CHAPTER

Our Perception is Biased

1

Our perception of the world around us is not a true depiction of what is actually there. Our perceptions are heavily biased by at least three factors:

- ***The past:*** our experience
- ***The present:*** the current context
- ***The future:*** our goals

PERCEPTION BIASED BY EXPERIENCE

Experience—your past perceptions—can bias your current perception in several different ways.

Perceptual priming

Imagine that you own a large insurance company. You are meeting with a real estate manager, discussing plans for a new campus of company buildings. The campus consists of a row of five buildings, the last two with T-shaped courtyards providing light for the cafeteria and fitness center. If the real estate manager showed you the map in Figure 1.1, you would see five black shapes representing the buildings.

Now imagine that instead of a real estate manager, you are meeting with an advertising manager. You are discussing a new billboard ad to be placed in certain markets around the country. The advertising manager shows you the same image, but in this scenario the image is a sketch of the ad, consisting of a single word: LIFE. In this scenario, you see a word, clearly and unambiguously.

When your perceptual system has been *primed* to see building shapes, you see building shapes, and the white areas between the buildings barely register in your perception. When your perceptual system has been primed to see text, you see text, and the black areas between the letters barely register.

Designing with the Mind in Mind. http://dx.doi.org/10.1016/B978-0-12-407914-4.00001-4

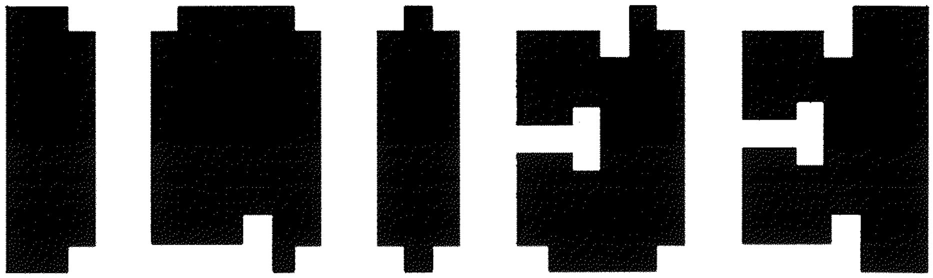


FIGURE 1.1

Building map or word? What you see depends on what you were told to see.

FIGURE 1.2

Image showing the effect of mental priming of the visual system. What do you see?

A relatively famous example of how priming the mind can affect perception is an image, supposedly by R. C. James,[1] that initially looks to most people like a random splattering of paint (see Fig. 1.2) similar to the work of the painter Jackson Pollack. Before reading further, look at the image.

Only after you are told that it is a Dalmatian dog sniffing the ground near a tree can your visual system organize the image into a coherent picture. Moreover, once you've seen the dog, it is hard to go back to seeing just a random collection of spots.

[1] Published in Lindsay and Norman (1972), Figure 3-17, p. 146.

These priming examples are visual, but priming can also bias other types of perception, such as sentence comprehension. For example, the headline "New Vaccine Contains Rabies" would probably be understood differently by people who had recently heard stories about contaminated vaccines than by people who had recently heard stories about successful uses of vaccines to fight diseases.

Familiar perceptual patterns or frames

Much of our lives are spent in familiar situations: the rooms in our homes, our yards, our routes to and from school or work, our offices, neighborhood parks, stores, restaurants, etc. Repeated exposure to each type of situation builds a pattern in our minds of what to expect to see there. These *perceptual patterns*, which some researchers call *frames*, include the objects or events that are usually encountered in that situation.

For example, you know most rooms in your home well enough that you need not constantly scrutinize every detail. You know how they are laid out and where most objects are located. You can probably navigate much of your home in total darkness. But your experience with homes is broader than your specific home. In addition to having a pattern for your home, your brain has one for homes in general. It biases your perception of all homes, familiar and new. In a kitchen, you expect to see a stove and a sink. In a bathroom, you expect to see a toilet, a sink, and a shower or a bathtub (or both).

Mental frames for situations bias our perception to see the objects and events expected in each situation. They are a mental shortcut: by eliminating the need for us to constantly scrutinize every detail of our environment, they help us get around in our world. However, mental frames also make us see things that aren't really there.

For example, if you visit a house in which there is no stove in the kitchen, you might nonetheless later recall seeing one, because your mental frame for kitchens has a strong stove component. Similarly, part of the frame for eating at a restaurant is paying the bill, so you might recall paying for your dinner even if you absentmindedly walked out without paying. Your brain also has frames for back yards, schools, city streets, business offices, supermarkets, dentist visits, taxis, air travel, and other familiar situations.

Anyone who uses computers, websites, or smartphones has frames for the desktop and files, web browsers, websites, and various types of applications and online services. For example, when they visit a new Web site, experienced Web users expect to see a site name and logo, a navigation bar, some other links, and maybe a search box. When they book a flight online, they expect to specify trip details, examine search results, make a choice, and make a purchase.

Because of the perceptual frames users of computer software and websites have, they often click buttons or links without looking carefully at them. Their perception of the display is based more on what their frame for the situation leads them to expect than on what is actually on the screen. This sometimes confounds software designers, who expect users to see what is on the screen—but that isn't how human vision works.

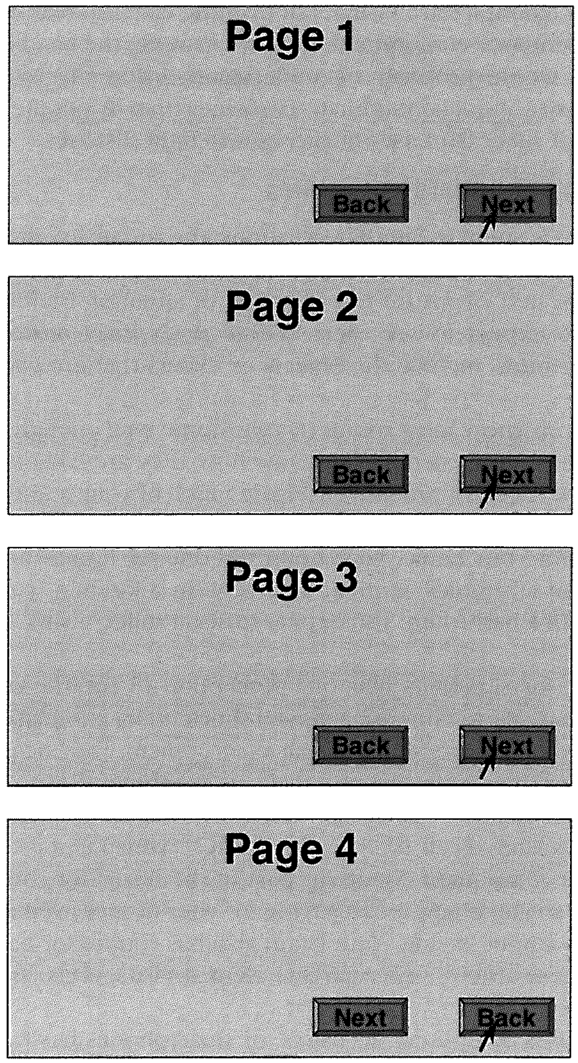


FIGURE 1.3

The "Next" button is perceived to be in a consistent location, even when it isn't.

For example, if the positions of the "Next" and "Back" buttons on the last page of a multistep dialog box[2] switched, many people would not immediately notice the switch (see Fig. 1.3). Their visual system would have been lulled into inattention by the consistent placement of the buttons on the prior several pages. Even after unintentionally going backward a few times, they might continue to perceive the buttons

[2] Multistep dialog boxes are called *wizards* in user-interface designer jargon.

in their standard locations. This is why consistent placement of controls is a common user-interface guideline, to ensure that reality matches the user's frame for the situation.

Similarly, if we are trying to find something but it is in a different place or looks different from usual, we might miss it even though it is in plain view because our mental frames tune us to look for expected features in expected locations. For example, if the "Submit" button on one form in a Web site is shaped differently or is a different color from those on other forms on the site, users might not find it. This expectation-induced blindness is discussed more later in this chapter in the "Perception Biased by Goals" section.

