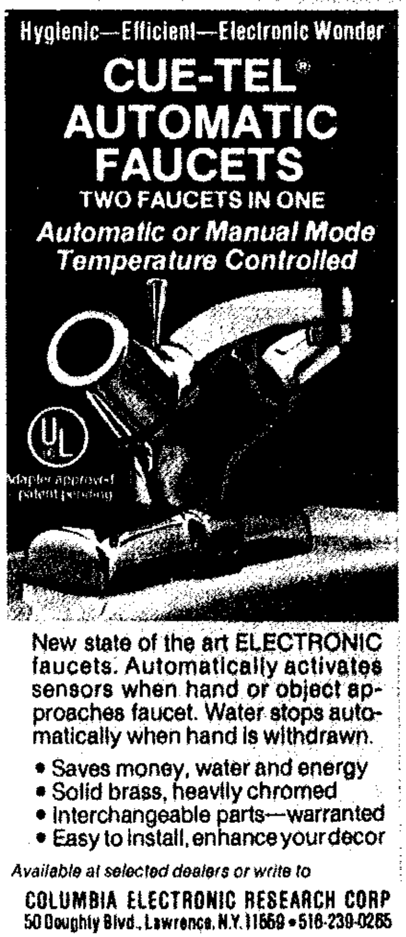
You might think that a water faucet would be pretty easy to design. After all, you merely want to start or stop the flow of water. But consider some of the problems. Suppose the faucets are for use in public places, where users may fail to turn them off. You can make a spring-operated faucet, which operates only as long as the handle is held. This automatically turns the faucet off; but it is difficult for users to hold the handle while wetting their hands. Ok, so you add a timer; then one push on the faucet handle yields five or ten seconds of water flow. But the extra complexity of the faucet design adds to the cost and lowers the reliability of the faucet. Furthermore, it is difficult to decide how long the water should stay on. Somehow it never seems like long enough for the user.

How about a foot-operated faucet, which overcomes the problems of springs and timers because the water stops as soon as the foot leaves the pedal (figure 6.4 *A*)? This solution requires slightly more elaborate plumbing, again raising the cost. It also makes the control invisible, violating a major design principle and making it difficult for a new user to find the control. How about a high-technology solution, with automatic sensors that turn on the water as soon as a hand is placed in the sink, turning it off as soon as it leaves (figure 6.4 *B*)? This solution has several problems. First, it is expensive. Second, it makes the controls invisible, causing difficulty for new users. And third, it is not easy to see how the user could control either the volume of water or the temperature. More on this faucet later.

Not all faucets are designed under the constraints of public faucets. At home, aesthetic considerations tend to dominate. Styles often reflect the social and economic class of the user. And different kinds of users have different requirements.

The same considerations hold true for most everyday things. The variety of possible solutions to the usual problems is enormous. The range of expression permitted the designer is vast. Moreover, the number of tiny details that must be accounted for is astounding. Pick up almost any manufactured item and examine its details with care. The little wiggly bends on a hairpin are essential in keeping it from slipping out of the hair: someone had to think of that, then design special equipment to create the bends. The felt-tip pen I am examining as I write has six different sizes on the pen body, two different sizes on the cap. The pen changes its taper at numerous spots, each change serving some function. Four different substances comprise the pen body (and I am not counting the ink, the container that holds the ink, or the felt





6.4 **Nonstandard Faucets.** There are often good reasons for using nonstandard means for operating faucets, but the result is that the user is apt to need help to operate them. *A* (above) shows the faucet and operating instructions from the sink in a British train. *B* (right) shows an advertisement for an automatic faucet: simply put the hand under it and the water comes out at a preset temperature and rate of flow. Convenient, but only for those who know the secret.

tip). The cap is made of two kinds of plastic and one kind of metal. The inside of the cap has a number of subtle indentations and internal structures that clearly match up with corresponding parts of the pen body, both to hold the cap on firmly and to prevent the felt tip from drying out. There are more parts and variables than I would ever have imagined.

The pen's designer must be aware of hundreds of requirements. Make the pen too thin, and it will not be strong enough to stand up to the hard use of schoolchildren. Make the middle section too thick, and it can neither be grasped properly by the fingers nor controlled with enough precision. Yet people with arthritic hands may need a thick body because they can't close their fingers entirely. Leave out the tiny hole near the tip, and pressure changes in the atmosphere will cause the ink to leak out. And what of those who use the pen as a measuring device or as a mechanical implement to pry, poke, stab, and twist? For example, the instructions for the clock in my automobile say to set it by depressing the recessed button with the tip of a ball-point pen. How could the pen designer have known about this? What obligation does the designer have to consider varied and obscure uses?

DESIGNING FOR SPECIAL PEOPLE

There is no such thing as the average person. This poses a particular problem for the designer, who usually must come up with a single design for everyone; the task is difficult when all sorts of people are expected to use the item. The designer can consult handbooks with tables that show average arm reach and seated height, how far the average person can stretch backward while seated, and how much room is needed for average hips, knees, and elbows. Physical anthropometry the field is called. With the data the designer can try to meet the size requirements for almost everyone, say for the 90th, 95th, or even the 99th percentile. Suppose you design a product for the 95th percentile, that is, for everyone except the 5 percent of people who are smaller or larger. You're leaving out a lot of people. If the United States has 250 million people, 5 percent is 12.5 million. Even if you design for the 99th percentile you'll leave out 1 percent of the population—2.5 million.

Consider typists. Typists need to have their hands comfortably poised above the keyboard. Because of the thickness of typewriters, typing tables are designed to be lower than work tables. Of course, what matters is not the table height or the keyboard thickness, but the distance from the normal position of the typist's hands to the keyboard, which is determined by several factors:

- How big the typist is: legs, chest, hands
- How high the table is
- How thick the keyboard is
- How high the chair is

What can the designer do? One solution is to make everything adjustable: chair height, height and angle of the typing table. In fact, good typing tables have several parts: a part for the keyboard, a part for the computer screen, a part that holds working papers. Let each part be separately adjustable in height and angle. Then everyone can be accommodated.

Some problems are not solved by adjustments. Left-handed people, for example, present special problems. Simple adjustments won't work, nor will averages: average a left-hander with a right-hander and what do you get? Here is where special products help—left-handed scissors and knives, left-handed rulers (figure 6.5). These special-purpose devices don't always work, of course, not when one device is to be used by many, or where the items are too large or expensive for each person to own or to carry around. In such cases the only solution is to make the device itself ambidextrous, even if that makes it a bit less efficient for each person.

