deep consideration and study of human needs to the design process, whatever the product or service, whatever the major focus.

Fundamental Principles of Interaction

Great designers produce pleasurable experiences. *Experience*: note the word. Engineers tend not to like it; it is too subjective. But when I ask them about their favorite automobile or test equipment, they will smile delightedly as they discuss the fit and finish, the sensation of power during acceleration, their ease of control while shifting or steering, or the wonderful feel of the knobs and switches on the instrument. Those are experiences.

Experience is critical, for it determines how fondly people remember their interactions. Was the overall experience positive, or was it frustrating and confusing? When our home technology behaves in an uninterpretable fashion we can become confused, frustrated, and even angry—all strong negative emotions. When there is understanding it can lead to a feeling of control, of mastery, and of satisfaction or even pride—all strong positive emotions. Cognition and emotion are tightly intertwined, which means that the designers must design with both in mind.

When we interact with a product, we need to figure out how to work it. This means discovering what it does, how it works, and what operations are possible: discoverability. Discoverability results from appropriate application of five fundamental psychological concepts covered in the next few chapters: *affordances*, *signifiers*, *constraints*, *mappings*, and *feedback*. But there is a sixth principle, perhaps most important of all: the *conceptual model* of the system. It is the conceptual model that provides true understanding. So I now turn to these fundamental principles, starting with affordances, signifiers, mappings, and feedback, then moving to conceptual models. Constraints are covered in Chapters 3 and 4.

AFFORDANCES

We live in a world filled with objects, many natural, the rest artificial. Every day we encounter thousands of objects, many of them new to us. Many of the new objects are similar to ones we already

know, but many are unique, yet we manage quite well. How do we do this? Why is it that when we encounter many unusual natural objects, we know how to interact with them? Why is this true with many of the artificial, human-made objects we encounter? The answer lies with a few basic principles. Some of the most important of these principles come from a consideration of affordances.

The term *affordance* refers to the relationship between a physical object and a person (or for that matter, any interacting agent, whether animal or human, or even machines and robots). An affordance is a relationship between the properties of an object and the capabilities of the agent that determine just how the object could possibly be used. A chair affords ("is for") support and, therefore, affords sitting. Most chairs can also be carried by a single person (they afford lifting), but some can only be lifted by a strong person or by a team of people. If young or relatively weak people cannot lift a chair, then for these people, the chair does not have that affordance, it does not afford lifting.

The presence of an affordance is jointly determined by the qualities of the object and the abilities of the agent that is interacting. This relational definition of affordance gives considerable difficulty to many people. We are used to thinking that properties are associated with objects. But affordance is not a property. An affordance is a relationship. Whether an affordance exists depends upon the properties of both the object and the agent.

Glass affords transparency. At the same time, its physical structure blocks the passage of most physical objects. As a result, glass affords seeing through and support, but not the passage of air or most physical objects (atomic particles can pass through glass). The blockage of passage can be considered an anti-affordance—the prevention of interaction. To be effective, affordances and anti-affordances have to be discoverable—perceivable. This poses a difficulty with glass. The reason we like glass is its relative invisibility, but this aspect, so useful in the normal window, also hides its anti-affordance property of blocking passage. As a result, birds often try to fly through windows. And every year, numerous people injure themselves when they walk (or run) through closed glass doors or large picture windows. If an affordance or anti-affordance cannot be perceived, some means of signaling its presence is required: I call this property a *signifier* (discussed in the next section).

The notion of affordance and the insights it provides originated with J. J. Gibson, an eminent psychologist who provided many advances to our understanding of human perception. I had interacted with him over many years, sometimes in formal conferences and seminars, but most fruitfully over many bottles of beer, late at night, just talking. We disagreed about almost everything. I was an engineer who became a cognitive psychologist, trying to understand how the mind works. He started off as a Gestalt psychologist, but then developed an approach that is today named after him: Gibsonian psychology, an ecological approach to perception. He argued that the world contained the clues and that people simply picked them up through "direct perception." I argued that nothing could be direct: the brain had to process the information arriving at the sense organs to put together a coherent interpretation. "Nonsense," he loudly proclaimed; "it requires no interpretation: it is directly perceived." And then he would put his hand to his ears, and with a triumphant flourish, turn off his hearing aids: my counterarguments would fall upon deaf ears—literally.

When I pondered my question—how do people know how to act when confronted with a novel situation—I realized that a large part of the answer lay in Gibson's work. He pointed out that all the senses work together, that we pick up information about the world by the combined result of all of them. "Information pickup" was one of his favorite phrases, and Gibson believed that the combined information picked up by all of our sensory apparatus—sight, sound, smell, touch, balance, kinesthetic, acceleration, body position—determines our perceptions without the need for internal processing or cognition. Although he and I disagreed about the role played by the brain's internal processing, his brilliance was in focusing attention on the rich amount of information present in the world. Moreover, the physical objects conveyed important information about how people could interact with them, a property he named "affordance."

Affordances exist even if they are not visible. For designers, their visibility is critical: visible affordances provide strong clues to the operations of things. A flat plate mounted on a door affords pushing. Knobs afford turning, pushing, and pulling. Slots are for inserting things into. Balls are for throwing or bouncing. Perceived affordances help people figure out what actions are possible without the need for labels or instructions. I call the signaling component of affordances *signifiers.*

SIGNIFIERS

Are affordances important to designers? The first edition of this book introduced the term *affordances* to the world of design. The design community loved the concept and affordances soon propagated into the instruction and writing about design. I soon found mention of the term everywhere. Alas, the term became used in ways that had nothing to do with the original.

Many people find affordances difficult to understand because they are relationships, not properties. Designers deal with fixed properties, so there is a temptation to say that the property is an affordance. But that is not the only problem with the concept of affordances.

Designers have practical problems. They need to know how to design things to make them understandable. They soon discovered that when working with the graphical designs for electronic displays, they needed a way to designate which parts could be touched, slid upward, downward, or sideways, or tapped upon. The actions could be done with a mouse, stylus, or fingers. Some systems responded to body motions, gestures, and spoken words, with no touching of any physical device. How could designers describe what they were doing? There was no word that fit, so they took the closest existing word—*affordance.* Soon designers were saying such things as, "I put an affordance there," to describe why they displayed a circle on a screen to indicate where the person should touch, whether by mouse or by finger. "No," I said, "that is not an affordance. That is a way of communicating where the touch should be. You are communicating where to do the touching: the affordance of touching exists on the entire screen: you are trying to signify *where* the touch should take place. That's not the same thing as saying *what* action is possible."

Not only did my explanation fail to satisfy the design community, but I myself was unhappy. Eventually I gave up: designers needed a word to describe what they were doing, so they chose *affordance*. What alternative did they have? I decided to provide a better answer: *signifiers*. Affordances determine what actions are possible. Signifiers communicate where the action should take place. We need both.

People need some way of understanding the product or service they wish to use, some sign of what it is for, what is happening, and what the alternative actions are. People search for clues, for any sign that might help them cope and understand. It is the sign that is important, anything that might signify meaningful information. Designers need to provide these clues. What people need, and what designers must provide, are signifiers. Good design requires, among other things, good communication of the purpose, structure, and operation of the device to the people who use it. That is the role of the signifier.

The term *signifier* has had a long and illustrious career in the exotic field of semiotics, the study of signs and symbols. But just as I appropriated *affordance* to use in design in a manner somewhat different than its inventor had intended, I use *signifier* in a somewhat different way than it is used in semiotics. For me, the term *signifier* refers to any mark or sound, any perceivable indicator that communicates appropriate behavior to a person.

Signifiers can be deliberate and intentional, such as the sign PUSH on a door, but they may also be accidental and unintentional, such as our use of the visible trail made by previous people walking through a field or over a snow-covered terrain to determine the best path. Or how we might use the presence or absence of people waiting at a train station to determine whether we have missed the train. (I explain these ideas in more detail in my book *Living with Complexity*.)

A.

B.

C.

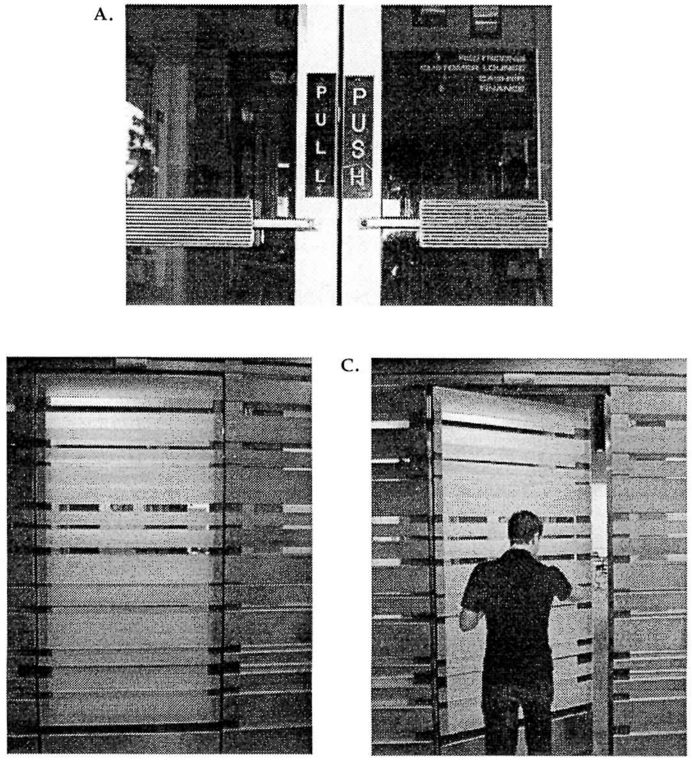


FIGURE 1.2. Problem Doors: Signifiers Are Needed. Door hardware can signal whether to push or pull without signs, but the hardware of the two doors in the upper photo, A, are identical even though one should be pushed, the other pulled. The flat, ribbed horizontal bar has the obvious perceived affordance of pushing, but as the signs indicate, the door on the left is to be pulled, the one on the right is to be pushed. In the bottom pair of photos, B and C, there are no visible signifiers or affordances. How does one know which side to push? Trial and error. When external signifiers—signs—have to be added to something as simple as a door, it indicates bad design. (Photographs by the author.)

The signifier is an important communication device to the recipient, whether or not communication was intended. It doesn't matter whether the useful signal was deliberately placed or whether it is incidental: there is no necessary distinction. Why should it matter whether a flag was placed as a deliberate clue to wind direction (as is done at airports or on the masts of sailboats) or was there as an advertisement or symbol of pride in one's country (as is done on public buildings). Once I interpret a flag's motion to indicate wind direction, it does not matter why it was placed there.

Consider a bookmark, a deliberately placed signifier of one's place in reading a book. But the physical nature of books also makes a bookmark an accidental signifier, for its placement also indicates how much of the book remains. Most readers have learned to use this accidental signifier to aid in their enjoyment of the reading. With few pages left, we know the end is near. And if the reading is torturous, as in a school assignment, one can always console oneself by knowing there are "only a few more pages to get through." Electronic book readers do not have the physical structure of paper books, so unless the software designer deliberately provides a clue, they do not convey any signal about the amount of text remaining.