Habituation

A third way in which experience biases perception is called *habituation*. Repeated exposure to the same (or highly similar) perceptions dulls our perceptual system's sensitivity to them. Habituation is a very low-level phenomenon of our nervous system: it occurs at a neuronal level. Even primitive animals like flatworms and amoeba, with very simple nervous systems, habituate to repeated stimuli (e.g., mild electric shocks or light-flashes). People, with our complex nervous systems, habituate to a range of events, from low-level ones like a continually beeping tone, to medium-level ones like a blinking ad on a Web site, to high-level ones like a person who tells the same jokes at every party or a politician giving a long, repetitious speech.

We experience habituation in computer usage when the same error messages or "Are you sure?" confirmation messages appear again and again. People initially notice them and perhaps respond, but eventually click them closed reflexively without bothering to read them.

Habituation is also a factor in a recent phenomenon variously labeled "social media burnout" (Nichols, 2013), "social media fatigue," or "Facebook vacations" (Rainie et al., 2013): newcomers to social media sites and tweeting are initially excited by the novelty of microblogging about their experiences, but sooner or later get tired of wasting time reading tweets about every little thing that their "friends" do or see—for example, "Man! Was that ever a great salmon salad I had for lunch today."

Attentional blink

Another low-level biasing of perception by past experience occurs just after we spot or hear something important. For a very brief period following the recognition—between 0.15 and 0.45 second—we are nearly deaf and blind to other visual stimuli, even though our ears and eyes stay functional. Researchers call this the *attentional blink* (Raymond et al., 1992, Stafford and Webb, 2005).[3] It is thought to be caused by the brain's perceptual and attention mechanisms being briefly fully occupied with processing the first recognition.

[3] Chapter 14 discusses the attentional blink interval in the context of other perceptual intervals.

A classic example: You are in a subway car as it enters a station, planning to meet two friends at that station. As the train arrives, your car passes one of your friends, and you spot him briefly through your window. In the next split second, your window passes your other friend, but you fail to notice her because her image hit your retina during the attentional blink that resulted from your recognition of your first friend.

When people use computer-based systems and online services, attentional blink can cause them to miss information or events if things appear in rapid succession. A popular modern technique for making documentary videos is to present a series of still photographs in rapid succession.[4] This technique is highly prone to attentional blink effects: if an image really captures your attention (e.g., it has a strong meaning for you), you will probably miss one or more of the immediately following images. In contrast, a captivating image in an auto-running slideshow (e.g., on a Web site or an information kiosk) is unlikely to cause attentional blink (i.e., missing the next image) because each image typically remains displayed for several seconds.

PERCEPTION BIASED BY CURRENT CONTEXT

When we try to understand how our visual perception works, it is tempting to think of it as a bottom-up process, combining basic features such as edges, lines, angles, curves, and patterns into figures and ultimately into meaningful objects. To take reading as an example, you might assume that our visual system first recognizes shapes as letters and then combines letters into words, words into sentences, and so on.

But visual perception—reading in particular—is not strictly a bottom-up process. It includes top-down influences too. For example, the word in which a character appears may affect how we identify the character (see Fig. 1.4).

Similarly, our overall comprehension of a sentence or a paragraph can even influence what words we see in it. For example, the same letter sequence can be read as different words depending on the meaning of the surrounding paragraph (see Fig. 1.5).

Contextual biasing of vision need not involve reading. The Müller-Lyer illusion is a famous example (see Fig. 1.6): the two horizontal lines are the same length, but the outward-pointing "fins" cause our visual system to see the top line as longer than the

FIGURE 1.4

The same character is perceived as H or A depending on the surrounding letters.

[4] For an example, search YouTube for "history of the world in two minutes."

Fold napkins. *Polish silverware.* Wash dishes.

French napkins. *Polish silverware.* German dishes.

FIGURE 1.5

The same phrase is perceived differently depending on the list it appears in.

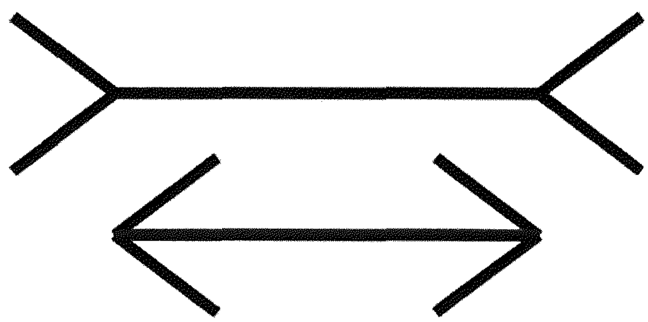


FIGURE 1.6

Müller–Lyer illusion: equal-length horizontal lines appear to have different lengths.

line with inward-pointing "fins." This and other optical illusions (see Fig. 1.7) trick us because our visual system does not use accurate, optimal methods to perceive the world. It developed through evolution, a semi-random process that layers jury-rigged—often incomplete and inaccurate—solutions on top of each other. It works fine most of the time, but it includes a lot of approximations, kludges, hacks, and outright "bugs" that cause it to fail in certain cases.

The examples in Figures 1.6 and 1.7 show vision being biased by *visual* context. However, biasing of perception by the current context works *between* different senses too. Perceptions in any of our five senses may affect simultaneous perceptions in any of our other senses. What we feel with our tactile sense can be biased by what we hear, see, or smell. What we see can be biased by what we hear, and what we hear can be biased by what we see. The following two examples of visual perception affect what we hear:

- ***McGurk effect.*** If you watch a video of someone saying "bah, bah, bah," then "dah, dah, dah," then "vah, vah, vah," but the audio is "bah, bah, bah" throughout, you will hear the syllable indicated by the speaker's lip movement rather than the syllable actually in the audio track.[5] Only by closing or averting your eyes do you hear the syllable as it really is. I'll bet you didn't know you could read lips, and in fact do so many times a day.

[5] Go to YouTube, search for "McGurk effect," and view (and hear) some of the resulting videos.

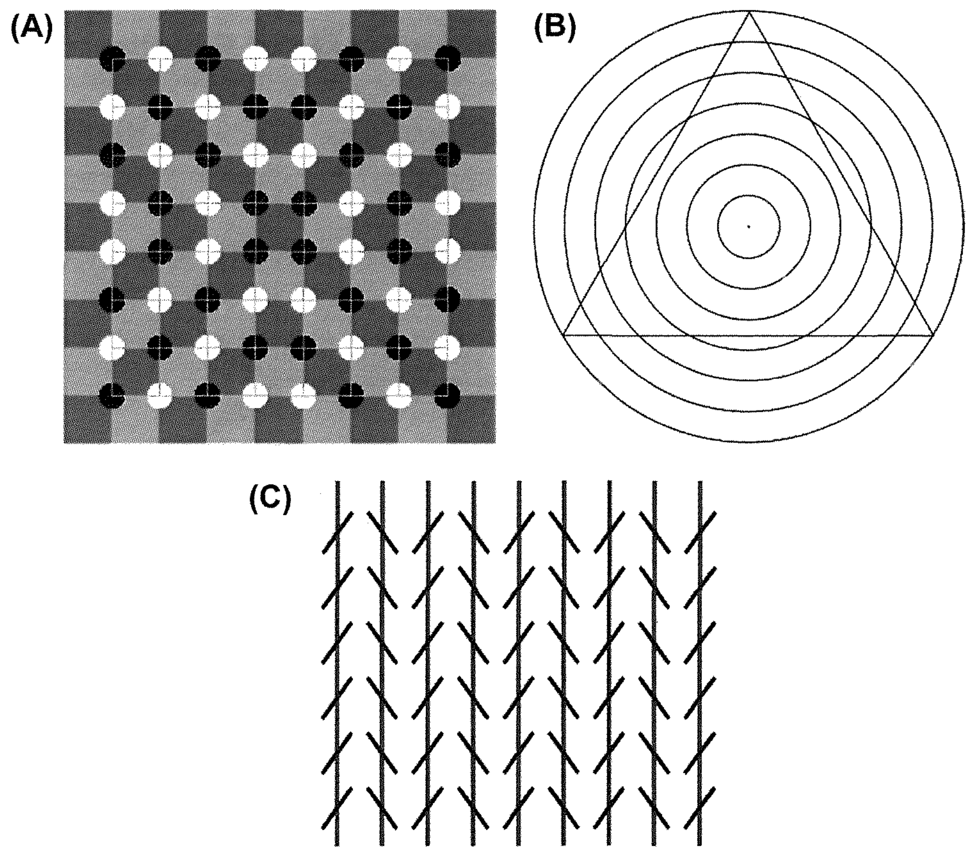


FIGURE 1.7

(A) The checkboard does *not* bulge in the middle; (B) the triangle sides are *not* bent; and (C) the red vertical lines are parallel.