Consider the special problems of the aged and infirm, the handicapped, the blind or near-blind, the deaf or hard of hearing, the very short or very tall, or the foreign. Wheelchairs, for example, cannot easily manipulate curbs, stairs, or narrow aisles. As we age, our physical agility decreases, our reaction time slows, our visual skills deteriorate,

6.5 Left-handed Ruler. Writing from left to right with the left hand means that you cover what you write, making rulers hard to use, smearing the ink. A left-handed pen is a pen with fast-drying ink. This ruler for left-handers has the numbers going from right to left. One solution to the problem of diversity among individuals is to produce specialized objects.

and our ability to attend to several things at once or to switch rapidly among competing events decreases.

High-speed highways pose special problems for the aged. An automobile traveling at high speed on a crowded highway at dusk is already pushing the limit of the driver's capabilities. The elderly are pushed beyond their limit. The solution adopted by many elderly drivers is to travel very slowly, to adjust their speed to what their processing can handle comfortably. Unfortunately, the slow driver poses a hazard to other drivers: on high-speed highways, things are considerably safer if everyone travels at approximately the same speed. I see no simple solution to this problem. In many cities, especially in the United States, there is no easy way to get from one place to another except by private automobile. Yet the elderly can't be expected to stay home. The solution has got to be either increased public transportation, or supplied drivers, or perhaps special streets or highway lanes with slower speed limits. Automated cars, the dream of science fiction writers and city planners, may still one day come about; they would take care of this problem.

Those of you who are young, do not smirk. Our abilities begin to deteriorate relatively early, starting in our mid-twenties. By our mid-forties our eyes can no longer adjust sufficiently to focus over the entire range of distances, so most of us need reading glasses or bifocals. Bifocals make it harder to do fine work, harder to use computer terminals (whose screens seem to be designed for twenty-year-olds).

I type these words seated in front of my computer terminal, head tilted upward at an uncomfortable angle so that I can see the screen out of the bottom half of my eyeglasses. I can't figure out how to get comfortable. Lower the screen and it gets in the way of my typing. Use special "computer" glasses adjusted for screen size and distance, and I can't read all the notes and outlines scattered about me at various distances. Fortunately, I can change the size of the type that appears on the screen. I use a twelve-point font, one whose letters are comfortably large. Alas, this is a tradeoff, for the larger the letters on the screen, the less material can fit. Change to nine-point font and I can see 78 percent more material (33 percent more lines, each with 33 percent more words): a non-trivial difference when I'm trying to write long sections. But the letters are 33 percent smaller, making it harder both to read and to correct them. At least my computer allows flexibility in type size; most do not.

By the time we're sixty, enough stray material has scattered about in our eyes that visual contrast is diminished, enough to be one of the major reasons that airline pilots are forced to retire at this age. At the age of sixty a person is still in good mental and physical shape, and the accumulated wisdom of the years leads to superior performance in many tasks. But physical strength is lessened, the agility of the body decreased, and the speed of some operations lessened. In a world where the average age is increasing, sixty is still relatively young: most sixty-year-olds have another twenty years to live, many have forty. We need to design with these people in mind—think of it as designing with our future selves in mind.

There is no simple solution, no one size fits all. But designing for flexibility helps. Flexibility in the size of the images on computer screens, in the sizes, heights, and angles of tables and chairs. Flexibility on our highways, perhaps making sure there are alternative routes with different speed limits. Fixed solutions will invariably fail with some people; flexible solutions at least offer a chance for those with special needs.

SELECTIVE ATTENTION: THE PROBLEM OF FOCUS

The ability of conscious attention is limited: focus on one thing and you reduce your attention to others. Psychologists call the phenomenon "selective attention." Excessive focus leads to a kind of tunnel vision, where peripheral items are ignored.

I watched a consumer show on British television on toasters that caught fire when the bread was too dry. The consumer representatives pointed out that people often inserted their fingers, a fork, or a knife into the toaster to extract the toast. This was very dangerous (even more dangerous in Britain than in the United States because the voltage is 240 volts, not the 120 of the United States). Yet some toasters had exposed wires very close to the top, quite reachable by the finger or the metal utensil. The consumer representatives argued that manufacturers should not have placed the wires so close to the opening.

The manufacturers denied that their toasters were dangerous. "Why," they asked, "would someone stick their fingers or a knife into a toaster?" Certainly the instructions warned them not to. Certainly they must know it is dangerous. To the designer, such an action is so unthinkable that prevention did not enter into the design considerations.

Consider the matter from the user's point of view. The person sees a problem—stuck or burning toast—and focuses on the solution—to extract it. The danger does not come to mind. To my own surprise, I did the same thing the very next day. I inserted two crumpets into the toaster; a few minutes later, smoke was pouring out. Quick, I ran over to the toaster, popped up the crumpets as far as they would go, and then quickly (but carefully?) inserted a knife blade into the toaster, down the side, to lift them out. What was I doing?

Selective attention: attend to the immediate problem, forget the rest. Sure I was being careful, but that is probably what the people who electrocuted themselves also believed. It just didn't seem dangerous, that's all.

The same story is repeated over and over again. Underwater divers focus so much on struggling to the surface that they fail to release the lead weights (on a special easy-to-release belt) that are keeping them underwater. People who are fleeing a fire push hard against a door, harder and harder, failing to recognize that the door opens by pulling. Someone is trapped behind a door, pushing against the left side when it opens from the right. Motorcyclists have their helmets strapped to their bike, not their head. People don't use seatbelts, or they drive too fast, because it is inconvenient to do otherwise and because they don't see the danger.

When there is a problem, people are apt to focus on it to the exclusion of other factors. The designer must design for the problem case, making other factors more salient, or easier to get to, or perhaps less necessary. That's what the forcing functions of chapter 5 were all about. Make the power cutoff switch of the toaster a forcing function, so that a person can't stick something into the toaster without flipping a power cutoff switch (which should be easy to get to and use). Or change the design of the wiring and heating elements so that lethal elements cannot be reached from outside, no matter what flesh or metal gets put into the toaster.

A corollary principle is that designers must guard against the problems of focus in their own design. Did their attention to one set of variables cause them to neglect another? Did safety suffer for usability? Usability for aesthetics? Aesthetics for manufacturability?

The Faucet: A Case History of Design Difficulties

It may be hard to believe that an everyday water faucet could need an instruction manual. I saw one, this time at the meeting of the British Psychological Society in Sheffield, England. The participants stayed in dormitories. Upon checking into one of these, the Ranmoor House, a guest was given a pamphlet that gave useful information: where the churches were, the times of meals, the location of the post office, and how to work the taps (faucets). "The taps on the washhand basin are operated by pushing down gently."