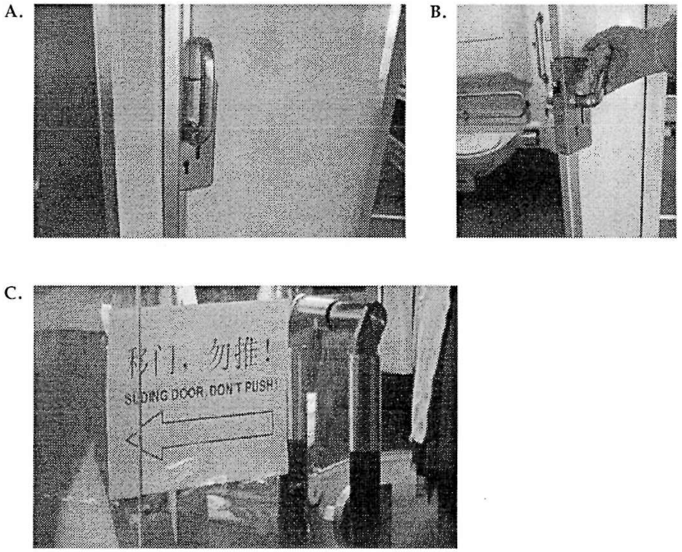


FIGURE 1.3. **Sliding Doors: Seldom Done Well.** Sliding doors are seldom signified properly. The top two photographs show the sliding door to the toilet on an Amtrak train in the United States. The handle clearly signifies "pull," but in fact, it needs to be rotated and the door slid to the right. The owner of the store in Shanghai, China, Photo C, solved the problem with a sign. "DON'T PUSH!" it says, in both English and Chinese. Amtrak's toilet door could have used a similar kind of sign. (Photographs by the author.)

Whatever their nature, planned or accidental, signifiers provide valuable clues as to the nature of the world and of social activities. For us to function in this social, technological world, we need to develop internal models of what things mean, of how they operate. We seek all the clues we can find to help in this enterprise, and in this way, we are detectives, searching for whatever guidance we might find. If we are fortunate, thoughtful designers provide the clues for us. Otherwise, we must use our own creativity and imagination.

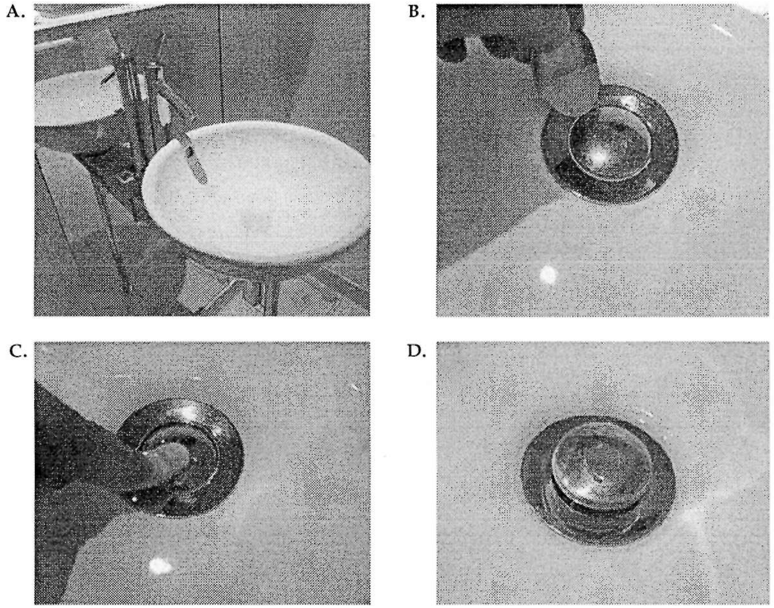
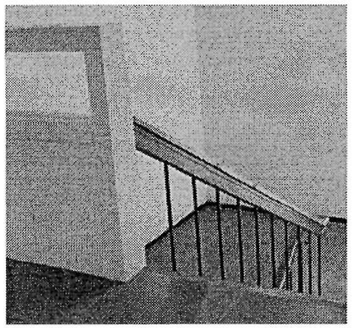


FIGURE 1.4. **The Sink That Would Not Drain: Where Signifiers Fail.** I washed my hands in my hotel sink in London, but then, as shown in Photo A, was left with the question of how to empty the sink of the dirty water. I searched all over for a control: none. I tried prying open the sink stopper with a spoon (Photo B): failure. I finally left my hotel room and went to the front desk to ask for instructions. (Yes, I actually did.) "Push down on the stopper," I was told. Yes, it worked (Photos C and D). But how was anyone to ever discover this? And why should I have to put my clean hands back into the dirty water to empty the sink? The problem here is not just the lack of signifier, it is the faulty decision to produce a stopper that requires people to dirty their clean hands to use it. (Photographs by the author.)

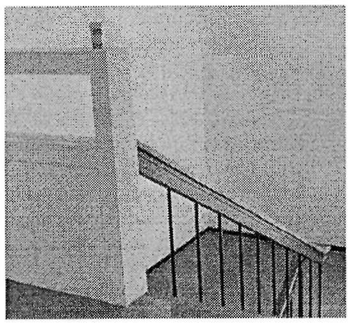
Affordances, perceived affordances, and signifiers have much in common, so let me pause to ensure that the distinctions are clear.

Affordances represent the possibilities in the world for how an agent (a person, animal, or machine) can interact with something. Some affordances are perceivable, others are invisible. Signifiers are signals. Some signifiers are signs, labels, and drawings placed in the world, such as the signs labeled "push," "pull," or "exit" on doors, or arrows and diagrams indicating what is to be acted upon or in which direction to gesture, or other instructions. Some signifiers are simply the perceived affordances, such as the handle of a door or the physical structure of a switch. Note that some perceived affordances may not be real: they may look like doors or places to push, or an impediment to entry, when in fact they are not. These are misleading signifiers, oftentimes accidental but sometimes purposeful, as when trying to keep people from doing actions for which they are not qualified, or in games, where one of the challenges is to figure out what is real and what is not.

A.



B.



C.

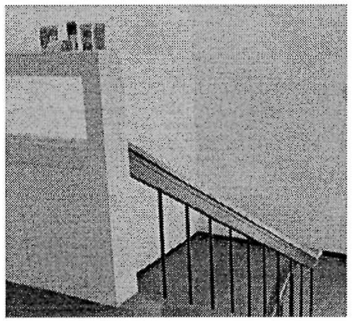


FIGURE 1.5. **Accidental Affordances Can Become Strong Signifiers.** This wall, at the Industrial Design department of KAIST, in Korea, provides an anti-affordance, preventing people from falling down the stair shaft. Its top is flat, an accidental by-product of the design. But flat surfaces afford support, and as soon as one person discovers it can be used to dispose of empty drink containers, the discarded container becomes a signifier, telling others that it is permissible to discard their items there. (Photographs by the author.)

My favorite example of a misleading signifier is a row of vertical pipes across a service road that I once saw in a public park. The pipes obviously blocked cars and trucks from driving on that road: they were good examples of anti-affordances. But to my great surprise, I saw a park vehicle simply go through the pipes. Huh? I walked over and examined them: the pipes were made of rubber, so vehicles could simply drive right over them. A very clever signifier, signaling a blocked road (via an apparent anti-affordance) to the average person, but permitting passage for those who knew.

To summarize:

- Affordances are the possible interactions between people and the environment. Some affordances are perceivable, others are not.
- Perceived affordances often act as signifiers, but they can be ambiguous.
- Signifiers signal things, in particular what actions are possible and how they should be done. Signifiers must be perceivable, else they fail to function.

In design, signifiers are more important than affordances, for they communicate how to use the design. A signifier can be words, a graphical illustration, or just a device whose perceived affordances are unambiguous. Creative designers incorporate the signifying part of the design into a cohesive experience. For the most part, designers can focus upon signifiers.

Because affordances and signifiers are fundamentally important principles of good design, they show up frequently in the pages of this book. Whenever you see hand-lettered signs pasted on doors, switches, or products, trying to explain how to work them, what to do and what not to do, you are also looking at poor design.

AFFORDANCES AND SIGNIFIERS: A CONVERSATION

A designer approaches his mentor. He is working on a system that recommends restaurants to people, based upon their preferences and those of their friends. But in his tests, he discovered that people never used all of the features. "Why not?" he asks his mentor.

(With apologies to Socrates.)

DESIGNER	MENTOR
I'm frustrated; people aren't using our application properly.	Can you tell me about it?
The screen shows the restaurant that we recommend. It matches their preferences, and their friends like it as well. If they want to see other recommendations, all they have to do is swipe left or right. To learn more about a place, just swipe up for a menu or down to see if any friends are there now. People seem to find the other recommendations, but not the menus or their friends? I don't understand.	Why do you think this might be?
I don't know. Should I add some affordances? Suppose I put an arrow on each edge and add a label saying what they do.	That is very nice. But why do you call these affordances? They could already do the actions. Weren't the affordances already there?
Yes, you have a point. But the affordances weren't visible. I made them visible.	Very true. You added a signal of what to do.
Yes, isn't that what I said?	Not quite—you called them affordances even though they afford nothing new: they signify what to do and where to do it. So call them by their right name: *"signifiers."*
Oh, I see. But then why do designers care about affordances? Perhaps we should focus our attention on signifiers.	You speak wisely. Communication is a key to good design. And a key to communication is the signifier.
Oh. Now I understand my confusion. Yes, a signifier is what signifies. It is a sign. Now it seems perfectly obvious.	Profound ideas are always obvious once they are understood.

MAPPING

Mapping is a technical term, borrowed from mathematics, meaning the relationship between the elements of two sets of things. Suppose there are many lights in the ceiling of a classroom or auditorium and a row of light switches on the wall at the front of the

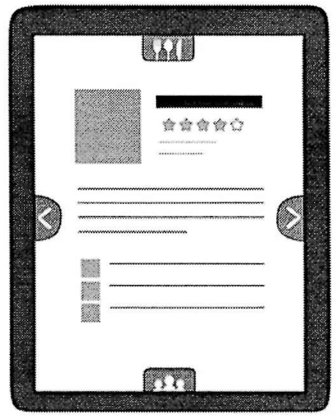


FIGURE 1.6. Signifiers on a Touch Screen. The arrows and icons are signifiers: they provide signals about the permissible operations for this restaurant guide. Swiping left or right brings up new restaurant recommendations. Swiping up reveals the menu for the restaurant being displayed; swiping down, friends who recommend the restaurant.

room. The mapping of switches to lights specifies which switch controls which light.

Mapping is an important concept in the design and layout of controls and displays. When the mapping uses spatial correspondence between the layout of the controls and the devices being controlled, it is easy to determine how to use them. In steering a car, we rotate the steering wheel clockwise to cause the car to turn right: the top of the wheel moves in the same direction as the car. Note that other choices could have been made. In early cars, steering was controlled by a variety of devices, including tillers, handlebars, and reins. Today, some vehicles use joysticks, much as in a computer game. In cars that used tillers, steering was done much as one steers a boat: move the tiller to the left to turn to the right. Tractors, construction equipment such as bulldozers and cranes, and military tanks that have tracks instead of wheels use separate controls for the speed and direction of each track: to turn right, the left track is increased in speed, while the right track is slowed or even reversed. This is also how a wheelchair is steered.

All of these mappings for the control of vehicles work because each has a compelling conceptual model of how the operation of the control affects the vehicle. Thus, if we speed up the left wheel of a wheelchair while stopping the right wheel, it is easy to imagine the chair's pivoting on the right wheel, circling to the right. In a small boat, we can understand the tiller by realizing that pushing the tiller to the left causes the ship's rudder to move to the right and the resulting force of the water on the rudder slows down the right side of the boat, so that the boat rotates to the right. It doesn't matter whether these conceptual models are accurate: what matters is that they provide a clear way of remembering and understanding the mappings. The relationship between a control and its results is easiest to learn wherever there is an understandable mapping between the controls, the actions, and the intended result.