- ***Ventriloquism.*** Ventriloquists don't throw their voice; they just learn to talk without moving their mouths much. Viewers' brains perceive the talking as coming from the nearest moving mouth: that of the ventriloquist's puppet (Eagleman, 2012).

An example of the opposite—hearing biasing vision—is the illusory flash effect. When a spot is flashed *once* briefly on a display but is accompanied by *two* quick beeps, it appears to flash twice. Similarly, the perceived rate of a blinking light can be adjusted by the frequency of a repeating click (Eagleman, 2012).

Later chapters explain how visual perception, reading, and recognition function in the human brain. For now, I will simply say that the pattern of neural activity that corresponds to recognizing a letter, a word, a face, or any object includes input from neural activity stimulated by the context. This context includes other nearby perceived objects and events, and even reactivated memories of previously perceived objects and events.

Context biases perception not only in people but also in lower animals. A friend of mine often brought her dog with her in her car when running errands. One day, as she drove into her driveway, a cat was in the front yard. The dog saw it and began barking. My friend opened the car door and the dog jumped out and ran after the cat, which turned and jumped through a bush to escape. The dog dove into the bush but missed the cat. The dog remained agitated for some time afterward.

Thereafter, for as long as my friend lived in that house, whenever she arrived at home with her dog in the car, he would get excited, bark, jump out of the car as soon as the door was opened, dash across the yard, and leap into the bush. There was no cat, but that didn't matter. Returning home in the car was enough to make the dog see one—perhaps even smell one. However, walking home, as the dog did after being taken for his daily walk, did not evoke the "cat mirage."

PERCEPTION BIASED BY GOALS

In addition to being biased by our past experience and the present context, our perception is influenced by our goals and plans for the future. Specifically, our goals:

- ***Guide*** our perceptual apparatus, so we sample what we need from the world around us.
- ***Filter*** our perceptions: things unrelated to our goals tend to be filtered out preconsciously, never registering in our conscious minds.

For example, when people navigate through software or a Web site, seeking information or a specific function, they don't read carefully. They scan screens quickly and superficially for items that seem related to their goal. They don't simply ignore items unrelated to their goals; they often don't even notice them.

To see this, glance at Figure 1.8 and look for scissors, and then immediately flip back to this page. Try it now.

Did you spot the scissors? Now, without looking back at the toolbox, can you say whether there is a screwdriver in the toolbox too?

Our goals filter our perceptions in other perceptual senses as well as in vision. A familiar example is the "cocktail party" effect. If you are conversing with someone at a crowded party, you can focus your attention to hear mainly what he or she is saying even though many other people are talking near you. The more interested you are in the conversation, the more strongly your brain filters out surrounding chatter. If you are bored by what your conversational partner is saying, you will probably hear much more of the conversations around you.

The effect was first documented in studies of air-traffic controllers, who were able to carry on a conversation with the pilots of their assigned aircraft even though

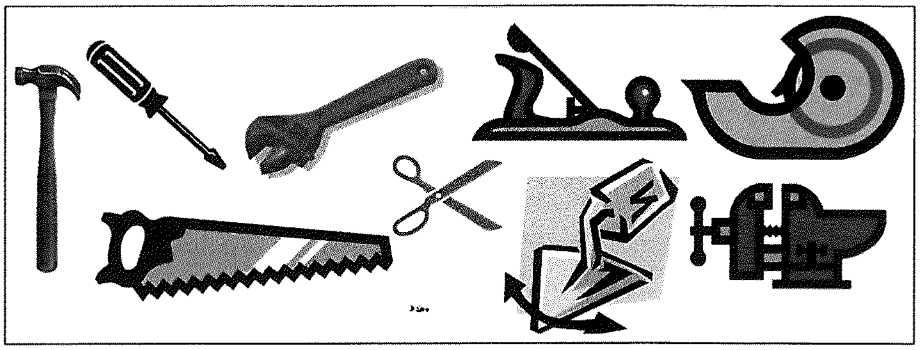


FIGURE 1.8

Toolbox: Are there scissors here?

many different conversations were occurring simultaneously on the same radio frequency, coming out of the same speaker in the control room (Arons, 1992). Research suggests that our ability to focus on one conversation among several simultaneous ones depends not only on our interest level in the conversation, but also on objective factors, such as the similarity of voices in the cacophony, the amount of general "noise" (e.g., clattering dishes or loud music), and the predictability of what your conversational partner is saying (Arons, 1992).

This filtering of perception by our goals is particularly true for adults, who tend to be more focused on goals than children are. Children are more stimulus-driven: their perception is less filtered by their goals. This characteristic makes them more distractible than adults, but it also makes them less biased as observers.

A parlor game demonstrates this age difference in perceptual filtering. It is similar to the Figure 1.8 exercise. Most households have a catch-all drawer for kitchen implements or tools. From your living room, send a visitor to the room where the catch-all drawer is, with instructions to fetch you a specific tool, such as measuring spoons or a pipe wrench. When the person returns with the tool, ask whether another specific tool was in the drawer. Most adults will not know what else was in the drawer. Children—if they can complete the task without being distracted by all the cool stuff in the drawer—will often be able to tell you more about what else was there.

Perceptual filtering can also be seen in how people navigate websites. Suppose I put you on the homepage of New Zealand's University of Canterbury (see Fig. 1.9) and asked you to find information about financial support for postgraduate students in the computer science department. You would scan the page and probably quickly click one of the links that share words with the goal that I gave you: Departments (top left), Scholarships (middle), then Postgraduate Students (bottom left) or Postgraduate (right). If you're a "search" person, you might instead go right to the Search box (top right), type words related to the goal, and click "Go."

Whether you browse or search, it is likely that you would leave the homepage without noticing that you were randomly chosen to win $100 (bottom right). Why? Because that was not related to your *goal*.

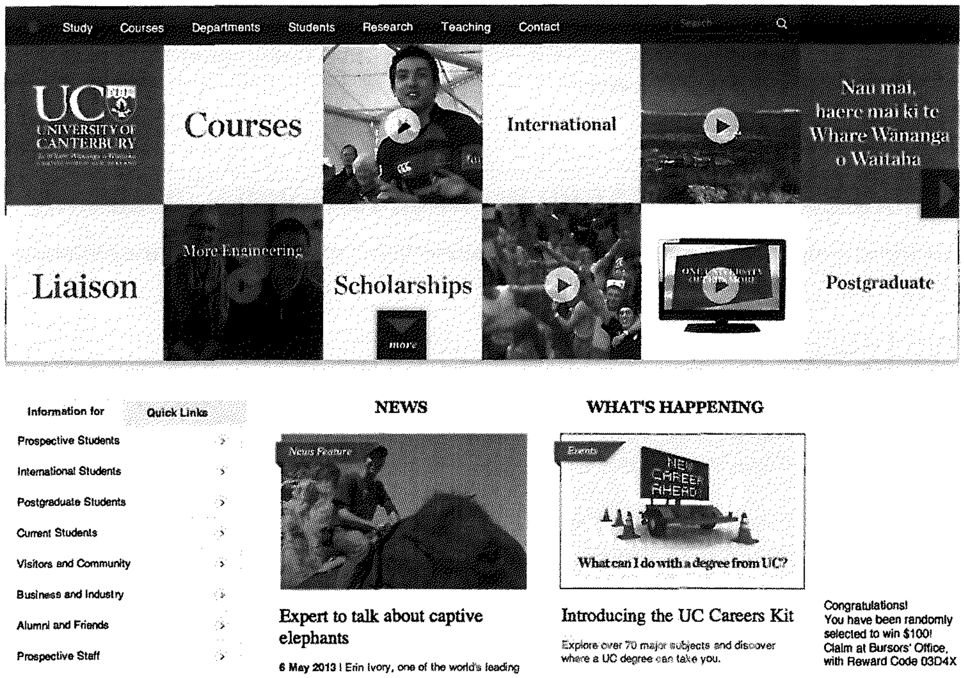


FIGURE 1.9

University of Canterbury Web site: navigating sites requires perceptual filtering.