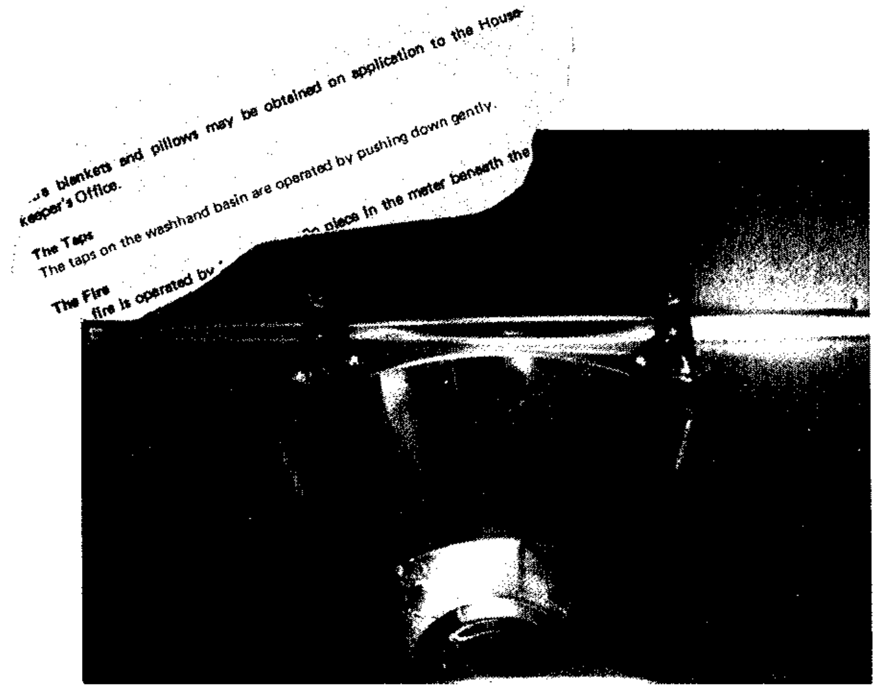
When it was my turn to speak at the conference, I asked the audience about those taps. How many had trouble using it? Polite, restrained titterings from the audience. How many tried to turn the handle? A large show of hands. How many had to seek help? A few honest folks raised their hands. Afterward, one woman came up to me and said that she had given up and had to walk up and down the halls until she found someone who could explain the taps to her.

A simple sink, a simple-looking faucet. But it looks like it should be turned, not pushed (figure 6.6 A*). If you want the faucet to be pushed, make it look like it should be pushed. It can be done: the airlines do it right (figure 6.6* B*).*

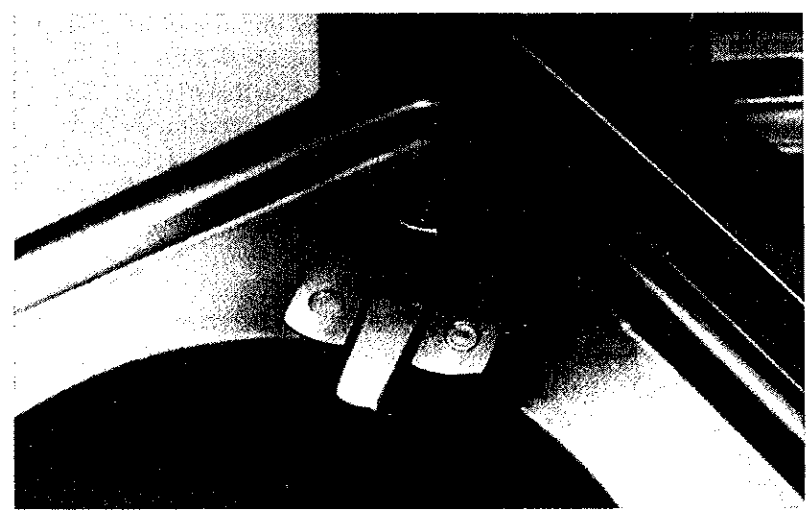
Pity the poor house porters, always getting calls for help about the faucets. So instructions were put in the orientation sheet. Who would ever think of having to read instructions before using a faucet? At least put them on the faucets, where they can't be missed. But when simple things need instructions, it is a certain sign of poor design.

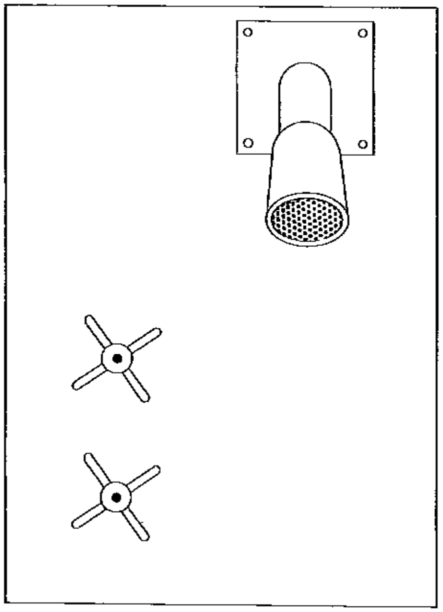
Why are faucets so hard to get right? Let us take a closer look at the two major variables (they will give us quite enough to do). The person who uses the faucets cares about two things: the water temperature and volume. Two things to control. We should be able to do that with two controls, one for each. Except that water comes in two pipes, hot and cold, and so the two things that are easiest to control—volume of hot water and volume of cold water—are not the two things we want to have controlled. Hence the designer's dilemma.

There are three problems; two relate to the mapping of intentions to actions, and the third is the problem of evaluation:



6.6 Contrasting Designs for "Push" Faucets. The faucets *A* (above) in the Ranmoor dormitory at the University of Sheffield give little clue to their mode of operation. As a result, occupants must be supplied with the instruction sheet for "the taps". The faucets *B* (below) on the sink of a commercial airline are designed properly. Pushing is clearly indicated. No instruction manual is required.





6.7 Vertical Faucets. The world standard is that hot is on the left, cold on the right. What do you do here? Why would anyone dream up this scheme?

- Which faucet controls the hot, which the cold?
- What do you do to the faucet to make it increase or decrease the water flow?
- How do you determine if the volume or temperature is correct?

The two mapping problems are solved through cultural conventions, or constraints. It is a worldwide convention that the left faucet should be hot, the right cold. It is also a universal convention that screw threads are made to tighten with clockwise turning, loosen with counterclockwise. You turn off a faucet by tightening a screw thread (tightening a washer against its seat), thereby shutting off the flow of water. So clockwise turning shuts off the water, counterclockwise turns it on.

Unfortunately, the constraints do not always hold. Most of the English people I asked were not aware that left/hot, right/cold was a convention; it is violated too often to be considered a convention in England. But the convention isn't universal in the United States. Look at the picture of a shower control from my own university (figure 6.7). Here we have vertical faucets. Vertical? If left is the standard for hot, how does that translate to vertical arrangements? Is hot the top or the bottom? Weird.