Natural mapping, by which I mean taking advantage of spatial analogies, leads to immediate understanding. For example, to move an object up, move the control up. To make it easy to determine which control works which light in a large room or auditorium, arrange the controls in the same pattern as the lights. Some natural mappings are cultural or biological, as in the universal standard that moving the hand up signifies more, moving it down signifies less, which is why it is appropriate to use vertical position to represent intensity or amount. Other natural mappings follow from the principles of perception and allow for the natural grouping or patterning of controls and feedback. Groupings and proximity are important principles from Gestalt psychology that can be used to map controls to function: related controls should be grouped together. Controls should be close to the item being controlled.

Note that there are many mappings that feel "natural" but in fact are specific to a particular culture: what is natural for one culture is not necessarily natural for another. In Chapter 3, I discuss how

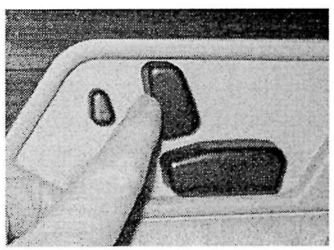


FIGURE 1.7. **Good Mapping: Automobile Seat Adjustment Control.** This is an excellent example of natural mapping. The control is in the shape of the seat itself: the mapping is straightforward. To move the front edge of the seat higher, lift up on the front part of the button. To make the seat back recline, move the button back. The same principle could be applied to much more common objects. This particular control is from Mercedes-Benz, but this form of mapping is now used by many automobile companies. (Photograph by the author.)

different cultures view time, which has important implications for some kinds of mappings.

A device is easy to use when the set of possible actions is visible, when the controls and displays exploit natural mappings. The principles are simple but rarely incorporated into design. Good design takes care, planning, thought, and an understanding of how people behave.

FEEDBACK

Ever watch people at an elevator repeatedly push the Up button, or repeatedly push the pedestrian button at a street crossing? Ever drive to a traffic intersection and wait an inordinate amount of time for the signals to change, wondering all the time whether the detection circuits noticed your vehicle (a common problem with bicycles)? What is missing in all these cases is feedback: some way of letting you know that the system is working on your request.

Feedback—communicating the results of an action—is a well-known concept from the science of control and information theory. Imagine trying to hit a target with a ball when you cannot see the target. Even as simple a task as picking up a glass with the hand requires feedback to aim the hand properly, to grasp the glass, and to lift it. A misplaced hand will spill the contents, too hard a grip will break the glass, and too weak a grip will allow it to fall. The human nervous system is equipped with numerous feedback mechanisms, including visual, auditory, and touch sensors, as well as vestibular and proprioceptive systems that monitor body position and muscle and limb movements. Given the importance of feedback, it is amazing how many products ignore it.

Feedback must be immediate: even a delay of a tenth of a second can be disconcerting. If the delay is too long, people often give up, going off to do other activities. This is annoying to the people, but it can also be wasteful of resources when the system spends considerable time and effort to satisfy the request, only to find that the intended recipient is no longer there. Feedback must also be informative. Many companies try to save money by using inexpensive lights or sound generators for feedback. These simple light flashes or beeps are usually more annoying than useful. They tell us that something has happened, but convey very little information about what has happened, and then nothing about what we should do about it. When the signal is auditory, in many cases we cannot even be certain which device has created the sound. If the signal is a light, we may miss it unless our eyes are on the correct spot at the correct time. Poor feedback can be worse than no feedback at all, because it is distracting, uninformative, and in many cases irritating and anxiety-provoking.

Too much feedback can be even more annoying than too little. My dishwasher likes to beep at three a.m. to tell me that the wash is done, defeating my goal of having it work in the middle of the night so as not to disturb anyone (and to use less expensive electricity). But worst of all is inappropriate, uninterpretable feedback. The irritation caused by a "backseat driver" is well enough known that it is the staple of numerous jokes. Backseat drivers are often correct, but their remarks and comments can be so numerous and continuous that instead of helping, they become an irritating distraction. Machines that give too much feedback are like backseat drivers. Not only is it distracting to be subjected to continual flashing lights, text announcements, spoken voices, or beeps and boops, but it can be dangerous. Too many announcements cause people to ignore all of them, or wherever possible, disable all of them, which means that critical and important ones are apt to be missed. Feedback is essential, but not when it gets in the way of other things, including a calm and relaxing environment.

Poor design of feedback can be the result of decisions aimed at reducing costs, even if they make life more difficult for people. Rather than use multiple signal lights, informative displays, or rich, musical sounds with varying patterns, the focus upon cost reduction forces the design to use a single light or sound to convey multiple types of information. If the choice is to use a light, then one flash might mean one thing; two rapid flashes, something else. A long flash might signal yet another state; and a long flash followed by a brief one, yet another. If the choice is to use a sound, quite often the least expensive sound device is selected, one that

can only produce a high-frequency beep. Just as with the lights, the only way to signal different states of the machine is by beeping different patterns. What do all these different patterns mean? How can we possibly learn and remember them? It doesn't help that every different machine uses a different pattern of lights or beeps, sometimes with the same patterns meaning contradictory things for different machines. All the beeps sound alike, so it often isn't even possible to know which machine is talking to us.

Feedback has to be planned. All actions need to be confirmed, but in a manner that is unobtrusive. Feedback must also be prioritized, so that unimportant information is presented in an unobtrusive fashion, but important signals are presented in a way that does capture attention. When there are major emergencies, then even important signals have to be prioritized. When every device is signaling a major emergency, nothing is gained by the resulting cacophony. The continual beeps and alarms of equipment can be dangerous. In many emergencies, workers have to spend valuable time turning off all the alarms because the sounds interfere with the concentration required to solve the problem. Hospital operating rooms, emergency wards. Nuclear power control plants. Airplane cockpits. All can become confusing, irritating, and life-endangering places because of excessive feedback, excessive alarms, and incompatible message coding. Feedback is essential, but it has to be done correctly. Appropriately.

CONCEPTUAL MODELS

A conceptual model is an explanation, usually highly simplified, of how something works. It doesn't have to be complete or even accurate as long as it is useful. The files, folders, and icons you see displayed on a computer screen help people create the conceptual model of documents and folders inside the computer, or of apps or applications residing on the screen, waiting to be summoned. In fact, there are no folders inside the computer—those are effective conceptualizations designed to make them easier to use. Sometimes these depictions can add to the confusion, however. When reading e-mail or visiting a website, the material appears to be on the device, for that is where it is displayed and manipulated. But in fact, in many cases the actual material is "in the cloud," located on some distant machine. The conceptual model is of one, coherent image, whereas it may actually consist of parts, each located on different machines that could be almost anywhere in the world. This simplified model is helpful for normal usage, but if the network connection to the cloud services is interrupted, the result can be confusing. Information is still on their screen, but users can no longer save it or retrieve new things: their conceptual model offers no explanation. Simplified models are valuable only as long as the assumptions that support them hold true.

There are often multiple conceptual models of a product or device. People's conceptual models for the way that regenerative braking in a hybrid or electrically powered automobile works are quite different for average drivers than for technically sophisticated drivers, different again for whoever must service the system, and yet different again for those who designed the system.

Conceptual models found in technical manuals and books for technical use can be detailed and complex. The ones we are concerned with here are simpler: they reside in the minds of the people who are using the product, so they are also "mental models." Mental models, as the name implies, are the conceptual models in people's minds that represent their understanding of how things work. Different people may hold different mental models of the same item. Indeed, a single person might have multiple models of the same item, each dealing with a different aspect of its operation: the models can even be in conflict.

Conceptual models are often inferred from the device itself. Some models are passed on from person to person. Some come from manuals. Usually the device itself offers very little assistance, so the model is constructed by experience. Quite often these models are erroneous, and therefore lead to difficulties in using the device.

The major clues to how things work come from their perceived structure—in particular from signifiers, affordances, constraints, and mappings. Hand tools for the shop, gardening, and the house tend to make their critical parts sufficiently visible that concep-

FIGURE 1.8. Junghans Mega 1000 Digital Radio Controlled Watch. There is no good conceptual model for understanding the operation of my watch. It has five buttons with no hints as to what each one does. And yes, the buttons do different things in their different modes. But it is a very nice-looking watch, and always has the exact time because it checks official radio time stations. (The top row of the display is the date: Wednesday, February 20, the eighth week of the year.) (Photograph by the author.)

tual models of their operation and function are readily derived. Consider a pair of scissors: 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 both affordances—they allow the fingers to be inserted—and signifiers—they indicate where the fingers are to go. The sizes of the holes provide constraints to limit the possible fingers: a big hole suggests several fingers; a small hole, only one. The mapping between holes and fingers—the set of possible operations—is signified and constrained by the holes. Moreover, the operation is not sensitive to finger placement: if you use the wrong fingers (or the wrong hand), the scissors still work, although not as comfortably. You can figure out the scissors because their operating parts are visible and the implications clear. The conceptual model is obvious, and there is effective use of signifiers, affordances, and constraints.

What happens when the device does not suggest a good conceptual model? Consider my digital watch with five buttons: two along the top, two along the bottom, and one on the left side (Figure 1.8). What is each button 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. Moreover, the buttons have multiple ways of being used. Two of the buttons do different things when pushed quickly or when kept depressed for several seconds. Some operations require simultaneous depression of several of the buttons. The only way to tell how to work the watch is to read the manual, over and over again. With the scissors, moving the handle makes the blades move. The watch provides no visible relationship between the buttons and the possible actions, no discernible relationship between the actions and the end results. I really like the watch: too bad I can't remember all the functions.

Conceptual models are valuable in providing understanding, in predicting how things will behave, and in figuring out what to do when things do not go as planned. 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, just 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.

I used to own an ordinary, two-compartment refrigerator—nothing very fancy about it. The problem was that I couldn't set the temperature properly. There were only two things to do: adjust the temperature of the freezer compartment and adjust the tempera-

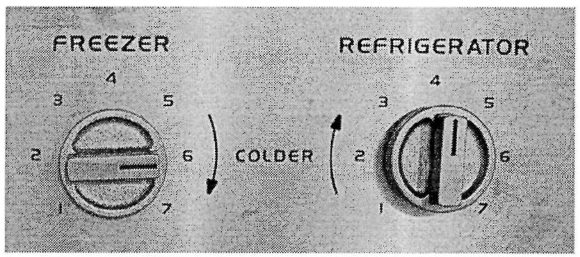


FIGURE 1.9. Refrigerator Controls. Two compartments—fresh food and freezer—and two controls (in the fresh food unit). Your task: Suppose the freezer is too cold, the fresh food section just right. How would you adjust the controls so as to make the freezer warmer and keep the fresh food the same? (Photograph by the author.)

ture of the fresh food compartment. And there were two controls, one labeled "freezer," the other "refrigerator." What's the problem?

Oh, perhaps I'd better warn you. The two controls are not independent. The freezer control also affects the fresh food temperature, and the fresh food control also affects the freezer. Moreover, the manual warns that one should "always allow twenty-four (24) hours for the temperature to stabilize whether setting the controls for the first time or making an adjustment."