What is the mechanism by which our current goals bias our perception? There are two:

- ***Influencing where we look.*** Perception is active, not passive. Think of your perceptual senses not as simply *filtering* what comes to you, but rather as *reaching out* into the world and *pulling in* what you need to perceive. Your hands, your primary touch sensors, literally do this, but the rest of your senses do it too. You constantly move your eyes, ears, hands, feet, body, and attention so as to sample exactly the things in your environment that are most relevant to what you are doing or about to do (Ware, 2008). If you are looking on a Web site for a campus map, your eyes and pointer-controlling hand are attracted to anything that might lead you to that goal. You more or less ignore anything unrelated to your goal.

- ***Sensitizing our perceptual system to certain features.*** When you are looking for something, your brain can prime your perception to be especially sensitive to features of what you are looking for (Ware, 2008). For example, when you are looking for a red car in a large parking lot, red cars will seem to pop out as you scan the lot, and cars of other colors will barely register in your consciousness, even though you do in some sense see them. Similarly, when you are trying to find your spouse in a dark, crowded room, your brain "programs" your auditory system to be especially sensitive to the combination of frequencies that make up his or her voice.

TAKING BIASED PERCEPTION INTO ACCOUNT WHEN DESIGNING

All these sources of perceptual bias of course have implications for user-interface design. Here are three.

Avoid ambiguity

Avoid ambiguous information displays, and test your design to verify that all users interpret the display in the same way. Where ambiguity is unavoidable, either rely on standards or conventions to resolve it, or prime users to resolve the ambiguity in the intended way.

For example, computer displays often shade buttons and text fields to make them look raised in relation to the background surface (see Fig. 1.10). This appearance relies on a convention, familiar to most experienced computer users, that the light source is at the top left of the screen. If an object were depicted as lit by a light source in a different location, users would not see the object as raised.

Be consistent

Place information and controls in consistent locations. Controls and data displays that serve the same function on different pages should be placed in the same position on each page on which they appear. They should also have the same color, text fonts, shading, and so on. This consistency allows users to spot and recognize them quickly.

Understand the goals

Users come to a system with goals they want to achieve. Designers should understand those goals. Realize that users' goals may vary, and that their goals strongly influence what they perceive. Ensure that at every point in an interaction, the information users need is available, prominent, and maps clearly to a possible user goal, so users will notice and use the information.

FIGURE 1.10

Buttons on computer screens are often shaded to make them look three dimensional, but the convention works only if the light source is assumed to be on the top left.

CHAPTER

Our Vision is Optimized to See Structure

2

Early in the twentieth century, a group of German psychologists sought to explain how human visual perception works. They observed and catalogued many important visual phenomena. One of their basic findings was that human vision is holistic: our visual system automatically imposes structure on visual input and is wired to perceive whole shapes, figures, and objects rather than disconnected edges, lines, and areas. The German word for "shape" or "figure" is *Gestalt,* so these theories became known as the *Gestalt principles of visual perception.*

Today's perceptual and cognitive psychologists regard the Gestalt theory of perception as more of a *descriptive* framework than an *explanatory* and *predictive* theory. Today's theories of visual perception tend to be based heavily on the neurophysiology of the eyes, optic nerve, and brain (see Chapters 4–7).

Not surprisingly, the findings of neurophysiological researchers support the observations of the Gestalt psychologists. We really are—along with other animals—"wired" to perceive our surroundings in terms of whole objects (Stafford and Webb, 2005; Ware, 2008). Consequently, the Gestalt principles are still valid—if not as a fundamental explanation of visual perception, at least as a framework for describing it. They also provide a useful basis for guidelines for graphic design and user-interface design (Soegaard, 2007).

For present purposes, the most important Gestalt principles are Proximity, Similarity, Continuity, Closure, Symmetry, Figure/Ground, and Common Fate. The following sections describe each principle and provide examples from both static graphic design and user-interface design.

GESTALT PRINCIPLE: PROXIMITY

The Gestalt principle of *Proximity* is that the relative distance between objects in a display affects our perception of whether and how the objects are organized into

Designing with the Mind in Mind. http://dx.doi.org/10.1016/B978-0-12-407914-4.00002-6

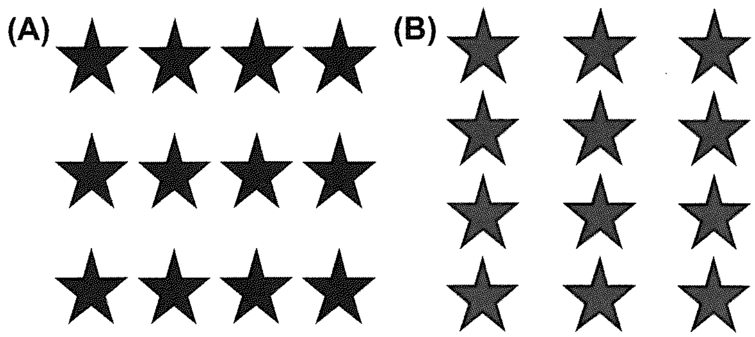


FIGURE 2.1

Proximity: items that are closer appear grouped as rows (A) and columns (B).

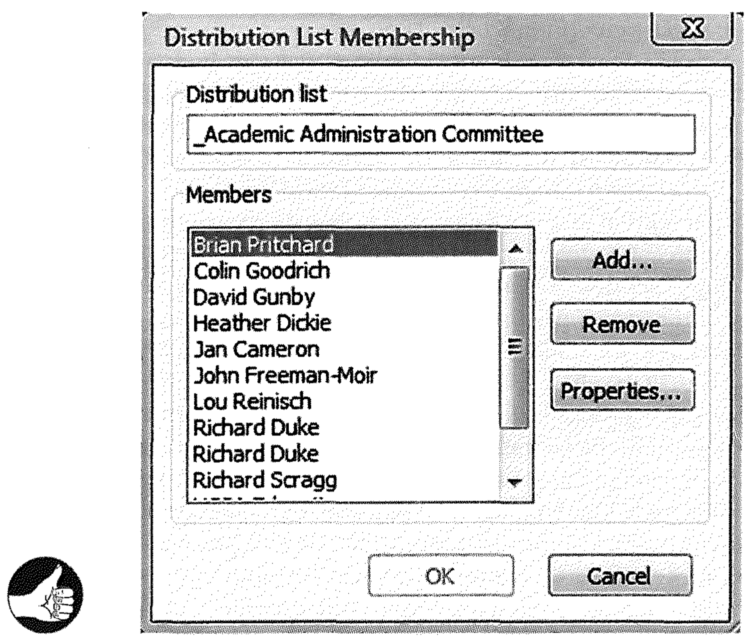


FIGURE 2.2

In Outlook's Distribution List Membership dialog box, list buttons are in a group box, separate from the control buttons.

subgroups. Objects that are near each other (relative to other objects) appear grouped, while those that are farther apart do not.

In Figure 2.1A, the stars are closer together horizontally than they are vertically, so we see three rows of stars, while the stars in Figure 2.1B are closer together vertically than they are horizontally, so we see three columns.

The Proximity principle has obvious relevance to the layout of control panels or data forms in software, Web sites, and electronic appliances. Designers often separate groups of on-screen controls and data displays by enclosing them in group boxes or by placing separator lines between groups (see Fig. 2.2).

However, according to the Proximity principle, items on a display can be visually grouped simply by spacing them closer to each other than to other controls, without group boxes or visible borders (see Fig. 2.3). Many graphic design experts recommend this approach to reduce visual clutter and code size in a user interface (Mullet and Sano, 1994).