Sometimes a designer messes with the convention on purpose. The human body has a mirror-image symmetry, says this pseudopsychologist. So if the left hand moves clockwise, why, the right hand should move counterclockwise. Watch out, your plumber or architect may install a bathroom fixture where clockwise rotation turns the hot water off and the cold water on. Or is it the other way around? No matter, as you try to control the water temperature, soap running down over your eyes, groping to change the water control with one hand, soap or shampoo clutched in the other, you are guaranteed to get it wrong. The water is freezing, so you try to increase the amount of hot. You will probably turn on the shower, or the bath, or open the drain (or shut it), or turn off the hot water completely, or scald yourself.

Whoever invented that mirror-image nonsense should be forced to take a shower. Yes, there is some logic to it. To be a bit fair to the inventor of the scheme, it does work reasonably well as long as you always use the faucets by placing both hands on them at the same time, adjusting both controls simultaneously. It fails miserably, however, when one hand is used to alternate between the two controls. Then you cannot remember which direction does what.

What about the evaluation problem? Feedback in the use of most faucets is rapid and direct, so turning them the wrong way is easy to discover and correct—the evaluate-action cycle is easy to traverse. As a result, the discrepancy from normal rules is often not noticed. Unless you are in the shower and the feedback occurs when you scald yourself.

Older sinks have two separate spouts. Here evaluation is difficult. You can wave your hand rapidly back and forth between the spouts, hoping thereby to get a good mix of temperatures, or you can fill up the basin, adjusting the amount of hot and cold water so that the accumulating mixture reaches the desired temperature. Usually you settle for anything in the neighborhood. Each problem alone isn't a big deal. But the total sum of all the trivial mal-design unnecessarily adds to the trauma of everyday life.

Now consider the modern single-spout, single-control faucet. Technology to the rescue. Move the control one way, it adjusts temperature. Move it another, it adjusts volume. Hurrah! We control exactly the variables of interest, and the mixing spout solves the evaluation problem.

Yes, these new faucets are beautiful. Sleek, elegant, prize winning.

Unusable. They solved one set of problems only to create yet another. The mapping problems now predominate.

- Which control is associated with which action?
- What operations do you apply to the controls?

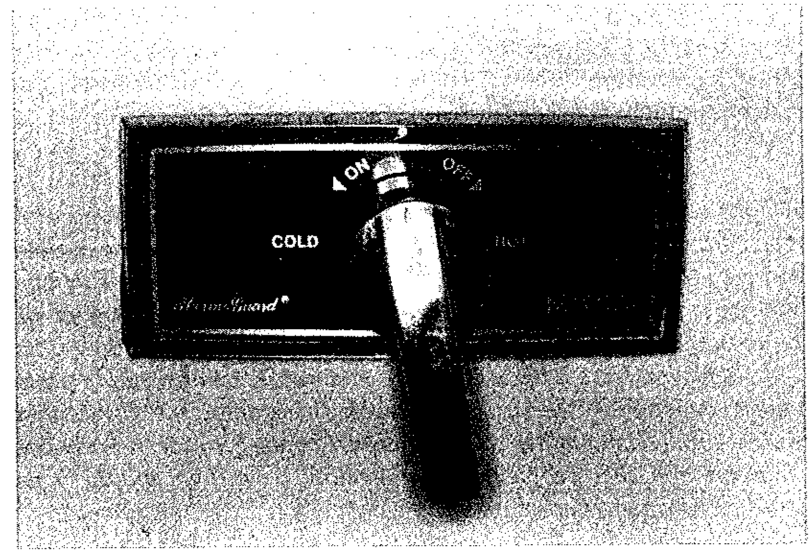
The problem is that it is very difficult to figure out which part of the sleek faucet is the control. And even if you figure that out, it is hard to figure out in which direction it moves. And once you figure that out, it is hard to figure out which direction controls which action. And when these fancy, multipurpose, sleek designs also control the basin plug and the diversion of water to shower or bath, disaster awaits.

There are two problems here. First, in the name of elegance, the moving parts sometimes meld invisibly into the faucet structure, making it nearly impossible even to find the controls, let alone figure out which way they move or what they control. Second, in the name of novelty, the new designs have forfeited the power of cultural constancy. Users don't want each new design to use a different method for controlling the water. Users need standardization. If all makers of faucets could agree on a standard set of motions to control amount and temperature (how about up and down to control amount—up meaning increase—and left and right to control temperature—left meaning hot?), then we could all learn the standards once, and forever afterward use the knowledge for every new faucet we encountered.

> If you can't put the knowledge on the device, then develop a cultural constraint: standardize what has to be kept in the head.

There could be small variations in the standard. Suppose a designer wanted temperature to be controlled by a knob that turned rather than a lever that moved left and right. Fortunately, there is a natural mapping that relates turning to direction: a clockwise turning is the same as moving to the right—getting colder—and a counterclockwise turning is the same as moving to the left—hotter.

Technological development never ceases. There is yet another solution to the control problem, one that has a slight virtue over the others: it is cheaper. One control turns the water on or off and lets you adjust either temperature or volume, but not both (figure 6.8). All you have to do is to locate the control and operate it. Think of all the mental energy and confusion you have been saved. We finally have a control that is truly easy to use. Success.



6.8 Simpler Faucets. In *A* (above) the mapping problem is solved—the faucet is presumed to be easy to use. The problem is that you cannot control the amount of water. On top of that, once the knob has been turned 180°, it is no longer clear which way to turn it in order to make the water hotter or colder. Faucet *B* (below) couldn't be simpler. It certainly is easy to use. Of course, you can only turn it on; you get a fixed temperature and a fixed volume of water.

Wait, we really do want to control both amount and temperature independently. This solution gives us only one control. So we can adjust temperature, but we get out whatever amount of water the designer thought was good for us. Or we can adjust the amount while getting an arbitrary temperature. The story of progress.

Some variants on this faucet control only on or off: you have control over neither volume nor temperature. Sometimes there is no visible means to turn on the water. How does the novice user realize that one is supposed to wave the hands under the faucet? There is no sign of the required operation, no relevant information in the world.

Perhaps you have a big sign: "Do not adjust controls, simply place hand under spout." The sign ruins the elegance, doesn't it? Interesting choice—understandability or elegance. Of course, if such faucets became common, then people would know how to use them and the signs could come down. Someday.

Two Deadly Temptations for the Designer

Let's return to the designer's problems. I've mentioned the time and economic pressures on them. Now let me tell you of two deadly temptations that await the unwary, temptations that lead toward products that are overly complex, products that drive users to distraction—I call these creeping featurism and the worshipping of false images.

CREEPING FEATURISM

I recently attended a demonstration of a new word processing program, held in a large, crowded auditorium. A representative from the company sat in front of the computer, a video projector putting a large image of the computer screen onto the movie screen. The audience was skeptical: they were experts and knew the limitations of such programs. The demonstrator was smooth and convincing, composing an outline, expanding it into text, indenting the paragraphs, numbering them, changing their styles, flipping into a drawing program, drawing a figure, inserting the drawing into the text with the text flowing neatly around the drawing. "You want two columns?" asked the demonstrator. "Here it is. Three columns? Four? Just name it." The screen flowed: three columns of text neatly lined up, illustrations just where they ought to be, page headers, footers, paragraph numbers, boldface italics. Large type, small type, footnotes neatly displayed at the end of columns. You could even highlight just the things that had been changed in the last revision. You could leave notes for yourself or a co-author, notes that would appear on the screen but need not be printed in the final text.