It was extremely difficult to regulate the temperature of my old refrigerator. Why? Because the controls suggest a false conceptual model. Two compartments, two controls, which implies that each control is responsible for the temperature of the compartment that carries its name: this conceptual model is shown in Figure 1.10A. It is 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: this conceptual model is shown in Figure 1.10B. In addition, there must be a temperature sensor, but there is no way of knowing where it is located. With the conceptual model suggested by the controls,

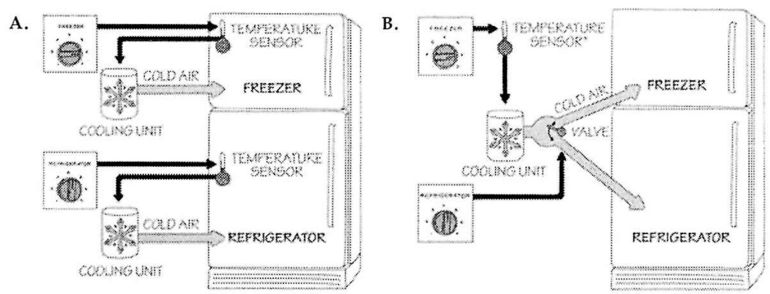


FIGURE 1.10. Two Conceptual Models for a Refrigerator. The conceptual model A is provided by the system image of the refrigerator as gleaned from the controls. Each control determines the temperature of the named part of the refrigerator. This means that each compartment has its own temperature sensor and cooling unit. This is wrong. The correct conceptual model is shown in B. There is no way of knowing where the temperature sensor is located so it is shown outside the refrigerator. The freezer control determines the freezer temperature (so is this where the sensor is located?). The refrigerator control determines how much of the cold air goes to the freezer and how much to the refrigerator.

adjusting the temperatures is almost impossible and always frustrating. Given the correct model, life would be much easier.

Why did the manufacturer suggest the wrong conceptual model? We will never know. In the twenty-five years since the publication of the first edition of this book, I have had many letters from people thanking me for explaining their confusing refrigerator, but never any communication from the manufacturer (General Electric). 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 was impossible to set the controls. And even though I am convinced I knew the correct model, I still couldn't accurately adjust the temperatures because the refrigerator design made it impossible to discover which control was for the temperature sensor, which for the relative proportion of cold air, and in which compartment the sensor was located. The lack of immediate feedback for the actions did not help: it took twenty-four hours to see whether the new setting was appropriate. I shouldn't have to keep a laboratory notebook and do controlled experiments just to set the temperature of my refrigerator.

I am happy to say that I no longer own that refrigerator. Instead I have one that has two separate controls, one in the fresh food compartment, one in the freezer compartment. Each control is nicely calibrated in degrees and labeled with the name of the compartment it controls. The two compartments are independent: setting the temperature in one has no effect on the temperature in the other. This solution, although ideal, does cost more. But far less expensive solutions are possible. With today's inexpensive sensors and motors, it should be possible to have a single cooling unit with a motor-controlled valve controlling the relative proportion of cold air diverted to each compartment. A simple, inexpensive computer chip could regulate the cooling unit and valve position so that the temperatures in the two compartments match their targets. A bit more work for the engineering design team? Yes, but the results would be worth it. Alas, General Electric is still selling refrigerators with the very same controls and mechanisms that cause so much confusion. The photograph in Figure 1.9 is from a contemporary refrigerator, photographed in a store while preparing this book.

The System Image

People create mental models of themselves, others, the environment, and the things with which they interact. These are conceptual models formed through experience, training, and instruction. These models serve as guides to help achieve our goals and in understanding the world.

How do we form an appropriate conceptual model for the devices we interact with? We cannot talk to the designer, so we rely upon whatever information is available to us: what the device looks like, what we know from using similar things in the past, what was told to us in the sales literature, by salespeople and advertisements, by articles we may have read, by the product website and instruction manuals. I call the combined information available to us the *system image.* 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.

As illustrated in Figure 1.11, the designer of the product and the person using the product form somewhat disconnected vertices of a triangle. The designer's conceptual model is the designer's conception of the product, occupying one vertex of the triangle. The product itself is no longer with the designer, so it is isolated as a second vertex, perhaps sitting on the user's kitchen counter. The system image is what can be perceived from the physical structure that has been built (including documentation, instructions, signifiers, and any information available from websites and help lines). The user's conceptual model comes from the system image, through interaction with the product, reading, searching for online information, and from whatever manuals are provided. The designer expects the user's model to be identical to the design model, but because designers cannot communicate directly with users, the entire burden of communication is on the system image.

FIGURE 1.11. The Designer's Model, the User's Model, and the System Image. The designer's conceptual model is the designer's conception of the look, feel, and operation of a product. The system image is what can be derived from the physical structure that has been built (including documentation). The user's mental model is developed through interaction with the product and the system image. Designers expect the user's model to be identical to their own, but because they cannot communicate directly with the user, the burden of communication is with the system image.

Figure 1.11 indicates why communication is such an important aspect of good design. No matter how brilliant the product, if people cannot use it, it will receive poor reviews. It is up to the designer to provide the appropriate information to make the product understandable and usable. Most important is the provision of a good conceptual model that guides the user when thing go wrong. With a good conceptual model, people can figure out what has happened and correct the things that went wrong. Without a good model, they struggle, often making matters worse.

Good conceptual models are the key to understandable, enjoyable products: good communication is the key to good conceptual models.

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 increase our difficulty and frustration with technology. The design problem posed by technological advances is enormous. Consider the wristwatch. A few decades ago, watches were simple. All you had to do was set the time and keep the watch wound. The standard control was the stem: a knob at the side of the watch. Turning the knob would wind the spring that provided power to the watch movement. Pulling out the knob and turning it rotated the hands. The operations were easy to learn and easy to do. There was a reasonable relationship between the

KNOWING WHAT TO DO: CONSTRAINTS, DISCOVERABILITY, AND FEEDBACK

How do we determine how to operate something that we have never seen before? We have no choice but to combine knowledge in the world with that in the head. Knowledge in the world includes perceived affordances and signifiers, the mappings between the parts that appear to be controls or places to manipulate and the resulting actions, and the physical constraints that limit what can be done. Knowledge in the head includes conceptual models; cultural, semantic, and logical constraints on behavior; and analogies between the current situation and previous experiences with other situations. Chapter 3 was devoted to a discussion of how we acquire knowledge and use it. There, the major emphasis was upon the knowledge in the head. This chapter focuses upon the knowledge in the world: how designers can provide the critical information that allows people to know what to do, even when experiencing an unfamiliar device or situation.

Let me illustrate with an example: building a motorcycle from a Lego set (a children's construction toy). The Lego motorcycle shown in Figure 4.1 has fifteen pieces, some rather specialized. Of those fifteen pieces, only two pairs are alike—two rectangles with the word *police* on them, and the two hands of

A. B.

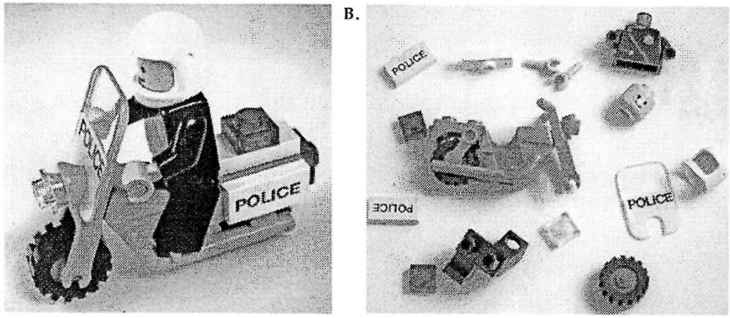


FIGURE 4.1. **Lego Motorcycle.** The toy Lego motorcycle is shown assembled (A) and in pieces (B). It has fifteen pieces so cleverly constructed that even an adult can put them together. The design exploits constraints to specify just which pieces fit where. Physical constraints limit alternative placements. Cultural and semantic constraints provide the necessary clues for further decisions. For example, cultural constraints dictate the placement of the three lights (red, blue, and yellow) and semantic constraints stop the user from putting the head backward on the body or the pieces labeled "police" upside down.

the policeman. Other pieces match one another in size and shape but are different colors. So, a number of the pieces are physically interchangeable—that is, the physical constraints are not sufficient to identify where they go—but the appropriate role for every single piece of the motorcycle is still unambiguously determined. How? By combining cultural, semantic, and logical constraints with the physical ones. As a result, it is possible to construct the motorcycle without any instructions or assistance.

In fact, I did the experiment. I asked people to put together the parts; they had never seen the finished structure and were not even told that it was a motorcycle (although it didn't take them long to figure this out). Nobody had any difficulty.

The visible 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. Cultural and semantic constraints provided strong restrictions on what would make sense for all but one of the remaining pieces, and with just one piece left and only one place it could possibly go, simple logic dictated the

placement. These four classes of constraints—physical, cultural, semantic, and logical—seem to be universal, appearing in a wide variety of situations.

Constraints are powerful clues, limiting the set of possible actions. The thoughtful use of constraints in design lets people readily determine the proper course of action, even in a novel situation.

Four Kinds of Constraints: Physical, Cultural, Semantic, and Logical

PHYSICAL CONSTRAINTS

Physical limitations constrain possible operations. Thus, a large peg cannot fit into a small hole. With the Lego motorcycle, the windshield would fit in only one place. 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, a physical constraint prevents a wrong action from succeeding only after it has been tried.

The traditional cylindrical battery, Figure 4.2A, lacks sufficient physical constraints. It can be put into battery compartments in two orientations: one that is correct, the other of which can damage the equipment. The instructions in Figure 4.2B show that polarity is important, yet the inferior signifiers inside the battery compartment makes it very difficult to determine the proper orientation for the batteries.

Why not design a battery with which it would be impossible to make an error: use physical constraints so that the battery will fit only if properly oriented. Alternatively, design the battery or the electrical contacts so that orientation doesn't matter.

Figure 4.3 shows a battery that has been designed so that orientation is irrelevant. Both ends of the battery are identical, with the

A. B.

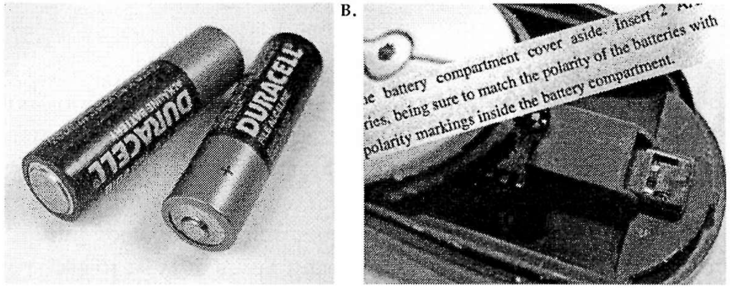
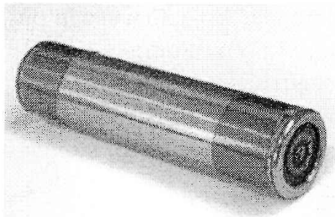


FIGURE 4.2. Cylindrical Battery: Where Constraints Are Needed. Figure A shows the traditional cylindrical battery that requires correct orientation in the slot to work properly (and to avoid damaging the equipment). But look at Figure B, which shows where two batteries are to be installed. The instructions from the manual are shown as an overlay to the photograph. They seem simple, but can you see into the dark recess to figure out which end of each battery goes where? Nope. The lettering is black against black: slightly raised shapes in the dark plastic.

FIGURE 4.3. Making Battery Orientation Irrelevant. This photograph shows a battery whose orientation doesn't matter; it can be inserted into the equipment in either possible direction. How? Each end of the battery has the same three concentric rings, with the center one on both ends being the "plus" terminal and the middle one being the "minus" terminal.



positive and negative terminals for the battery being its center and middle rings, respectively. The contact for the positive polarity is designed so it contacts only the center ring. Similarly, the contact for negative polarity touches only the middle ring. Although this seems to solve the problem, I have only seen this one example of such a battery: they are not widely available or used.

Another alternative is to invent battery contacts that allow our existing cylindrical batteries to be inserted in either orientation yet still work properly: Microsoft has invented this kind of contact, which it calls InstaLoad, and is attempting to convince equipment manufacturers to use it.

A third alternative is to design the shape of the battery so that it can fit in only one way. Most plug-in components do this well, using shapes, notches, and protrusions to constrain insertion

to a single orientation. So why can't our everyday batteries be the same?