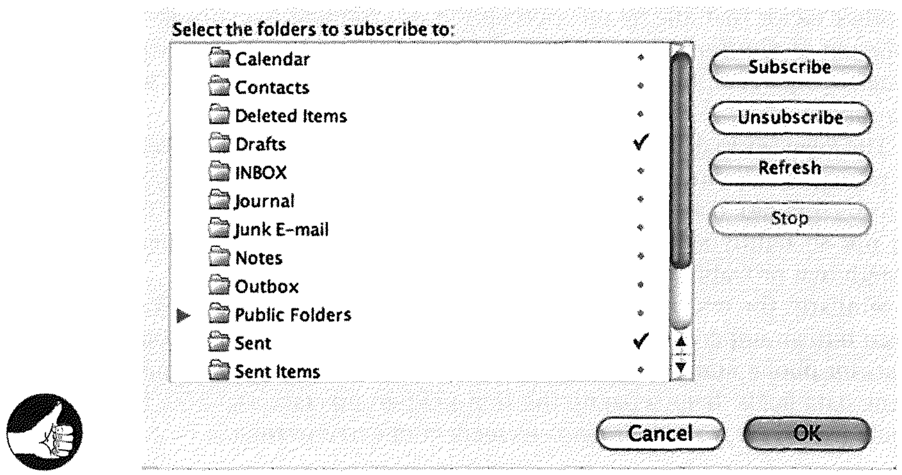


FIGURE 2.3

In Mozilla Thunderbird's Subscribe Folders dialog box, controls are grouped using the Proximity principle.

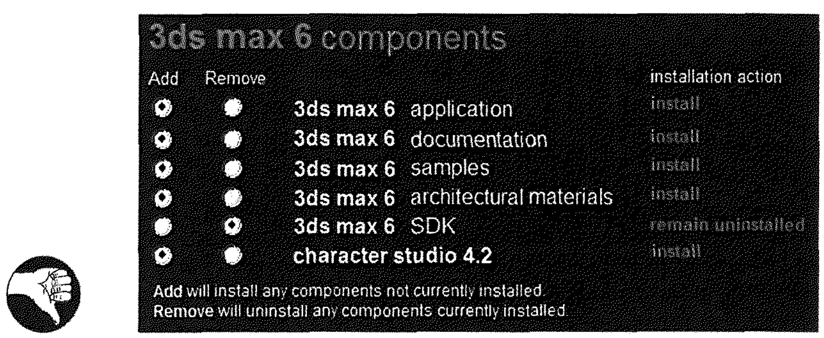


FIGURE 2.4

In Discreet's Software Installer, poorly spaced radio buttons look grouped in vertical columns.

Conversely, if controls are poorly spaced (e.g., if connected controls are too far apart) people will have trouble perceiving them as related, making the software harder to learn and remember. For example, the Discreet Software Installer displays six horizontal pairs of radio buttons, each representing a two-way choice, but their spacing, due to the Proximity principle, makes them appear to be two vertical sets of radio buttons, each representing a six-way choice, at least until users try them and learn how they operate (see Fig. 2.4).

GESTALT PRINCIPLE: SIMILARITY

Another factor that affects our perception of grouping is expressed in the Gestalt principle of *Similarity*, where objects that look similar appear grouped, all other things being equal. In Figure 2.5, the slightly larger, "hollow" stars are perceived as a group.

The Page Setup dialog box in Mac OS applications uses the Similarity and Proximity principles to convey groupings (see Fig. 2.6). The three very similar and tightly spaced Orientation settings are clearly intended to appear grouped. The three menus are not so tightly spaced but look similar enough that they appear related even though that probably wasn't intended.

Similarly, the text fields in a form at book publisher Elsevier's Web site are organized into an upper group of eight for the name and address, a group of three split fields for phone numbers, and two single text fields. The four menus, in addition to being data fields, help separate the text field groups (see Fig. 2.7). By contrast, the labels are too far from their fields to seem connected to them.

FIGURE 2.5

Similarity: items appear grouped if they look more similar to each other than to other objects.

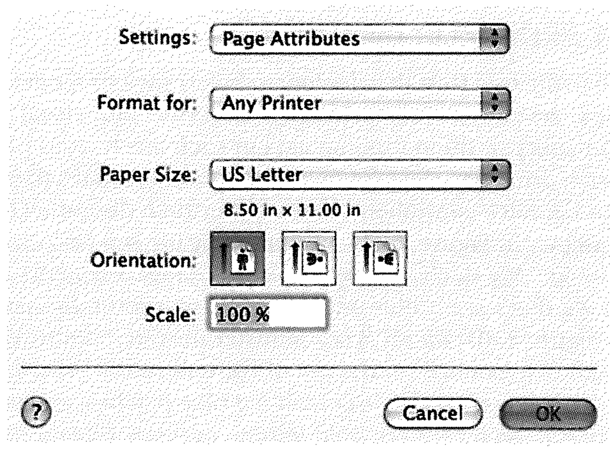


FIGURE 2.6

Mac OS Page Setup dialog box. The Similarity and Proximity principles are used to group the Orientation settings.

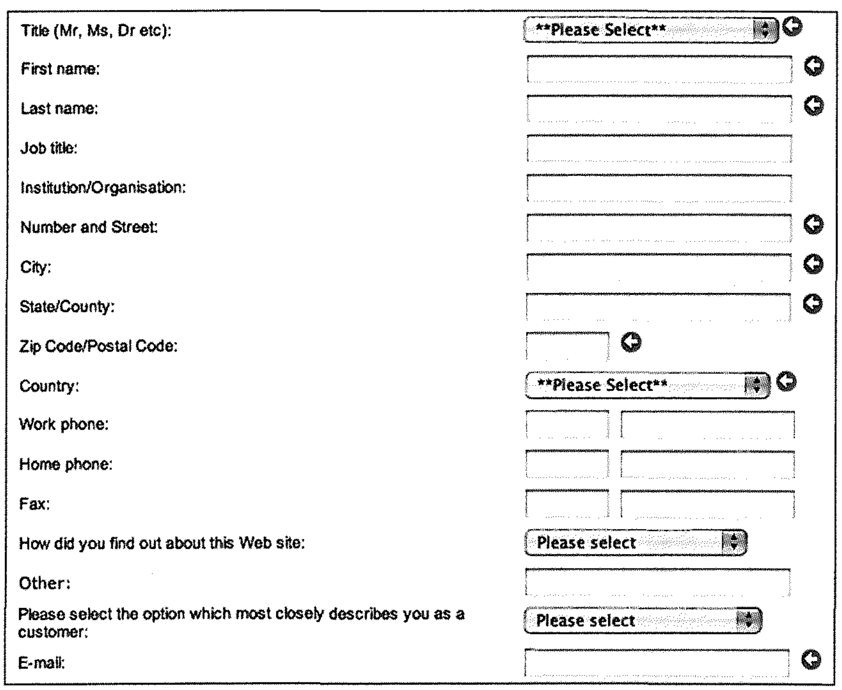


FIGURE 2.7

Similarity makes the text fields appear grouped in this online form at Elsevier.com.

GESTALT PRINCIPLE: CONTINUITY

In addition to the two Gestalt principles concerning our tendency to organize objects into groups, several Gestalt principles describe our visual system's tendency to resolve ambiguity or fill in missing data in such a way as to perceive whole objects. The first such principle, the principle of *Continuity,* states that our visual perception is biased to perceive continuous forms rather than disconnected segments.

For example, in Figure 2.8A, we automatically see two crossing lines—one blue and one orange. We don't see two separate orange segments and two separate blue ones, and we don't see a blue-and-orange V on top of an upside-down orange-and-blue V. In Figure 2.8B, we see a sea monster in water, not three pieces of one.

A well-known example of the use of the continuity principle in graphic design is the IBM® logo. It consists of disconnected blue patches, and yet it is not at all ambiguous; it is easily seen as three bold letters, perhaps viewed through something like venetian blinds (see Fig. 2.9).