The audience applauded. They called out for their favorite features. Usually the demonstrator would say, "Yes, I am glad you asked, here it is," and whiz, bang, wave of hands, click of keys, swish of the mouse, and the screen would display the latest called-for feature. Sometimes the demonstrator would say, "Not yet, it will be in the second release—in a few months."

Creeping featurism is the tendency to add to the number of features that a device can do, often extending the number beyond all reason. There is no way that a program can remain usable and understandable by the time it has all of those special-purpose features. The word processor I use on my home computer comes with a 340-page reference manual, plus a 150-page introductory manual intended for first-time users (who probably can't understand the reference manual until they have first read the learning manual). EMACS, the text editor I use on my university computer, comes with a 250-page manual, which would be longer if you weren't assumed to be expert at many things.

How can users cope? How can users protect themselves from themselves? After all, as the story of the demonstration illustrates, it is the users who request the features; the designers are simply obliging them. But each new set of features adds immeasurably to the size and complexity of the system. More and more things have to be made invisible, in violation of all the principles of design. No constraints, no affordances; invisible, arbitrary mappings. And all because the users have demanded features.

Creeping featurism is a disease, fatal if not treated promptly. There are some cures, but, as usual, the best approach is to practice preventive medicine. The problem is that the disease comes so naturally, so innocently. Analyze a task, and you see how it can be made easier. Why, adding features seems so virtuous, following the very preachings of this book, simply trying to make life easier for everyone. But with extra features comes extra complexity. Each new feature adds yet another control, or display, or button, or instruction. Complexity probably increases as the square of the features: double the number of features, quadruple the complexity. Provide ten times as many features, multiply the complexity by one hundred.

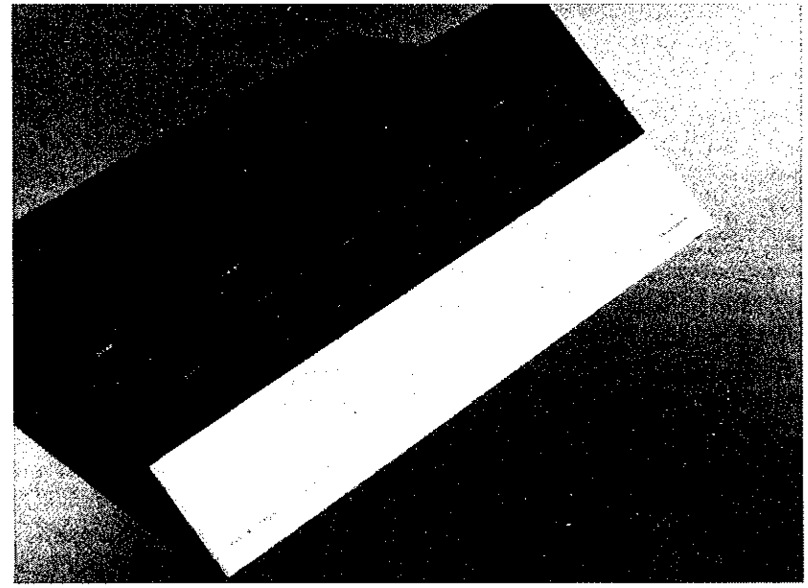
There are two paths to treating featurism. One is avoidance, or at the least, great restraint. Yes, allow features that seem absolutely necessary, but steel yourself to the rigors of doing without the rest. Once a device has multiple functions, there is no way to avoid having multiple controls and operations, multiple pages of instructions, multiple difficulties and confusions.

The second path is organization. Organize, modularize, use the strategy of divide and conquer. Suppose we take each set of features and hide them away in separate locations, perhaps with dividing barriers between sets. The technical word is *modularization*. Create separate functional modules, each with a limited set of controls, each specialized for some different aspect of the task. The virtue is that each separate module has limited properties, limited features. Yet the sum total of features in the device is unchanged. The proper division of a complex set of controls into modules allows you to conquer complexity (as can be seen in figure 6.9).

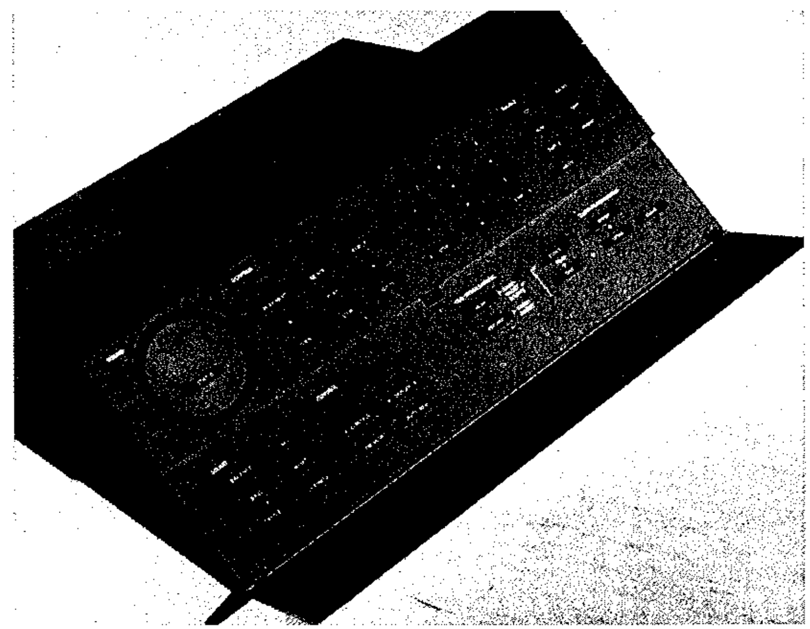
HE WORSHIPPING OF FALSE IMAGES

The designer—and user—may further be tempted to worship complexity. Some of my students did a study of office copying machines. They discovered that the most expensive, most feature-laden machines were best sellers among law firms. Did the firms need the extra features of the machines? No. It turns out that they liked to put them in the front offices where clients were waiting—impressive machines, with flashing lights and pretty displays. The firm gained an aura of being modern and up to date, capable of dealing with the rigors of modern high technology. The fact that the machines were too complex to be mastered by most of the people in the firms was irrelevant: the copiers did not even have to be used—appearance alone did the job. Ah, yes, the worshipping of false images, in this case, by the customers.