Why does inelegant design persist for so long? This is called the *legacy problem*, and it will come up several times in this book. Too many devices use the existing standard—that is the legacy. If the symmetrical cylindrical battery were changed, there would also have to be a major change in a huge number of products. The new batteries would not work in older equipment, nor the old batteries in new equipment. Microsoft's design of contacts would allow us to continue to use the same batteries we are used to, but the products would have to switch to the new contacts. Two years after Microsoft's introduction of InstaLoad, despite positive press, I could find no products that use them—not even Microsoft products.

Locks and keys suffer from a similar problem. Although it is usually easy to distinguish the smooth top part of a key from its jagged underside, it is difficult to tell from the lock just which orientation of the key is required, especially in dark environments. Many electrical and electronic plugs and sockets have the same problem. Although they do have physical constraints to prevent improper insertion, it is often extremely difficult to perceive their correct orientation, especially when keyholes and electronic sockets are in difficult-to-reach, dimly lit locations. Some devices, such as USB plugs, are constrained, but the constraint is so subtle that it takes much fussing and fumbling to find the correct orientation. Why aren't all these devices orientation insensitive?

It is not difficult to design keys and plugs that work regardless of how they are inserted. Automobile keys that are insensitive to the orientation have long existed, but not all manufacturers use them. Similarly, many electrical connectors are insensitive to orientation, but again, only a few manufacturers use them. Why the resistance? Some of it results from the legacy concerns about the expense of massive change. But much seems to be a classic example of corporate thinking: "This is the way we have always done things. We don't care about the customer." It is, of course, true that difficulty in inserting keys, batteries, or plugs is not a big enough issue to affect the decision of whether to purchase something, but still, the lack of attention to customer needs on even simple things is often symptomatic of larger issues that have greater impact.

Note that a superior solution would be to solve the fundamental need—solving the root need. After all, we don't really care about keys and locks: what we need is some way of ensuring that only authorized people can get access to whatever is being locked. Instead of redoing the shapes of physical keys, make them irrelevant. Once this is recognized, a whole set of solutions present themselves: combination locks that do not require keys, or keyless locks that can be operated only by authorized people. One method is through possession of an electronic wireless device, such as the identification badges that unlock doors when they are moved close to a sensor, or automobile keys that can stay in the pocket or carrying case. Biometric devices could identify the person through face or voice recognition, fingerprints, or other biometric measures, such as iris patterns. This approach is discussed in Chapter 3, page 91.

CULTURAL CONSTRAINTS

Each culture has a set of allowable actions for social situations. Thus, in our own culture 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 a strange room, at a strange party, with 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 many of the problems we have with new machines: there are as yet no universally 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 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 proposed that in these cases we follow "scripts" that can guide the sequence of behavior. The sociologist Erving 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 of a culture.

The next time you are in an elevator, try violating cultural norms and see how uncomfortable that makes you and the other people in the elevator. It doesn't take much: Stand facing the rear. Or look directly at some of the passengers. In a bus or streetcar, give your seat to the next athletic-looking person you see (the act is especially effective if you are elderly, pregnant, or disabled).

In the case of the Lego motorcycle of Figure 4.1, cultural constraints determine the locations of the three lights of the motorcycle, which are otherwise physically interchangeable. Red is the culturally defined standard for a brake light, which is placed in the rear. And a police vehicle often has a blue flashing light on top. As for the yellow piece, this is an interesting example of cultural change: few people today remember that yellow used to be a standard headlight color in Europe and a few other locations (Lego comes from Denmark). Today, European and North American standards require white headlights. As a result, figuring out that the yellow piece represents a headlight on the front of the motorcycle is no longer as easy as it used to be. Cultural constraints are likely to change with time.

SEMANTIC CONSTRAINTS

Semantics is the study of meaning. Semantic constraints are those that 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. But just as cultural constraints can change with time, so, too, can semantic ones. Extreme sports push the boundaries of what we think of as meaningful and sensible. New technologies change the meanings of things. And creative people continually change how we interact with our technologies and one another. When cars become fully automated, communicating among themselves with wireless networks, what will be the meaning of the red lights on the rear of the auto? That the car is braking? But for whom would the signal be intended? The other cars would already know. The red light would become meaningless, so it could either be removed or it could be redefined to indicate some other condition. The meanings of today may not be the meanings of the future.

LOGICAL CONSTRAINTS

The blue light of the Lego motorcycle presents a special problem. Many people had no knowledge that would help, but after all the other pieces had been placed on the motorcycle, there was only one piece left, only one possible place to go. The blue light was logically constrained.

Logical constraints are often used by home dwellers who undertake repair jobs. Suppose you take apart a leaking faucet to replace a washer, but when you put the faucet together again, you discover a part left over. Oops, obviously there was an error: the part should have been installed. This is an example of a logical constraint.

The natural mappings discussed in Chapter 3 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 orientation of the lights and the switches differ, the natural mapping is destroyed.

CULTURAL NORMS, CONVENTIONS, AND STANDARDS

Every culture has its own conventions. Do you kiss or shake hands when meeting someone? If kissing, on which cheek, and how many times? Is it an air kiss or an actual one? Or perhaps you bow, junior

person first, and lowest. Or raise hands, or perhaps press them together. Sniff? It is possible to spend a fascinating hour on the Internet exploring the different forms of greetings used by different cultures. It is also amusing to watch the consternation when people from more cool, formal countries first encounter people from warm-hearted, earthy countries, as one tries to bow and shake hands and the other tries to hug and kiss even total strangers. It is not so amusing to be one of those people: being hugged or kissed while trying to shake hands or bow. Or the other way around. Try kissing someone's cheek three times (left, right, left) when the person expects only one. Or worse, where he or she expects a handshake. Violation of cultural conventions can completely disrupt an interaction.

Conventions are actually a form of cultural constraint, usually associated with how people behave. Some conventions determine what activities should be done; others prohibit or discourage actions. But in all cases, they provide those knowledgeable of the culture with powerful constraints on behavior.

Sometimes these conventions are codified into international standards, sometimes into laws, and sometimes both. In the early days of heavily traveled streets, whether by horses and buggies or by automobiles, congestion and accidents arose. Over time, conventions developed about which side of the road to drive on, with different conventions in different countries. Who had precedence at crossings? The first person to get there? The vehicle or person on the right, or the person with the highest social status? All of these conventions have applied at one time or another. Today, worldwide standards govern many traffic situations: Drive on only one side of the street. The first car into an intersection has precedence. If both arrive at the same time, the car on the right (or left) has precedence. When merging traffic lanes, alternate cars—one from that lane, then one from this. The last rule is more of an informal convention: it is not part of any rule book that I am aware of, and although it is very nicely obeyed in the California streets on which I drive, the very concept would seem strange in some parts of the world.

Sometimes conventions clash. In Mexico, when two cars approach a narrow, one-lane bridge from opposite directions, if a car blinks its headlights, it means, "I got here first and I'm going over the bridge." In England, if a car blinks its lights, it means, "I see you: please go first." Either signal is equally appropriate and useful, but not if the two drivers follow different conventions. Imagine a Mexican driver meeting an English driver in some third country. (Note that driving experts warn against using headlight blinks as signals because even within any single country, either interpretation is held by many drivers, none of whom imagines someone else might have the opposite interpretation.)

Ever get embarrassed at a formal dinner party where there appear to be dozens of utensils at each place setting? What do you do? Do you drink that nice bowl of water or is it for dipping your fingers to clean them? Do you eat a chicken drumstick or slice of pizza with your fingers or with a knife and fork?

Do these issues matter? Yes, they do. Violate conventions and you are marked as an outsider. A rude outsider, at that.

Applying Affordances, Signifiers, and Constraints to Everyday Objects

Affordances, signifiers, mappings, and constraints can simplify our encounters with everyday objects. Failure to properly deploy these cues leads to problems.

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. To operate a door, we have to find 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, a signifier, 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 by using the signifiers, in part guided by constraints.

Doors come in amazing variety. Some open only if a button is pushed, and some don't indicate how 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, to pull, push, slide, lift, ring a bell, insert a card, type a password, smile, rotate, bow, dance, or, perhaps, just ask. Somehow, when a device as simple as a door has to have a sign telling you whether to pull, push, or slide, then it is a failure, poorly designed.

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 incorporate clear and unambiguous clues—signifiers. 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.

Flat plates or bars can clearly and unambiguously signify both the proper action and its location, for their affordances constrain the possible actions to that of pushing. Remember the discussion of the fire door and its panic bar in Chapter 2 (Figure 2.5, page 60)? The panic bar, with its large horizontal surface, often with a secondary color on the part intended to be pushed, provides a good example of an unambiguous signifier. It very nicely constrains improper behavior when panicked people press against the door as they attempt to flee a fire. The best push bars offer both visible affordances that act as physical constraints on the action, and also a visible signifier, thereby unobtrusively specifying *what* to do and *where* to do it.

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. 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 awkward to use.

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 how 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 cabinet door that opens outward by being pushed inward. The push releases the catch and energizes 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 do not wish to mar the smooth surface of the door. One of the cabinets in my home has one of these latches in its glass door. Because the glass affords visibility of the shelves inside, it is obvious that there is no room for the door to open inward; therefore, to push the door seems contradictory. New and infrequent users of this door usually reject pushing and open it by pulling, which often requires them to use fingernails, knife blades, or more ingenious methods to pry it open. A similar, counterintuitive type of design was the source of my difficulties in emptying the dirty water from my sink in a London hotel (Figure 1.4, page 17).

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). In some transit systems, the passenger is supposed to operate

the door, but in others this is forbidden. The frequent traveler is continually confronted with this kind of situation: the behavior that is appropriate in one place is inappropriate in another, even in situations that appear to be identical. Known cultural norms can create comfort and harmony. Unknown norms can lead to discomfort and confusion.

THE PROBLEM WITH SWITCHES

When I give talks, quite often 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 industry could be dangerous. In many control rooms, row upon row of identical-looking switches confront the operators. How do they avoid the occasional error, confusion, or accidental bumping against the wrong control? Or mis-aim? They don't. Fortunately, industrial settings are usually pretty robust. A few errors every now and then are not important—usually.

One type of popular small airplane has identical-looking switches for flaps and for 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 had been known for fifty years. Why were these 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 to determine what type of device they control; for example, flaps or landing gear. The second is the mapping problem, discussed extensively in Chapters 1 and 3; for example, when there are many lights and an array of switches, 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 appear in offices, auditoriums, and industrial locations than in homes.

With complex installations, where there are numerous lights and switches, the light controls seldom fit the needs of the situation. When I give talks, I need a way to dim the light hitting the projection screen so that images are visible, but keep enough light on the audience so that they can take notes (and I can monitor their reaction to the talk). This kind of control is seldom provided. Electricians are not trained to do task analyses.

Whose fault is this? Probably nobody's. Blaming a person is seldom appropriate or useful, a point I return to in Chapter 5. The problem is probably due to the difficulties of coordinating the different professions involved in installing light controls.

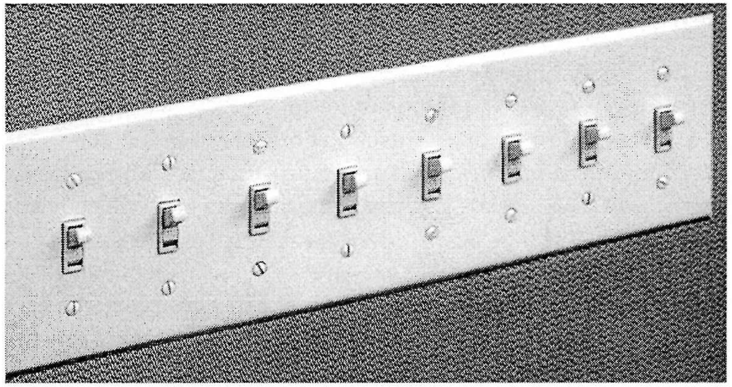


FIGURE 4.4. **Incomprehensible Light Switches.** Banks of switches like this are not uncommon in homes. There is no obvious mapping between the switches and the lights being controlled. I once had a similar panel in my home, although with only six switches. Even after years of living in the house, I could never remember which to use, so I simply put all the switches either up (on) or down (off). How did I solve the problem? See Figure 4.5.