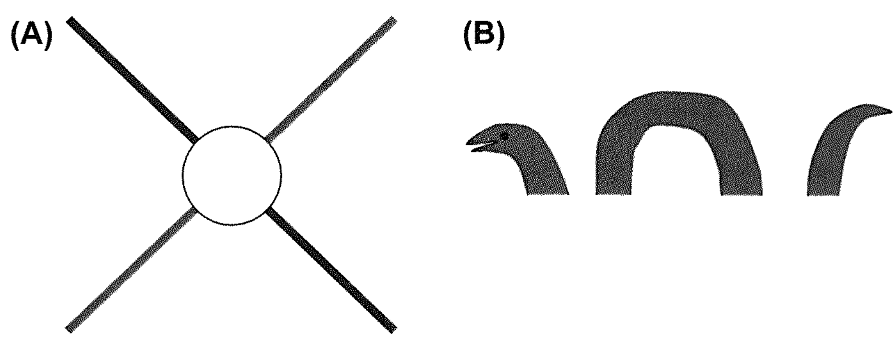


FIGURE 2.8

Continuity: Human vision is biased to see continuous forms, even adding missing data if necessary.

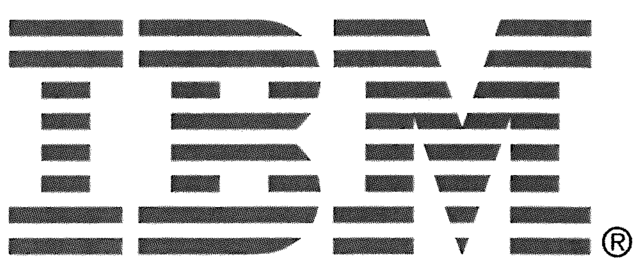


FIGURE 2.9

The IBM company logo uses the Continuity principle to form letters from disconnected patches.

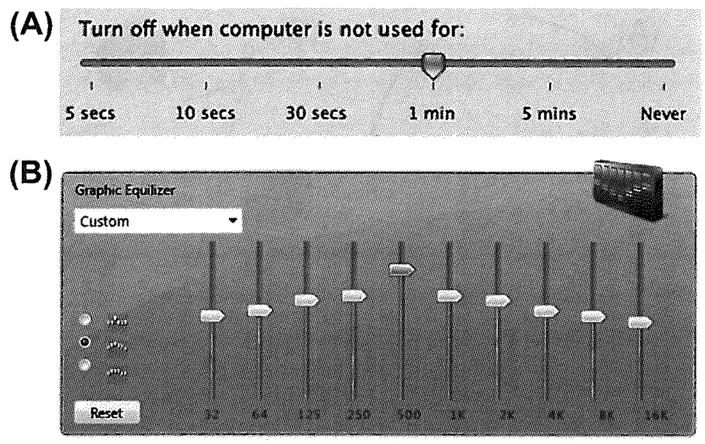


FIGURE 2.10

Continuity: we see a slider as a single slot with a handle somewhere on it, not as two slots separated by a handle: (A) Mac OS and (B) ComponentOne.

Slider controls are a user-interface example of the Continuity principle. We see a slider as depicting a single range controlled by a handle that appears somewhere on the slider, not as two separate ranges separated by the handle (see Fig. 2.10A). Even displaying different colors on each side of a slider's handle doesn't completely "break" our perception of a slider as one continuous object, although ComponentOne's choice of strongly contrasting colors (gray vs. red) certainly strains that perception a bit (see Fig. 2.10B).

GESTALT PRINCIPLE: CLOSURE

Related to Continuity is the Gestalt principle of *Closure*, which states that our visual system automatically tries to close open figures so that they are perceived as whole objects rather than separate pieces. Thus, we perceive the disconnected arcs in Figure 2.11A as a circle.

Our visual system is so strongly biased to see objects that it can even interpret a totally blank area as an object. We see the combination of shapes in Figure 2.11B as a white triangle overlapping another triangle and three black circles, even though the figure really only contains three V shapes and three black pac-men.

The Closure principle is often applied in graphical user interfaces (GUIs). For example, GUIs often represent collections of objects (e.g., documents or messages) as *stacks* (see Fig. 2.12). Just showing one whole object and the edges of others "behind" it is enough to make users perceive a stack of objects, all whole.

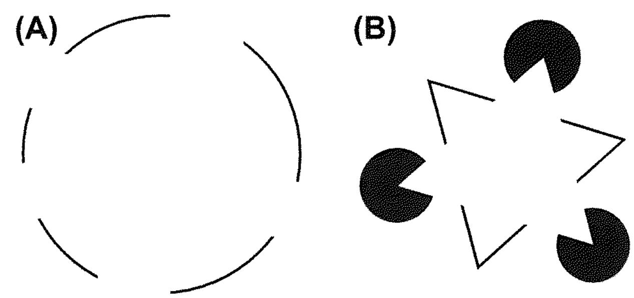


FIGURE 2.11

Closure: Human vision is biased to see whole objects, even when they are incomplete.

FIGURE 2.12

Icons depicting stacks of objects exhibit the Closure principle: partially visible objects are perceived as whole.

GESTALT PRINCIPLE: SYMMETRY

A third fact about our tendency to see objects is captured in the Gestalt principle of *Symmetry*. It states that we tend to parse complex scenes in a way that reduces the complexity. The data in our visual field usually has more than one possible interpretation, but our vision automatically organizes and interprets the data so as to simplify it and give it symmetry.

For example, we see the complex shape on the far left of Figure 2.13 as two overlapping diamonds, not as two touching corner bricks or a pinch-waist octahedron with a square in its center. A pair of overlapping diamonds is simpler than the other two interpretations shown on the right—it has fewer sides and more symmetry than the other two interpretations.

In printed graphics and on computer screens, our visual system's reliance on the symmetry principle can be exploited to represent three-dimensional objects on a two-dimensional display. This can be seen in a cover illustration for Paul Thagard's book *Coherence in Thought and Action* (Thagard, 2002; see Fig. 2.14) and in a three-dimensional depiction of a cityscape (see Fig. 2.15).

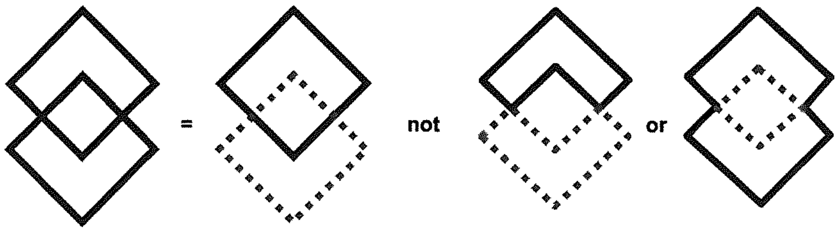


FIGURE 2.13

Symmetry: the human visual system tries to resolve complex scenes into combinations of simple, symmetrical shapes.

FIGURE 2.14

The cover of the book *Coherence in Thought and Action* (Thagard, 2002) uses the symmetry, Closure, and Continuity principles to depict a cube.

GESTALT PRINCIPLE: FIGURE/GROUND

The next Gestalt principle that describes how our visual system structures the data it receives is *Figure/Ground*. This principle states that our mind separates the visual field into the figure (the foreground) and ground (the background). The foreground consists of the elements of a scene that are the object of our primary attention, and the background is everything else.

The Figure/Ground principle also specifies that the visual system's parsing of scenes into figure and ground is influenced by characteristics of the scene. For example, when a small object or color patch overlaps a larger one, we tend to perceive the smaller object as the figure and the larger object as the ground (see Fig. 2.16).