A colleague told me of her difficulties with her home audio/television set. It was comprised of separate components, each alone not too complex. But the combination was so overwhelming that she could not



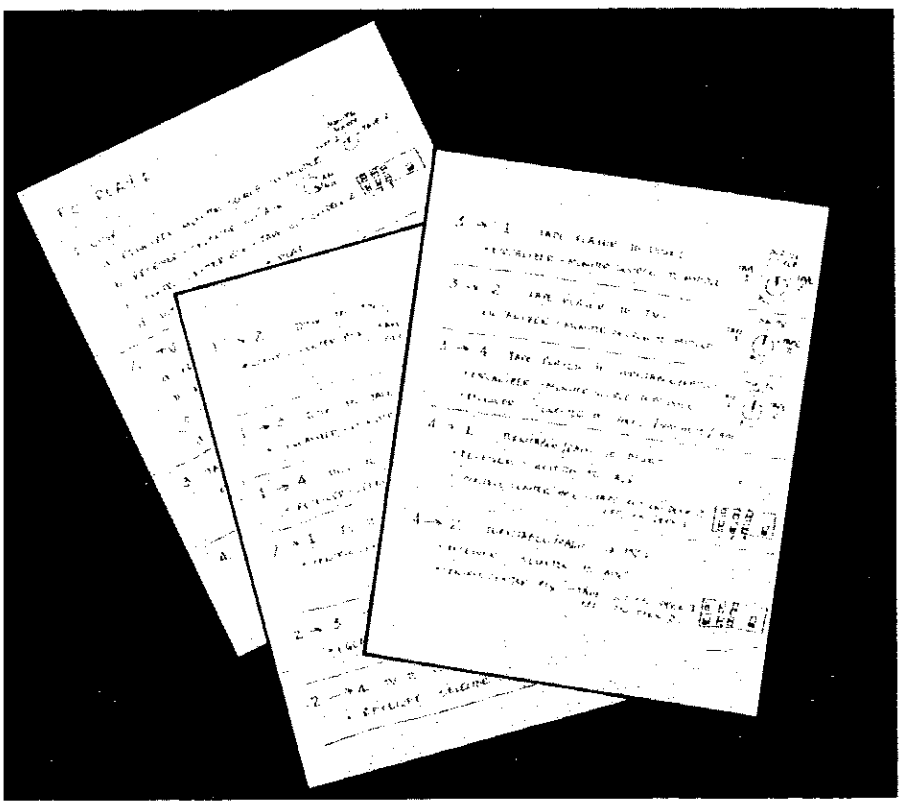
6.9 Overcoming Complexity through Organization. The remote control device *A* (above) for the Bang & Olufsen audio set (there are no controls on the set itself) serves numerous features and options. The controls are made simple through several principles. First, the buttons are grouped into logical, functional modules. Second, the display on the remote gives good feedback about the operation. Third, infrequently used controls are hidden beneath a panel *B* (below), which reduces the visual complexity in normal use but is available when needed.



use it. Her solution was to work through each of the operations she wished to perform and write explicit instructions for herself (figure 6.10). And even with these instructions, operation was not easy. Here the culprit clearly is the interactions among components. Imagine having to write several pages of instructions in order to use your own audio set!

In the case of the overly complex audio/television set, the components were from different manufacturers. Nonetheless, they were intended to be purchased and used individually. I have seen equal complexity in components from a single manufacturer. Some salespeople try to create the impression that this is how it has to be, that anyone with any technical competence can manage to work the devices. No,

6.10 A Personal Instruction Manual. My colleague had to write out three pages of instructions to help herself set up any desired configuration of her audio/video components. Too many interacting parts, too much complexity.



that attitude won't work. The equipment is simply too complex, the interaction between components too overwhelming. There was nothing particularly elaborate about my colleague's equipment. This person was reasonably sophisticated in technical things—she has a Ph.D. in computer science—but was baffled by an everyday audio set.

One of the problems with audio/video equipment is that even if each component has been designed with care, the interaction between components causes problems. The tuner, cassette deck, television, VCR, CD player, and so on, all seem to be designed in relative isolation. Put them all together and there is chaos: an amazing proliferation of controls, lights, meters, and interconnections that can defeat even the most talented.

In this case, the false image is appearance of technical sophistication. This is the sin responsible for the extra complexity of many of our devices, from telephones and televisions to dishwashers and washing machines, from automobile dashboards to audiovisual sets. There is no remedy except through education. You might argue that this is a victimless sin, hurting only those who practice it, but this is not true. Manufacturers and designers produce products for what they perceive as their market demands; therefore, if enough people sin in this way—and the evidence is that they do—then all the rest of us must pay for the pleasures of a few. We pay in fancy, colorful-looking equipment that is nearly impossible to use.

The Foibles of Computer Systems

Now turn to the computer, an area where all the major difficulties of design can be found in profusion. In this realm the user is seldom considered. There is nothing particularly special about the computer; it is a machine, a human artifact, just like the other sorts of things we have looked at, and it poses few problems that we haven't encountered already. But designers of computer systems seem particularly oblivious to the needs of users, particularly susceptible to all the pitfalls of design. The professional design community is seldom called in to help with computer products. Instead, design is left in the hands of engineers and programmers, people who usually have no experience, and no expertise in designing for people.

The abstract nature of the computer poses a particular challenge for the designer. The computer works electronically, invisibly, with no sign of the actions it is performing. And it is instructed through an abstract language, one that specifies the internal flow of control and movement of information, but one that is not particularly suited for the needs of the user. Specialized programmers work in these languages to instruct the system to perform its operations. The task is complex, and programmers must have a variety of skills and talents. The design of a program requires a combination of expertise, including technical skills, knowledge of the task, and knowledge of the needs and abilities of the users.

Programmers should not be responsible for the computer's interaction with the user; that is not their expertise, nor should it be. Many existing programs for user applications are too abstract, requiring actions that make sense for the demands of the computer and to the computer professional but that are not cohesive, sensible, necessary, or understandable to the everyday user. To make the system easier to use and to understand requires a large amount of extra work. I sympathize with the problems of the programmer, but I cannot excuse the general lack of concern for the users.

IOW TO DO THINGS WRONG

Ever sit down to a typical computer? If so, you have encountered "the tyranny of the blank screen." The person sits in front of the computer screen, ready to begin. Begin what? How? The screen is either completely blank or contains noninformative symbols or words that give no hint of what is expected. There is a typewriterlike keyboard, but there is no reason to suppose that one key is preferable to any other. Anyway, isn't it true that one wrong keystroke can blow up the machine? Or destroy valuable data? Or accidentally get connected to some top-secret data bank and then be investigated by the Secret Service? Who knows what danger lurks in the keypress? It is almost as frightening as being taken to a party filled with strange people, being led to the center of the room and let go. Your host disappears, saying: "Make yourself at home. I'm sure there are lots of people you can talk to." Not me. I retreat to the fringes and try to find something to read.