I once lived in a wonderful house on the cliffs of Del Mar, California, designed for us by two young, award-winning architects. The house was wonderful, and the architects proved their worth by the spectacular placement of the house and the broad windows that overlooked the ocean. But they liked spare, neat, modern design to a fault. Inside the house were, among other things, neat rows of light switches: 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. Figure 4.4 shows an eight-switch bank that I found in a home I was visiting. Who could remember what each does? My home only had six switches, and that was bad enough. (Photographs of the switch plate from my Del Mar home are no longer available.)

The lack of clear communication among the people and organizations constructing parts of a system is perhaps the most common cause of complicated, confusing designs. A usable design starts with careful observations of how the tasks being supported are actually performed, followed by a design process that results in a good fit to the actual ways the tasks get performed. The technical name for this method is *task analysis.* The name for the entire process is *human-centered design* (HCD), discussed in Chapter 6.

The solutions to the problem posed by my Del Mar home require the natural mappings described in Chapter 3. With six light switches mounted in a one-dimensional array, vertically on the wall, there is no way they can map naturally to the two-dimensional, horizontal placement of the lights in the ceiling. 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 floor plan 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. You can see the result in Figure 4.5. We mounted a floor plan of the living room on a plate and oriented it to match the room. Switches were placed on the floor plan so that each switch was located in the area controlled

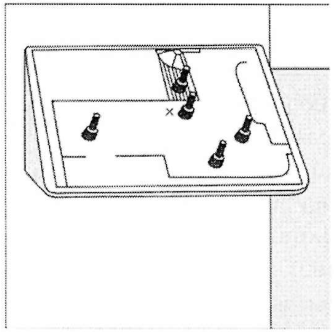


FIGURE 4.5. A Natural Mapping of Light Switches to Lights. This is how I mapped five switches to the lights in my living room. I placed small toggle switches that fit onto a plan of the home's living room, balcony, and hall, with each switch placed where the light was located. The X by the center switch indicates where this panel was located. The surface was tilted to make it easier to relate it to the horizontal arrangement of the lights, and the slope provided a natural anti-affordance, preventing people from putting coffee cups and drink containers on the controls.

by that switch. The plate was mounted with a slight tilt from the horizontal to make it easy to see and to make the mapping clear: had the plate been vertical, the mapping would still be ambiguous. The plate was tilted rather than horizontal to discourage people (us or visitors) from placing objects, such as cups, on the plate: an example of an anti-affordance. (We further simplified operations by moving the sixth switch to a different location where its meaning was clear and it did not confuse, because it stood alone.)

It is unnecessarily difficult to implement this spatial mapping of switches to lights: the required parts are not available. I had to hire a skilled technician to construct the wall-mounted box and install the special switches and control equipment. Builders and electricians need standardized components. Today, the switch boxes that are available to electricians are organized as rectangular boxes meant to hold a long, linear string of switches and to be mounted horizontally or vertically on the wall. To produce the appropriate spatial array, we would need a two-dimensional structure that could be mounted parallel to the floor, where the switches would be mounted on the top of the box, on the horizontal surface. The switch box should have a matrix of supports so that there can be free, relatively unrestricted placement of the switches in whatever pattern best suits the room. Ideally the box would use small switches, perhaps low-voltage switches that would control a separately mounted control structure that takes care of the lights (which is what I did in my home). Switches and lights could communicate

wirelessly instead of through the traditional home wiring cables. Instead of the standardized light plates for today's large, bulky switches, the plates should be designed for small holes appropriate to the small switches, combined with a way of inserting a floor plan on to the switch cover.

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. But these new switch boxes wouldn't have to stick out. They could be placed in indented openings in the walls: just as there is room inside the wall for the existing switch boxes, there is also room for an indented horizontal surface. Or the switches could be mounted on a little pedestal.

As a side note, in the decades that have passed since the first edition of this book was published, the section on natural mappings and the difficulties with light switches has received a very popular reception. Nonetheless, there are no commercial tools available to make it easy to implement these ideas in the home. I once tried to convince the CEO of the company whose smart home devices I had used to implement the controls of Figure 4.5, to use the idea. "Why not manufacture the components to make it easy for people to do this," I suggested. I failed.

Someday, we will get rid of the hard-wired switches, which require excessive runs of electrical cable, add to the cost and difficulties of home construction, and make remodeling of electrical circuits extremely difficult and time consuming. Instead, we will use Internet or wireless signals to connect switches to the devices to be controlled. In this way, controls could be located anywhere. They could be reconfigured or moved. We could have multiple controls for the same item, some in our phones or other portable devices. I can control my home thermostat from anywhere in the world: why can't I do the same with my lights? Some of the necessary technology does exist today in specialty shops and custom builders, but they will not come into widespread usage until major manufacturers make the necessary components and traditional electricians become comfortable with installing them. The tools for creating switch configurations that use good mapping principles could become standard and easy to apply. It will happen, but it may take considerable time.

Alas, like many things that change, new technologies will bring virtues and deficits. The controls are apt to be through touch-sensitive screens, allowing excellent natural mapping to the spatial layouts involved, but lacking the physical affordances of physical switches. They can't be operated with the side of the arm or the elbow while trying to enter a room, hands loaded with packages or cups of coffee. Touch screens are fine if the hands are free. Perhaps cameras that recognize gestures will do the job.

ACTIVITY-CENTERED CONTROLS

Spatial mapping of switches is not always appropriate. In many cases it is better to have switches that control activities: activity-centered control. Many auditoriums in schools and companies have computer-based controls, with switches labeled with such phrases as "video," "computer," "full lights," and "lecture." When carefully designed, with a good, detailed analysis of the activities to be supported, the mapping of controls to activities works extremely well: video requires a dark auditorium plus control of sound level and controls to start, pause, and stop the presentation. Projected images require a dark screen area with enough light in the auditorium so people can take notes. Lectures require some stage lights so the speaker can be seen. Activity-based controls are excellent in theory, but the practice is difficult to get right. When it is done badly, it creates difficulties.

A related but wrong approach is to be device-centered rather than activity-centered. When they are device-centered, different control screens cover lights, sound, computer, and video projection. This requires the lecturer to go to one screen to adjust the light, a different screen to adjust sound levels, and yet a different screen to advance or control the images. It is a horrible cognitive interruption to the flow of the talk to go back and forth among the screens, perhaps to pause the video in order to make a comment or answer a question. Activity-centered controls anticipate this need and put light, sound level, and projection controls all in one location.

I once used an activity-centered control, setting it to present my photographs to the audience. All worked well until I was asked a question. I paused to answer it, but wanted to raise the room lights so I could see the audience. No, the activity of giving a talk with visually presented images meant that room lights were fixed at a dim setting. When I tried to increase the light intensity, this took me out of "giving a talk" activity, so I did get the light to where I wanted it, but the projection screen also went up into the ceiling and the projector was turned off. The difficulty with activity-based controllers is handling the exceptional cases, the ones not thought about during design.

Activity-centered controls are the proper way to go, if the activities are carefully selected to match actual requirements. But even in these cases, manual controls will still be required because there will always be some new, unexpected demand that requires idiosyncratic settings. As my example demonstrates, invoking the manual settings should not cause the current activity to be canceled.

Constraints That Force the Desired Behavior

FORCING FUNCTIONS

Forcing functions are a form of physical constraint: situations in which the actions are constrained so that failure at one stage prevents the next step from happening. Starting a car has a forcing function associated with it—the driver must have some physical object that signifies permission to use the car. In the past, it was a physical key to unlock the car doors and also to be placed into the ignition switch, which allowed the key to turn on the electrical system and, if rotated to its extreme position, to activate the engine.

Today's cars have many means of verifying permission. Some still require a key, but it can stay in one's pocket or carrying case. More and more, the key is not required and is replaced by a card, phone, or some physical token that can communicate with the car. As long as only authorized people have the card (which is, of course, the same for keys), everything works fine. Electric or hybrid vehicles do not need to start the engines prior to moving the car, but the procedures are still similar: drivers must authenticate themselves by having a physical item in their possession. Because the vehicle won't start without the authentication proved by possession of the key, it is a forcing function.

Forcing functions are the extreme case of strong constraints that can prevent inappropriate behavior. Not every situation allows such strong constraints to operate, but the general principle can be extended to a wide variety of situations. In the field of safety engineering, forcing functions show up under other names, in particular as specialized methods for the prevention of accidents. Three such methods are interlocks, lock-ins, and lockouts.

INTERLOCKS

An interlock forces operations to take place in proper sequence. Microwave ovens and devices with interior exposure to high voltage use interlocks as forcing functions to prevent people from opening the door of the oven or disassembling the devices without first turning off the electric power: the interlock disconnects the power the instant the door is opened or the back is removed. In automobiles with automatic transmissions, an interlock prevents the transmission from leaving the Park position unless the car's brake pedal is depressed.

Another form of interlock is the "dead man's switch" in numerous safety settings, especially for the operators of trains, lawn mowers, chainsaws, and many recreational vehicles. In Britain, these are called the "driver's safety device." Many require that the operator hold down a spring-loaded switch to enable operation of the equipment, so that if the operator dies (or loses control), the switch will be released, stopping the equipment. Because some operators bypassed the feature by tying down the control (or placing a heavy weight on foot-operated ones), various schemes have been developed to determine that the person is really alive and alert. Some require a midlevel of pressure; some, repeated depressions and releases. Some require responses to queries. But in all cases,

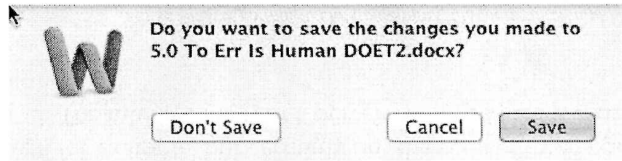


FIGURE 4.6 A Lock-In Forcing Function. This lock-in makes it difficult to exit a program without either saving the work or consciously saying not to. Notice that it is politely configured so that the desired operation can be taken right from the message.

they are examples of safety-related interlocks to prevent operation when the operator is incapacitated.

LOCK-INS

A lock-in keeps an operation active, preventing someone from prematurely stopping it. Standard lock-ins exist on many computer applications, where any attempt to exit the application without saving work is prevented by a message prompt asking whether that is what is really wanted (Figure 4. 6). These are so effective that I use them deliberately as my standard way of exiting. Rather than saving a file and then exiting the program, I simply exit, knowing that I will be given a simple way to save my work. What was once created as an error message has become an efficient shortcut.

Lock-ins can be quite literal, as in jail cells or playpens for babies, preventing a person from leaving the area.

