CHAPTER FOUR

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.

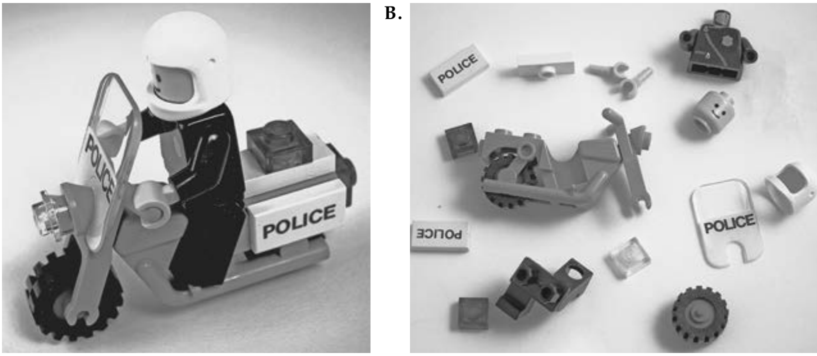

FIGURE 4.1. **Lego Motorcycle.** The toy Lego motorcycle is shown assembled (A) and in pieces (B). It has fifteen pieces so cleverly constructed that even an adult can put them together. The design exploits constraints to specify just which pieces fit where. Physical constraints limit alternative placements. 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.

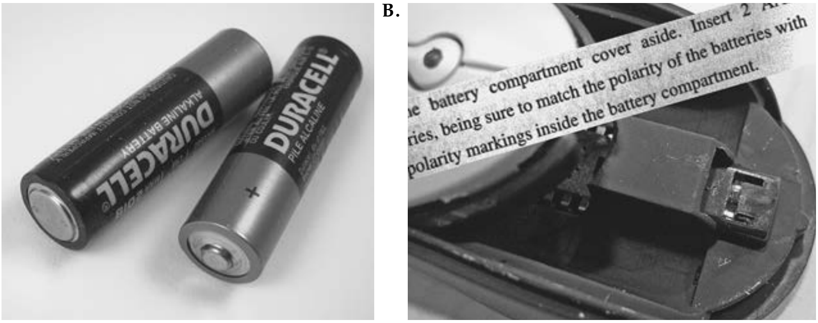

FIGURE 4.2. **Cylindrical Battery: Where Constraints Are Needed.** Figure A shows the traditional cylindrical battery that requires correct orientation in the slot to work properly (and to avoid damaging the equipment). But look at Figure B, which shows where two batteries are to be installed. The instructions 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.

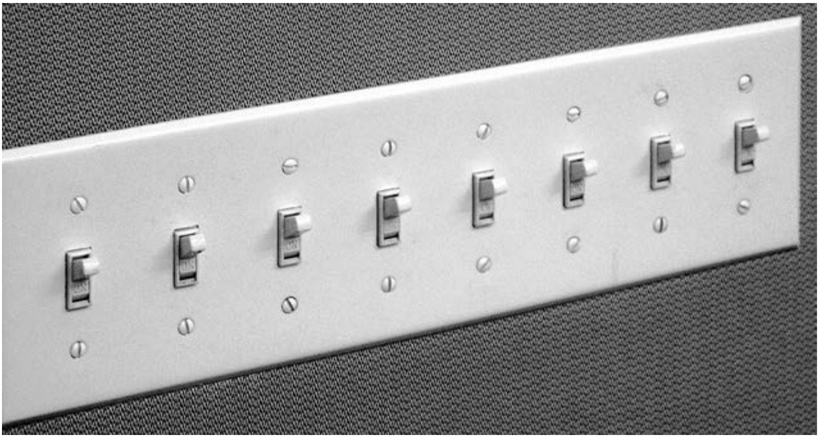

FIGURE 4.4. **Incomprehensible Light Switches.** Banks of switches like this are not uncommon in homes. There is no obvious mapping between the switches and the lights being controlled. I once had a similar panel in my home, although with only six switches. Even after years of living in the house, I 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

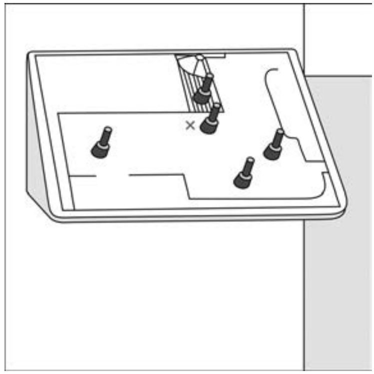

FIGURE 4.5. A Natural Mapping of Light Switches to Lights. This is how I mapped five switches to the lights in my living room. I placed small toggle switches that fit onto a plan of the home's living room, balcony, and hall, with each switch placed where the light was located. The X by the center switch 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,

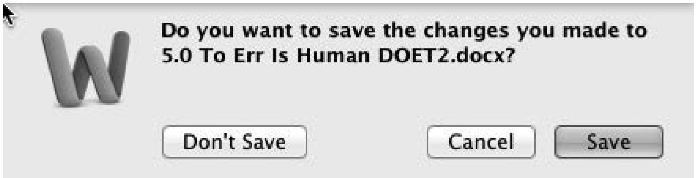

FIGURE 4.6 A Lock-In Forcing Function. This lock-in makes it difficult to exit a program without either saving the work or consciously saying not to. Notice that it is politely configured so that the desired operation can be taken right from the message.

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

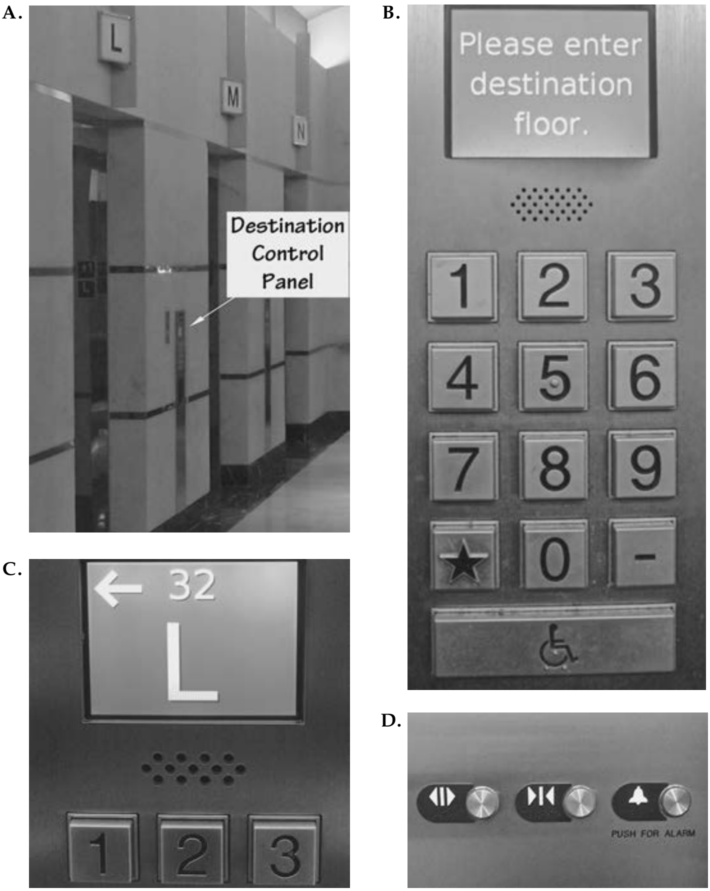

FIGURE 4.8. Destination-Control Elevators. In a destination-control system, the desired destination floor is entered into the control panel outside the elevators (A and B). After entering the destination floor into B, the display directs the traveler to the appropriate elevator, as shown in C, where "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?