What is the problem? Nothing special, just more of everything. The special powers of the computer can amplify all the usual problems to new levels of difficulty. If you set out to make something difficult to use, you could probably do no better than to copy the designers of modern computer systems. Do you want to do things wrong? Here is what to do:

- Make things invisible. Widen the Gulf of Execution: give no hints to the operations expected. Establish a Gulf of Evaluation: give no feedback, no visible results of the actions just taken. Exploit the tyranny of the blank screen.
- Be arbitrary. Computers make this easy. Use nonobvious command names or actions. Use arbitrary mappings between the intended action and what must actually be done.
- Be inconsistent: change the rules. Let something be done one way in one mode and another way in another mode. This is especially effective where it is necessary to go back and forth between the two modes.
- Make operations unintelligible. Use idiosyncratic language or abbreviations. Use uninformative error messages.
- Be impolite. Treat erroneous actions by the user as breaches of contract. Snarl. Insult. Mumble unintelligible verbiage.
- Make operations dangerous. Allow a single erroneous action to destroy invaluable work. Make it easy to do disastrous things. But put warnings in the manual; then, when people complain, you can ask, "But didn't you read the manual?"

This list is getting depressing, so let us turn to the good side. The computer has vast potential, more than enough to overcome all its problems. Because it has unlimited power, because it can accept almost any kind of control, and because it can create almost any kind of picture or sound, it has the potential to bridge the gulfs, to make life easier. If designed properly, systems can be tailored for (and by) each of us. But we must insist that the computer developers work for us—not for the technology, not for themselves. Programs and systems do exist that have shown us the potential; they take the user into account, and they make it easier for us to do our tasks—pleasurable, even. This is how it ought to be. Computers have the power not only to make everyday tasks easier, but to make them enjoyable as well.

IT'S NOT TOO LATE TO DO THINGS RIGHT

Computer technology is still young, still exploring its potential. The notion lingers that if you have not passed the secret rites of initiation into programming skills, you should not be allowed into the society of computer users. It is like the early days of the automobile: only the brave, the adventurous, and the mechanically sophisticated need apply.

Computer scientists have so far worked on developing powerful programming languages that make it possible to solve the technical problems of computation. Little effort has gone toward devising the languages of interaction. Every student programmer takes courses on the computational aspect of computers. But there are very few courses on the problems faced by the user; such courses are usually not required, and they are not easy to fit into the already crowded schedule of the fledgling computer scientist. As a result, most programmers fluently write computer programs that do wonderful things but that are unusable by the non-professional. Most programmers have never thought of the problems faced by the users. They are surprised to discover that their creations tyrannize the user. There is no longer any excuse for this. It is not that difficult to develop programs that make visible their actions, that allow the user to see what is going on, that make the set of possible actions visible, that display the current state of the system in a meaningful and clear way.[18]

Let me give some examples of excellent work, systems that do take into account the needs of the user. First, there is the spreadsheet, an accounting program that has changed the face of office bookkeeping. The first spreadsheet program, Visicalc, was so impressive that people bought computers just so they could use this one program. That is a strong argument for its usability. Spreadsheets have their problems; but on the whole, they allow people to work with numbers in a convenient way, with immediately visible results.

What did people like about the spreadsheet? The way it looked. You didn't seem to be using a computer—you were working on your problem. You organized the problem just the way you always would, except now it was easy to make changes, easy to see the results. Change one number and everything that depended on that number changed along with it, in just the proper way. What a painless way to do budget projections. All the benefits of the computer, without the technical impediments. In fact, the best computer programs are the ones in which the computer itself "disappears," in which you work directly on the problem without having to be aware of the computer.

Actually, Visicalc had numerous problems. The concept was brilliant, but the execution was flawed. I'm not complaining about the designers, for they were limited by the power of an earlier generation of personal computers. Today's personal computers are much more powerful, and the spreadsheet programs are much easier to use. But the program established the model: it felt as if you were working directly on the problem, not on a computer.

It is not easy to develop effective and usable computer systems. For one thing, it is expensive. Consider the principles described in this book: visibility, constraints, affordances, natural mappings, feedback. Applied to computer systems, these mean that, among other things, the computer must be capable of making things visible (or audible), which requires large and high quality visual displays, a variety of input devices, and plenty of computer memory. These require faster, more powerful computer circuits. And all this adds up to more expensive systems: more cost to manufacture, most cost to the consumer. It may not be immediately clear that the everyday users of computer systems are the ones who require the most powerful systems, with the most memory and the best displays. Professional programmers can get by with less, for they know how to deal with more complex interactions and less effective displays.

The first proper attempt to build an effective system was not a commercial success. This was the Xerox Star, a brainchild of the Xerox Corporation's Palo Alto Research Center. The developers recognized the importance of large, highly detailed display screens with plenty of graphics; they gave the machine the ability to have several different documents on the screen at the same time; and they introduced a pointing device—in this case, the "mouse"—for the user to specify a work area on the screen. The Xerox Star computer was a breakthrough in usable design.[19] But the system was too expensive and too slow. Users liked the power and the ease of operation, but they needed better performance. The benefits of easy to use commands were completely outweighed by the slow response speed. The display could not always keep up with typing, and requests for explanation (the "help" system) sometimes took so long that a user could go for a cup of coffee while waiting for an answer to even the simplest question. Xerox showed the way but suffered a common fate of pioneers: the spirit was willing but the implementation weak.

Fortunately for the consumer, the Apple Computer Company has followed through on Xerox's ideas, using the philosophy developed for *the Xerox Star (and hiring away some of Xerox's people) to produce first the Apple Lisa (also too slow and expensive and a failure in the marketplace) and then the Macintosh, a success story.*

The approach followed by Xerox has been well documented.[20] *The major goal was consistency of operations, to make things visible so that the available options could always be determined, and to test each idea with users at every step of the development process. These are all the important characteristics of good system design.*

Apple's Macintosh computer makes extensive use of visual displays. These eliminate the blank screen: the user can see what alternative actions are possible. The computer also makes the actions relatively easy to do, and it standardizes procedures so that methods learned for one program apply to most other programs. There is good feedback. Many actions are done by moving a mouse—a small, hand-held pointing device that causes a marker to move to the appropriate location on the screen. The mouse provides good mapping of action to result, and the use of menus—choices spelled out on the screen—makes the operations easy to perform. The Gulf of Execution and the Gulf of Evaluation are both securely bridged.

The Macintosh fails badly at many things, especially those for which it uses obscure combinations of keypresses to accomplish some task. Many of the problems arise from the use of the mouse. The mouse has one button, which simplifies its use but means that some actions must be specified by clicking the button several times or by simultaneously holding down various combinations of keys on the keyboard and clicking the mouse button. These actions violate the basic design philosophy. They are difficult to learn, difficult to remember, and difficult to do.