Some companies try to lock in customers by making all their products work harmoniously with one another but be incompatible with the products of their competition. Thus music, videos, or electronic books purchased from one company may be played or read on music and video players and e-book readers made by that company, but will fail with similar devices from other manufacturers. The goal is to use design as a business strategy: the consistency within a given manufacturer means once people learn the system, they will stay with it and hesitate to change. The confusion when using a different company's system further prevents customers from

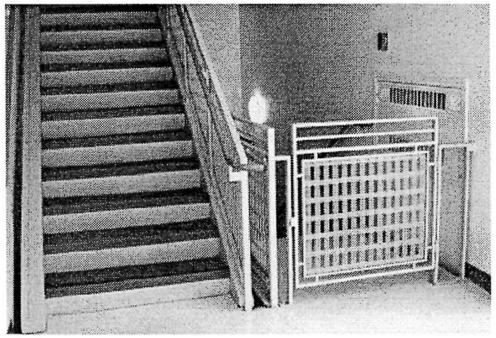


FIGURE 4.7. A Lockout Forcing Function for Fire Exit. The gate, placed at the ground floor of stairways, prevents people who might be rushing down the stairs to escape a fire from continuing into the basement areas, where they might get trapped.

changing systems. In the end, the people who must use multiple systems lose. Actually, everyone loses, except for the one manufacturer whose products dominate.

LOCKOUTS

Whereas a lock-in keeps someone in a space or prevents an action until the desired operations have been done, a lockout prevents someone from entering a space that is dangerous, or prevents an event from occurring. A good example of a lockout is found in stairways of public buildings, at least in the United States (Figure 4.7). In cases of fire, people have a tendency to flee in panic, down the stairs, down, down, down, past the ground floor and into the basement, where they might be trapped. The solution (required by the fire laws) is not to allow simple passage from the ground floor to the basement.

Lockouts are usually used for safety reasons. Thus, small children are protected by baby locks on cabinet doors, covers for electric outlets, and specialized caps on containers for drugs and toxic substances. The pin that prevents a fire extinguisher from being activated until it is removed is a lockout forcing function to prevent accidental discharge.

Forcing functions can be a nuisance in normal usage. The result is that many people will deliberately disable the forcing function, thereby negating its safety feature. The clever designer has to minimize the nuisance value while retaining the safety feature of the forcing function that guards against the occasional tragedy. The gate in Figure 4.7 is a clever compromise: sufficient restraint to make people realize they are leaving the ground floor, but not enough of an impediment to normal behavior that people will prop open the gate.

Other useful devices make use of a forcing function. In some public restrooms, a pull-down shelf is placed inconveniently on the wall just behind the cubicle door, held in a vertical position by a spring. You lower the shelf to the horizontal position, and the weight of a package or handbag keeps it there. The shelf's position is a forcing function. When the shelf is lowered, it blocks the door fully. So to get out of the cubicle, you have to remove whatever is on the shelf and raise it out of the way. Clever design.

Conventions, Constraints, and Affordances

In Chapter 1 we learned of the distinctions between affordances, perceived affordances, and signifiers. Affordances refer to the potential actions that are possible, but these are easily discoverable only if they are perceivable: perceived affordances. It is the signifier component of the perceived affordance that allows people to determine the possible actions. But how does one go from the perception of an affordance to understanding the potential action? In many cases, through conventions.

A doorknob has the perceived affordance of graspability. But knowing that it is the doorknob that is used to open and close doors is learned: it is a cultural aspect of the design that knobs, handles, and bars, when placed on doors, are intended to enable the opening and shutting of those doors. The same devices on fixed walls would have a different interpretation: they might offer support, for example, but certainly not the possibility of opening the wall. The interpretation of a perceived affordance is a cultural convention.

CONVENTIONS ARE CULTURAL CONSTRAINTS

Conventions are a special kind of cultural constraint. For example, the means by which people eat is subject to strong cultural constraints and conventions. Different cultures use different eating utensils. Some eat primarily with the fingers and bread. Some use elaborate serving devices. The same is true of almost every aspect of behavior imaginable, from the clothes that are worn; to the way one addresses elders, equals, and inferiors; and even to the order in which people enter or exit a room. What is considered correct and proper in one culture may be considered impolite in another.

Although conventions provide valuable guidance for novel situations, their existence can make it difficult to enact change: consider the story of destination-control elevators.

WHEN CONVENTIONS CHANGE: THE CASE OF DESTINATION-CONTROL ELEVATORS

> *Operating the common elevator seems like a no-brainer. Press the button, get in the box, go up or down, get out. But we've been encountering and documenting an array of curious design variations on this simple interaction, raising the question: Why?* (From Portigal & Norvaisas, 2011.)

This quotation comes from two design professionals who were so offended by a change in the controls for an elevator system that they wrote an entire article of complaint.

What could possibly cause such an offense? Was it really bad design or, as the authors suggest, a completely unnecessary change to an otherwise satisfactory system? Here is what happened: the authors had encountered a new convention for elevators called "Elevator Destination Control." Many people (including me) consider it superior to the one we are all used to. Its major disadvantage is that it is different. It violates customary convention. Violations of convention can be very disturbing. Here is the history.

When "modern" elevators were first installed in buildings in the late 1800s, they always had a human operator who controlled the speed and direction of the elevator, stopped at the appropriate floors, and opened and shut the doors. People would enter the elevator, greet the operator, and state which floor they wished to travel to. When the elevators became automated, a similar convention was followed. People entered the elevator and told the elevator what floor they were traveling to by pushing the appropriately marked button inside the elevator.

This is a pretty inefficient way of doing things. Most of you have probably experienced a crowded elevator where every person seems to want to go to a different floor, which means a slow trip for the people going to the higher floors. A destination-control elevator system groups passengers, so that those going to the same floor are asked to use the same elevator and the passenger load is distributed to maximize efficiency. Although this kind of grouping is only sensible for buildings that have a large number of elevators, that would cover any large hotel, office, or apartment building.

In the traditional elevator, passengers stand in the elevator hallway and indicate whether they wish to travel up or down. When an elevator arrives going in the appropriate direction, they get in and use the keypad inside the elevator to indicate their destination floor. As a result, five people might get into the same elevator each wanting a different floor. With destination control, the destination keypads are located in the hallway outside the elevators and there are no keypads inside the elevators (Figure 4.8A and D). People are directed to whichever elevator will most efficiently reach their floor. Thus, if there were five people desiring elevators, they might be assigned to five different elevators. The result is faster trips for everyone, with a minimum of stops. Even if people are assigned to elevators that are not the next to arrive, they will get to their destinations faster than if they took earlier elevators.

Destination control was invented in 1985, but the first commercial installation didn't appear until 1990 (in Schindler elevators). Now, decades later, it is starting to appear more frequently as developers of tall buildings discover that destination control yields better service to passengers, or equal service with fewer elevators.

Horrors! As Figure 4.8D confirms, there are no controls inside the elevator to specify a floor. What if passengers change their minds

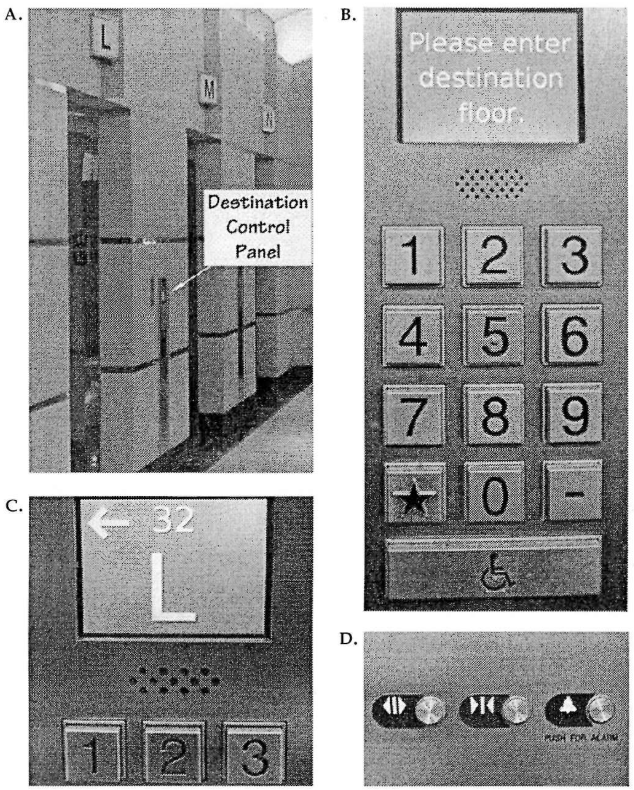


FIGURE 4.8. **Destination-Control Elevators.** In a destination-control system, the desired destination floor is entered into the control panel outside the elevators (A and B). After entering the destination floor into B, the display directs the traveler to the appropriate elevator, as shown in C, where "32" has been entered as the desired floor destination, and the person is directed to elevator "L" (the first elevator on the left, in A). There is no way to specify the floor from inside the elevator: Inside, the controls are only to open and shut the doors and an alarm (D). This is a much more efficient design, but confusing to people used to the more conventional system. (Photographs by the author.)

and wish to get off at a different floor? (Even my editor at Basic Books complained about this in a marginal note.) What then? What do you do in a regular elevator when you decide you really want to get off at the sixth floor just as the elevator passes the seventh floor? It's simple: just get off at the next stop and go to the destination control box in the elevator hall, and specify the intended floor.

PEOPLE'S RESPONSES TO CHANGES IN CONVENTIONS

People invariably object and complain whenever a new approach is introduced into an existing array of products and systems. Conventions are violated: new learning is required. The merits of the new system are irrelevant: it is the change that is upsetting. The destination control elevator is only one of many such examples. The metric system provides a powerful example of the difficulties in changing people's conventions.

The metric scale of measurement is superior to the English scale of units in almost every dimension: it is logical, easy to learn, and easy to use in computations. Today, over two centuries have passed since the metric system was developed by the French in the 1790s, yet three countries still resist its use: the United States, Liberia, and Myanmar. Even Great Britain has mostly switched, so the only major country left that uses the older English system of units is the United States. Why haven't we switched? The change is too upsetting for the people who have to learn the new system, and the initial cost of purchasing new tools and measuring devices seems excessive. The learning difficulties are nowhere as complex as purported, and the cost would be relatively small because the metric system is already in wide use, even in the United States.

Consistency in design is virtuous. It means that lessons learned with one system transfer readily to others. On the whole, consistency is to be followed. If a new way of doing things is only slightly better than the old, it is better to be consistent. But if there is to be a change, everybody has to change. Mixed systems are confusing to everyone. When a new way of doing things is vastly superior to another, then the merits of change outweigh the difficulty of change. Just because something is different does not mean it is bad. If we only kept to the old, we could never improve.

The Faucet: A Case History of Design

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 were lodged in dormitories. Upon checking into Ranmoor House, each guest was given a pamphlet that provided 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 them? Polite, restrained tittering 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 walked the halls until she found someone who could explain the taps to her. A simple sink, a simple-looking faucet. But it looks as if it should be turned, not pushed. If you want the faucet to be pushed, make it look as if it should be pushed. (This, of course, is similar to the problem I had emptying the water from the sink in my hotel, described in Chapter 1.)

Why is such a simple, standard item as a water faucet so difficult to get right? The person using a faucet cares about two things: water temperature and rate of flow. But water enters the faucet through two pipes, hot and cold. There is a conflict between the human need for temperature and flow and the physical structure of hot and cold.

There are several ways to deal with this:

- **Control both hot and cold water:** Two controls, one for hot water, the other cold.
- **Control only temperature:** One control, where rate of flow is fixed. Rotating the control from its fixed position turns on the water at some predetermined rate of flow, with the temperature controlled by the knob position.
- **Control only amount:** One control, where temperature is fixed, with rate of flow controlled by the knob position.
- **On-off.** One control turns the water on and off. This is how gesture-controlled faucets work: moving the hand under or away from the spout turns the water on or off, at a fixed temperature and rate of flow.
- **Control temperature and rate of flow.** Use two separate controls, one for water temperature, the other for flow rate. (I have never encountered this solution.)
- **One control for temperature and rate:** Have one integrated control, where movement in one direction controls the temperature and movement in a different direction controls the amount.