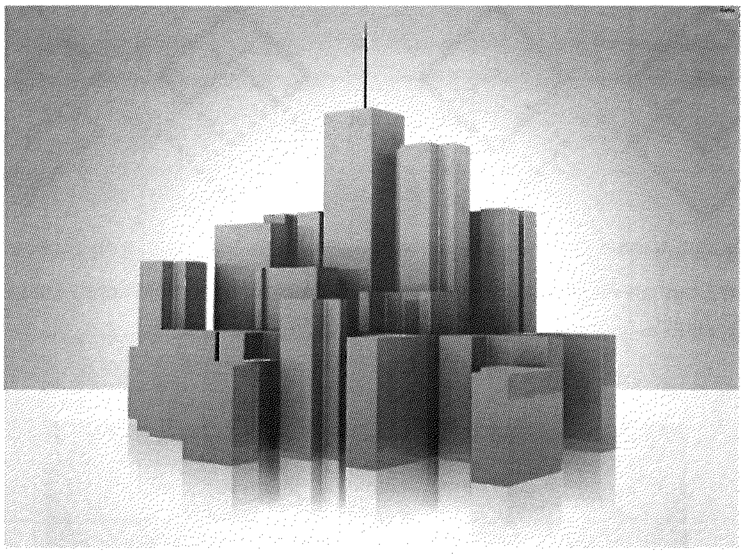


FIGURE 2.15

Symmetry: the human visual system parses very complex two-dimensional images into three-dimensional scenes.

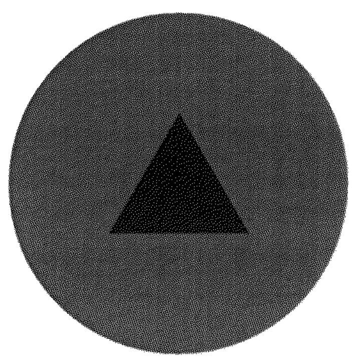


FIGURE 2.16

Figure/Ground: when objects overlap, we see the smaller as the figure and the larger as the ground.

However, our perception of figure versus ground is not completely determined by scene characteristics. It also depends on the viewer's focus of attention. Dutch artist M. C. Escher exploited this phenomenon to produce ambiguous images in which figure and ground switch roles as our attention shifts (see Fig. 2.17).

In user-interface and Web design, the Figure/Ground principle is often used to place an impression-inducing background "behind" the primary displayed content

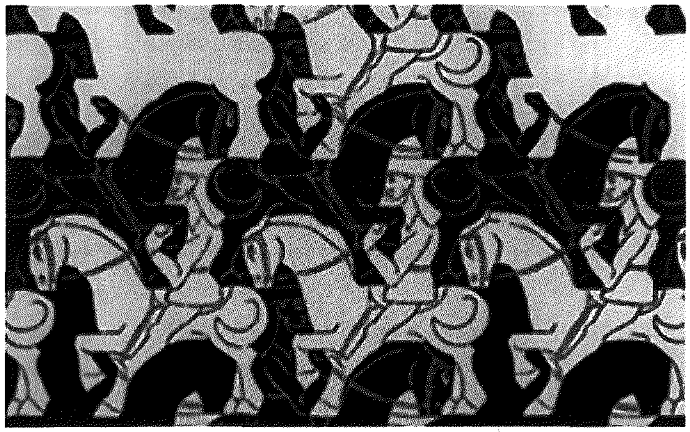


FIGURE 2.17
M. C. Escher exploited figure/ground ambiguity in his art.

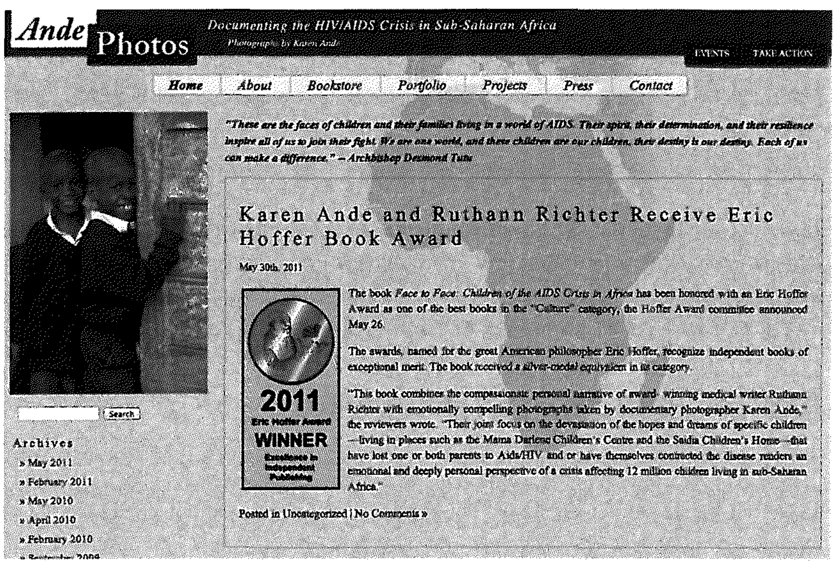


FIGURE 2.18
Figure/Ground is used at AndePhotos.com to display a thematic watermark "behind" the content.

(see Fig. 2.18). The background can convey information (e.g., the user's current location), or it can suggest a theme, brand, or mood for interpretation of the content.

Figure/Ground is also often used to pop up information over other content. Content that was formerly the figure—the focus of the users' attention—temporarily becomes the *background* for new information, which appears briefly as the new

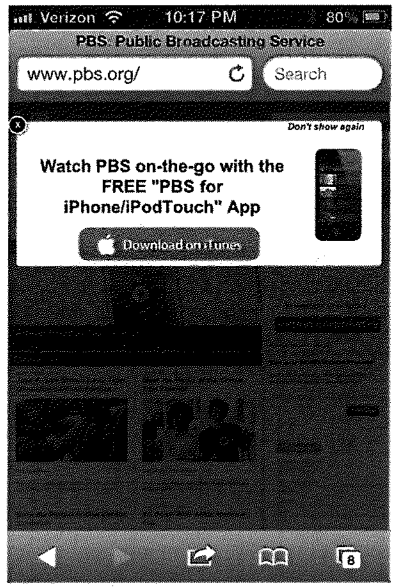


FIGURE 2.19

Figure/Ground is used at PBS.org's mobile Web site to pop up a call-to-action "over" the page content.

figure (see Fig. 2.19). This approach is usually better than temporarily *replacing* the old information with the new information, because it provides context that helps keep people oriented regarding their place in the interaction.

GESTALT PRINCIPLE: COMMON FATE

The previous six Gestalt principles concerned perception of static (unmoving) figures and objects. One final Gestalt principle—Common Fate—concerns moving objects. The cCommon Fate principle is related to the Proximity and Similarity principles—like them, it affects whether we perceive objects as grouped. The Common Fate principle states that objects that move together are perceived as grouped or related.

For example, in a display showing dozens of pentagons, if seven of them wiggled in synchrony, people would see them as a related group, even if the wiggling pentagons were separated from each other and looked no different from all the other pentagons (see Fig. 2.20).

Common motion—implying common fates—is used in some animations to show relationships between entities. For example, Google's GapMinder graphs animate dots representing nations to show changes over time in various factors of economic development. Countries that move together share development histories (see Fig. 2.21).

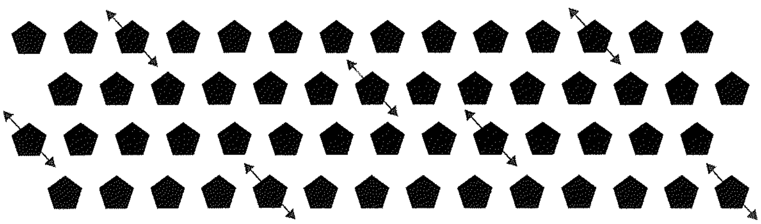


FIGURE 2.20

Common Fate: items appear grouped or related if they move together.