Ah, the buttons-on-the-mouse problem. How many buttons should the mouse have? Various models use one, two, or three, three being the most common number. Actually, some mice have more buttons; one design even has a chord keyboard on it. Fierce arguments rage over the correct number. The answer, of course, is that there is no correct answer. It is a tradeoff. Increase the number of buttons and you simplify some operations, but you also increase the complexity of the mapping problem. Even two buttons lead to an inconsistent mapping of functions to buttons. Reduce to one button and the mapping problem goes away, but so, too, does some of the functionality.

The Macintosh provides an example of what computer systems could be like. The design emphasizes visibility and feedback. Its "human interface guidelines" and its internal "toolbox" provide standards for the many programmers who design for it. It has emphasized consideration for the user. Yes, there are several serious drawbacks to the Macintosh: it is far from perfect. And it isn't unique. Still, for its relative success in making usability and understanding into primary design objectives, I'd give the Apple Macintosh a prize. If only I thought more of prizes.

COMPUTER AS CHAMELEON

The computer is unusual among machines in that its shape, form, and appearance are not fixed: they can be anything the designer wishes them to be. The computer can be like a chameleon, changing shape and outward appearance to match the situation. The operations of the computer can be soft, being done in appearance rather than substance. And the appearance can be reversed with a change of mind by the user. As users, we can create explorable systems that can be learned through experimentation, without fear of failure or damage. Furthermore, the computer can take on the appearance of the task; it can disappear behind a facade (its system image).

EXPLORABLE SYSTEMS: INVITING EXPERIMENTATION

One important method of making systems easier to learn and to use is to make them explorable, to encourage the user to experiment and learn the possibilities through active exploration. This is how many people learn about home appliances, or about a new stereo system, television set, or video game. Work the buttons while listening and looking to see what happens. The same can be true with computer systems. There are three requirements for a system to be explorable.

1. In each state of the system, the user must readily see and be able to do the allowable actions. The visibility acts as a suggestion, reminding the user of possibilities and inviting the exploration of new ideas and methods.
2. The effect of each action must be both visible and easy to interpret. This property allows users to learn the effects of each action, to

develop a good mental model of the system, and to learn the causal relationships between actions and outcomes. The system image plays a critical role in making such learning possible.

3. Actions should be without cost. When an action has an undesirable result, it must be readily reversible. This is especially important with computer systems. In the case of an irreversible action, the system should make clear what effect the contemplated action will have prior to its execution; there should be enough time to cancel the plan. Or the action should be difficult to do, nonexplorable. Most actions should be cost-free, explorable, discoverable.

TWO MODES OF COMPUTER USAGE

Compare two different ways of getting a task done. One way is to issue commands to someone else who does the actual work: call this "command mode" or "third-person" interaction. The other way is to do the operations yourself: call this "direct manipulation mode" or "first-person" interaction. The difference between these two is like the difference between being driven by a chauffeur and driving an automobile yourself. These two different modes exist with computers.[21]

Most computer systems offer command mode, third-person interactions. To use the computer, you type commands to it, using a special "command language" that you have to learn. Some computer systems offer direct manipulation, first-person interactions, good examples being the driving, flying, and sports games that are commonplace in arcades and on home machines. In these games, the feeling of direct control over the actions is an essential part of the task. This feeling of directness is also possible with everyday computer tasks, such as writing or bookkeeping. Spreadsheet programs and many text editors and word processing programs are good examples of direct manipulation systems used in business.

Both forms of interaction are needed. Third-person interaction is well suited for situations in which the job is laborious or repetitive, as well as those in which you can trust the system (or other person) to do the job for you properly. Sometimes it is nice to have a chauffeur. But if the job is critical, novel, or ill-specified, or if you do not yet know exactly what is to be done, then you need direct, first-person interaction. Now direct control is essential; an intermediary gets in the way.

But direct manipulation, first-person systems have their drawbacks. Although they are often easy to use, fun, and entertaining, it is often

difficult to do a really good job with them. They require the user to do the task directly, and the user may not be very good at it. Colored pencils and musical instruments are good examples of direct manipulation systems. But I, for one, am not a good artist or musician. When I want good art or good music, I need professional assistance. So, too, with many direct manipulation computer systems. I find that I often need first-person systems for which there is a backup intermediary, ready to take over when asked, available for advice when needed.

When I use a direct manipulation system—whether for text editing, drawing pictures, or creating and playing games—I do think of myself not as using a computer but as doing the particular task. The computer is, in effect, invisible. The point cannot be overstressed: make the computer system invisible. This principle can be applied with any form of system interaction, direct or indirect.

THE INVISIBLE COMPUTER OF THE FUTURE

Consider what the computer of the future might look like. Suppose I told you it wouldn't even be visible, that you wouldn't even know you were using one? What do I mean? Well, this is already true: you use computers when you use many modern automobiles, microwave ovens, and games. Or CD players and calculators. You don't notice the computer because you think of yourself as doing the task, not as using the computer.[22]

In the same sense, you don't go to the kitchen to use an electric motor; you go to use the refrigerator, or the blender, or the dishwasher. The motors are part of the task, even in the case of the blender, mixer, or food processor, which are essentially pure motors and the specialized attachments they drive.

The computer of the future is perhaps best illustrated by my imaginary perfect calendar. Suppose I am home one evening, deciding whether to accept an invitation to attend a conference next May. I pick up my appointment calendar and turn to the appropriate page. I tentatively decide that I can attend and pencil in the topic. The calendar flashes at me and displays a note reminding me that the university will still be in session during that period and that the trip overlaps my wife's birthday. I decide that the conference is important, so I make a note to check whether I can get someone to take over my classes and to see whether I can leave the conference early for the birthday. I close the *calendar and get back to other things. The next day, when I arrive at my office I find two notes on my message screen: one to find a substitute for my classes next May, the other to check with the conference organizers to see if I can leave early.*

This imaginary calendar looks like a calendar. It's about the size of a standard pad of paper, it opens up to display dates. But it really is a computer, so it can do things that today's appointment calendar cannot. It can, for example, present its information in different formats: it can display the pages compressed so that a whole year fits on one page; it can expand the display so that I see a single day in thirty-minute intervals. Because I frequently use my calendar in conjunction with my travels, the calendar is also an address book, notepad, and expense account record. Most important, it can also connect itself to my other systems (via a wireless infrared or electromagnetic channel). Thus, whatever I enter into the calendar gets transmitted to my office and home systems so that they are always in synchrony. If I make an appointment or change someone's address or telephone number on one system, the others get told. When I finish a trip, the expense record can be transferred to the expense account form. The computer is invisible, hidden beneath the surface; only the task is visible. Although I may actually be using a computer, I feel as if I am using my appointment calendar.

CHAPTER SEVEN

USER-CENTERED DESIGN

OFF THE LEASH By W.B. Park

"Darn these hooves! I hit the wrong switch again! Who designs these instrument panels, raccoons?"