Where there are two controls, one for hot water and one for cold, there are four mapping problems;

- Which knob controls the hot, which the cold?
- How do you change the temperature without affecting the rate of flow?
- How do you change the flow without affecting the temperature?
- Which direction increases water flow?

The 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, either. I once experienced shower controls that were placed vertically: Which one controlled the hot water, the top faucet or the bottom?

If the two faucet handles are round knobs, clockwise rotation of either should decrease volume. However, if each faucet has a single "blade" as its handle, then people don't think they are rotating the handles: they think that they are pushing or pulling. To maintain consistency, pulling either faucet should increase volume, even though this means rotating the left faucet counterclockwise and the right one clockwise. Although rotation direction is inconsistent, pulling and pushing is consistent, which is how people conceptualize their actions.

Alas, sometimes clever people are too clever for our good. Some well-meaning plumbing designers have decided that consistency should be ignored in favor of their own, private brand of psychology. The human body has mirror-image symmetry, say these pseudo-psychologists. 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 whose clockwise rotation has a different result with the hot water than with the cold.

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. If the water is too cold, the groping hand is just as likely to make the water colder as to make it scalding hot.

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 works as long as you always use two hands to adjust both faucets simultaneously. It fails miserably, however, when one hand is used to alternate between the two controls. Then you cannot remember which direction does what. Once again, notice that this can be corrected without replacing the individual faucets: just replace the handles with blades. It is psychological perceptions that matter—the conceptual model—not physical consistency.

The operation of faucets needs to be standardized so that the psychological conceptual model of operation is the same for all types of faucets. With the traditional dual faucet controls for hot and cold water, the standards should state:

- When the handles are round, both should rotate in the same direction to change water volume.
- When the handles are single blades, both should be pulled to change water volume (which means rotating in opposite directions in the faucet itself).

Other configurations of handles are possible. Suppose the handles are mounted on a horizontal axis so that they rotate vertically. Then what? Would the answer differ for single blade handles and round ones? I leave this as an exercise for the reader.

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 or freeze yourself. When the faucets are far removed from the spout, as is the case where the faucets are located in the center of the bathtub but the spouts high on an end wall, the delay between turning the faucets and the change in temperature can be quite long: I once timed a shower control to take 5 seconds. This makes setting the temperature rather difficult. Turn the faucet the wrong way and then dance around inside the shower while the water is scalding hot or freezing cold, madly turning the faucet in what you hope is the correct direction, hoping the temperature will stabilize quickly. Here the problem comes from the properties of fluid flow—it takes time for water to travel the 2 meters or so of pipe that might connect the faucets with the spout—so it is not easily remedied. But the problem is exacerbated by poor design of the controls.

Now let's turn to 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. The difficulty lies in a lack of standardization of the dimensions of control, and then, which direction of movement means what? Sometimes there is a knob that can be pushed or pulled, rotated clockwise or counterclockwise. But does the push or pull control volume or temperature? Is a pull more volume or less, hotter temperature or cooler? Sometimes there is a lever that moves side to side or forward and backward. Once again, which movement is volume, which temperature? And even then, which way is more (or hotter), which is less (or cooler)? The perceptually simple one-control faucet still has four mapping problems:

- What dimension of control affects the temperature?
- Which direction along that dimension means hotter?
- What dimension of control affects the rate of flow?
- Which direction along that dimension means more?

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. And then, different faucet designs use different solutions. One-control faucets ought to be superior because they control the psychological variables of interest. But because of the lack of standardization and awkward design (to call it "awkward" is being kind), they frustrate many people so much that they tend to be disliked more than they are admired.

Bath and kitchen faucet design ought to be simple, but can violate many design principles, including:

- Visible affordances and signifiers
- Discoverability
- Immediacy of feedback

Finally, many violate the principle of desperation:

- If all else fails, standardize.

Standardization is indeed the fundamental principle of desperation: when no other solution appears possible, simply design everything the same way, so people only have to learn once. 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 (that is, knowledge in the world), then develop a cultural constraint: standardize what has to be kept in the head. And remember the lesson from faucet rotation on page 153: The standards should reflect the psychological conceptual models, not the physical mechanics.

Standards simplify life for everyone. At the same time, they tend to hinder future development. And, as discussed in Chapter 6, there are often difficult political struggles in finding common agreement. Nonetheless, when all else fails, standards are the way to proceed.

Using Sound as Signifiers

Sometimes everything that is needed cannot 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 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 teakettle 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 simply 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 knowledge 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.

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. Experienced mechanics can diagnosis the condition of machinery just by listening. When sounds are generated artificially, if intelligently created using a rich auditory spectrum, with care to provide the subtle cues that are informative without being annoying, they can be as useful as sounds in the real world.

Sound is tricky. It can annoy and distract as easily as it can aid. Sounds that at one's first encounter are pleasant or cute easily become annoying rather than useful. 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 knowledge 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 knowledge, and if feedback from an action is expected to come from sound, silence can lead to problems.

WHEN SILENCE KILLS

It was a pleasant June day in Munich, Germany. I was picked up at my hotel and driven to the country with farmland on either side of the narrow, two-lane road. Occasional walkers strode by, and every so often a bicyclist passed. We parked the car on the shoulder of the road and joined a group of people looking up and down the road. "Okay, get ready," I was told. "Close your eyes and listen." I did so and about a minute later I heard a high-pitched whine, accompanied by a low humming sound: an automobile was approaching. As it came closer, I could hear tire noise. After the car had passed, I was asked my judgment of the sound. We repeated the exercise numerous times, and each time the sound was different. What was going on? We were evaluating sound designs for BMW's new electric vehicles.

Electric cars are extremely quiet. The only sounds they make come from the tires, the air, and occasionally, from the high-pitched whine of the electronics. Car lovers really like the silence. Pedestrians have mixed feelings, but the blind are greatly concerned. After all, the blind cross streets in traffic by relying upon the sounds of vehicles. That's how they know when it is safe to cross. And what is true for the blind might also be true for anyone stepping onto the street while distracted. If the vehicles don't make any sounds, they can kill. The United States National Highway Traffic Safety Administration determined that pedestrians are considerably more likely to be hit by hybrid or electric vehicles than by those that have an internal combustion engine. The greatest danger is when the hybrid or electric vehicles are moving slowly, when they are almost completely silent. The sounds of an automobile are important signifiers of its presence.

Adding sound to a vehicle to warn pedestrians is not a new idea. For many years, commercial trucks and construction equipment have had to make beeping sounds when backing up. Horns are required by law, presumably so that drivers can use them to alert pedestrians and other drivers when the need arises, although they are often used as a way of venting anger and rage instead. But adding a continuous sound to a normal vehicle because it would otherwise be too quiet, is a challenge.

What sound would you want? One group of blind people suggested putting some rocks into the hubcaps. I thought this was brilliant. The rocks would provide a natural set of cues, rich in meaning yet easy to interpret. The car would be quiet until the wheels started to turn. Then, the rocks would make natural, continuous scraping sounds at low speeds, change to the pitter-patter of falling stones at higher speeds, the frequency of the drops increasing with the speed of the car until the car was moving fast enough that the rocks would be frozen against the circumference of the rim, silent. Which is fine: the sounds are not needed for fast-moving vehicles because then the tire noise is audible. The lack of sound when the vehicle was not moving would be a problem, however.

The marketing divisions of automobile manufacturers thought that the addition of artificial sounds would be a wonderful branding opportunity, so each car brand or model should have its own unique sound that captured just the car personality the brand wished to convey. Porsche added loudspeakers to its electric car prototype to give it the same "throaty growl" as its gasoline-powered cars. Nissan wondered whether a hybrid automobile should sound like tweeting birds. Some manufacturers thought all cars should sound the same, with standardized sounds and sound levels, making it easier for everyone to learn how to interpret them. Some blind people thought they should sound like cars—you know, gasoline engines, following the old tradition that new technologies must always copy the old.

Skeuomorphic is the technical term for incorporating old, familiar ideas into new technologies, even though they no longer play a functional role. Skeuomorphic designs are often comfortable for traditionalists, and indeed the history of technology shows that new technologies and materials often slavishly imitate the old for no apparent reason except that is what people know how to do. Early automobiles looked like horse-driven carriages without the horses (which is also why they were called horseless carriages); early plastics were designed to look like wood; folders in computer file systems often look the same as paper folders, complete with tabs. One way of overcoming the fear of the new is to make it look like the old. This practice is decried by design purists, but in fact, it has its benefits in easing the transition from the old to the new. It gives comfort and makes learning easier. Existing conceptual models need only be modified rather than replaced. Eventually, new forms emerge that have no relationship to the old, but the skeuomorphic designs probably helped the transition.

When it came to deciding what sounds the new silent automobiles should generate, those who wanted differentiation ruled the day, yet everyone also agreed that there had to be some standards. It should be possible to determine that the sound is coming from an automobile, to identify its location, direction, and speed. No sound would be necessary once the car was going fast enough, in part because tire noise would be sufficient. Some standardization would be required, although with a lot of leeway. International standards committees started their procedures. Various countries, unhappy with the normally glacial speed of standards agreements and under pressure from their communities, started drafting legislation. Companies scurried to develop appropriate sounds, hiring experts in psychoacoustics, psychologists, and Hollywood sound designers.

The United States National Highway Traffic Safety Administration issued a set of principles along with a detailed list of requirements, including sound levels, spectra, and other criteria. The full document is 248 pages. The document states:

This standard will ensure that blind, visually-impaired, and other pedestrians are able to detect and recognize nearby hybrid and electric vehicles by requiring that hybrid and electric vehicles emit sound that pedestrians will be able to hear in a range of ambient environments and contain acoustic signal content that pedestrians will recognize as being emitted from a vehicle. The proposed standard establishes minimum sound requirements for hybrid and electric vehicles when operating under 30 kilometers per hour (km/h) (18 mph), when the vehicle's starting system is activated but the vehicle is stationary, and when the vehicle is operating in reverse. The agency chose a crossover speed of 30 km/h because this was the speed at which the sound levels of the hybrid and electric vehicles measured by the agency approximated the sound levels produced by similar internal combustion engine vehicles. (Department of Transportation, 2013.)

As I write this, sound designers are still experimenting. The automobile companies, lawmakers, and standards committees are still at work. Standards are not expected until 2014 or later, and then it will take considerable time to be deployed to the millions of vehicles across the world.

What principles should be used for the design sounds of electric vehicles (including hybrids)? The sounds have to meet several criteria:

- **Alerting.** The sound will indicate the presence of an electric vehicle.
- **Orientation.** The sound will make it possible to determine where the vehicle is located, a rough idea of its speed, and whether it is moving toward or away from the listener.
- **Lack of annoyance.** Because these sounds will be heard frequently even in light traffic and continually in heavy traffic, they must not be annoying. Note the contrast with sirens, horns, and backup signals, all of which are intended to be aggressive warnings. Such sounds are deliberately unpleasant, but because they are infrequent and for relatively short duration, they are acceptable. The challenge faced by electric vehicle sounds is to alert and orient, not annoy.

- **Standardization versus individualization.** Standardization is necessary to ensure that all electric vehicle sounds can readily be interpreted. If they vary too much, novel sounds might confuse the listener. Individualization has two functions: safety and marketing. From a safety point of view, if there were many vehicles present on the street, individualization would allow vehicles to be tracked. This is especially important at crowded intersections. From a marketing point of view, individualization can ensure that each brand of electric vehicle has its own unique characteristic, perhaps matching the quality of the sound to the brand image.

Stand still on a street corner and listen carefully to the vehicles around you. Listen to the silent bicycles and to the artificial sounds of electric cars. Do the cars meet the criteria? After years of trying to make cars run more quietly, who would have thought that one day we would spend years of effort and tens of millions of dollars to add sound?