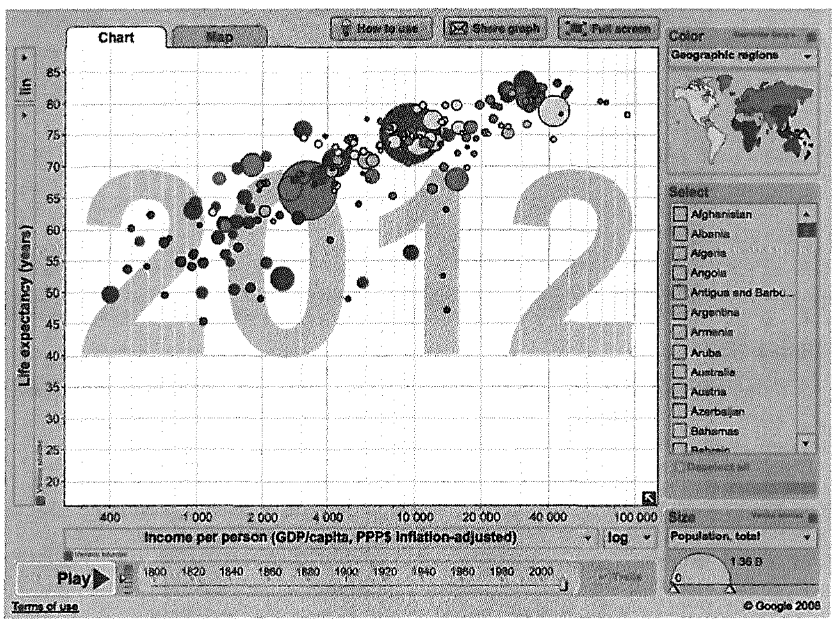


FIGURE 2.21

Common fate: GapMinder animates dots to show which nations have similar development histories (for details, animations, and videos, visit GapMinder.org).

GESTALT PRINCIPLES: COMBINED

Of course, in real-world visual scenes, the Gestalt principles work in concert, not in isolation. For example, a typical Mac OS desktop usually exemplifies six of the seven principles described here, excluding Common Fate): Proximity, Similarity, Continuity, Closure, Symmetry, and Figure/Ground (see Fig. 2.22). On a typical desktop, Common Fate is used (along with similarity) when a user selects several files or folders and drags them as a group to a new location (see Fig. 2.23).

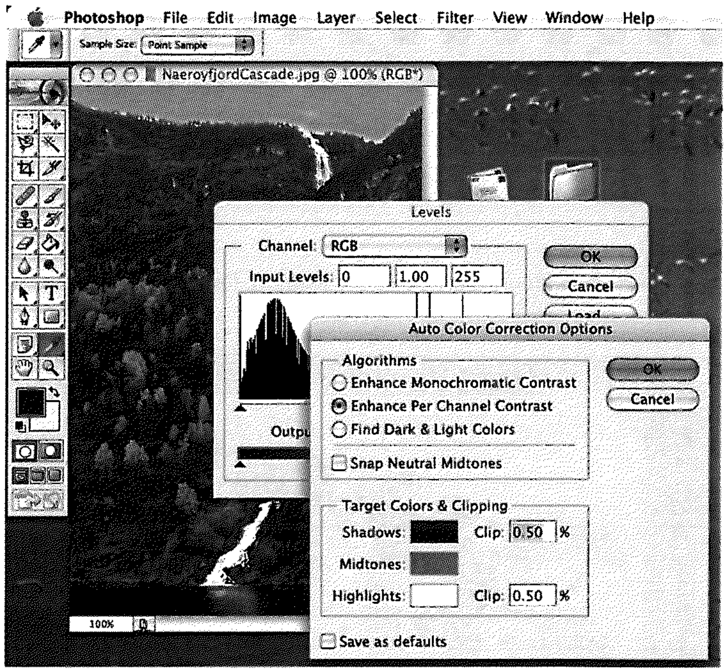


FIGURE 2.22

All of the Gestalt principles except Common Fate play a role in this portion of a Mac OS desktop.

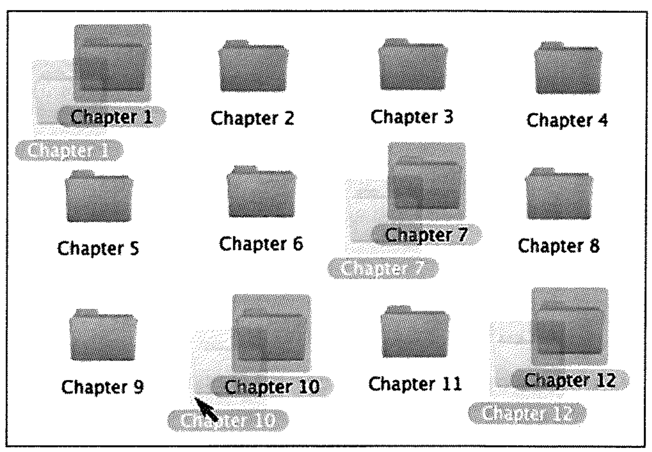


FIGURE 2.23

Similarity and Common Fate: when users drag folders that they have selected, common highlighting and motion make the selected folders appear grouped.

With all these Gestalt principles operating at once, *unintended* visual relationships can be implied by a design. A recommended practice, after designing a display, is to view it with each of the Gestalt principles in mind—Proximity, Similarity, Continuity, Closure, Symmetry, Figure/Ground, and Common Fate—to see if the design suggests any relationships between elements that you do *not* intend.

This page intentionally left blank

CHAPTER

We Seek and Use Visual Structure

3

Chapter 2 used the Gestalt principles of visual perception to show how our visual system is optimized to perceive structure. Perceiving structure in our environment helps us make sense of objects and events quickly. Chapter 2 also mentioned that when people are navigating through software or Web sites, they don't scrutinize screens carefully and read every word. They scan quickly for relevant information. This chapter presents examples to show that when information is presented in a terse, structured way, it is easier for people to scan and understand.

Consider two presentations of the same information about an airline flight reservation. The first presentation is unstructured prose text; the second is structured text in outline form (see Fig. 3.1). The structured presentation of the reservation can be scanned and understood much more quickly than the prose presentation.

Unstructured:

You are booked on United flight 237, which departs from Auckland at 14:30 on Tuesday 15 Oct and arrives at San Francisco at 11:40 on Tuesday 15 Oct.

Structured:

Flight: **United 237, Auckland → San Francisco**
Depart: **14:30 Tue 15 Oct**
Arrive: **11:40 Tue 15 Oct**

FIGURE 3.1

Structured presentation of airline reservation information is easier to scan and understand.

The more structured and terse the presentation of information, the more quickly and easily people can scan and comprehend it. Look at the Contents page from the California Department of Motor Vehicles (see Fig. 3.2). The wordy, repetitive links slow users down and "bury" the important words they need to see.

Designing with the Mind in Mind. http://dx.doi.org/10.1016/B978-0-12-407914-4.00003-8

Renewals, Duplicates, and Information Changes for Driver Licenses and/or ID Cards

- How to renew your driver license in person
- How to renew your driver license by mail
- How to renew your driver license by Internet
- How to renew your instruction permit
- How to apply for a duplicate driver license or identification (ID) card
- How to change your name on your driver license and/or identification (ID) card
- How to notify DMV of my change of address
- How to register for the organ donor gift of life program

FIGURE 3.2

Contents page at the California Department of Motor Vehicles (DMV) Web site buries the important information in repetitive prose.

Licenses & ID Cards:	Renewals, Duplicates, Changes		
• Renew license:	in person	by mail	by Internet
• Renew:	instruction permit		
• Apply for duplicate:	license	ID card	
• Change of:	name	address	
• Register as:	organ donor		

FIGURE 3.3

California DMV Web site Contents page with repetition eliminated and better visual structure.

Compare that with a terser, more structured hypothetical design that factors out needless repetition and marks as links only the words that represent options (see Fig. 3.3). All options presented in the actual Contents page are available in the revision, yet it consumes less screen space and is easier to scan.

Displaying search results is another situation in which structuring data and avoiding repetitive "noise" can improve people's ability to scan quickly and find what they seek. In 2006, search results at HP.com included so much repeated navigation data and metadata for each retrieved item that they were useless. By 2009, HP had eliminated the repetition and structured the results, making them easier to scan and more useful (see Fig. 3.4).

Of course, for information displays to be easy to scan, it is not enough merely to make them terse, structured, and nonrepetitious. They must also conform to the rules of graphic design, some of which were presented in Chapter 2.

For example, a prerelease version of a mortgage calculator on a real estate Web site presented its results in a table that violated at least two important rules of graphic design (see Fig. 3.5A). First, people usually read (online or offline) from top to bottom, but the labels for calculated amounts were *below* their corresponding values. Second, the labels were just as close to the value below as to their own

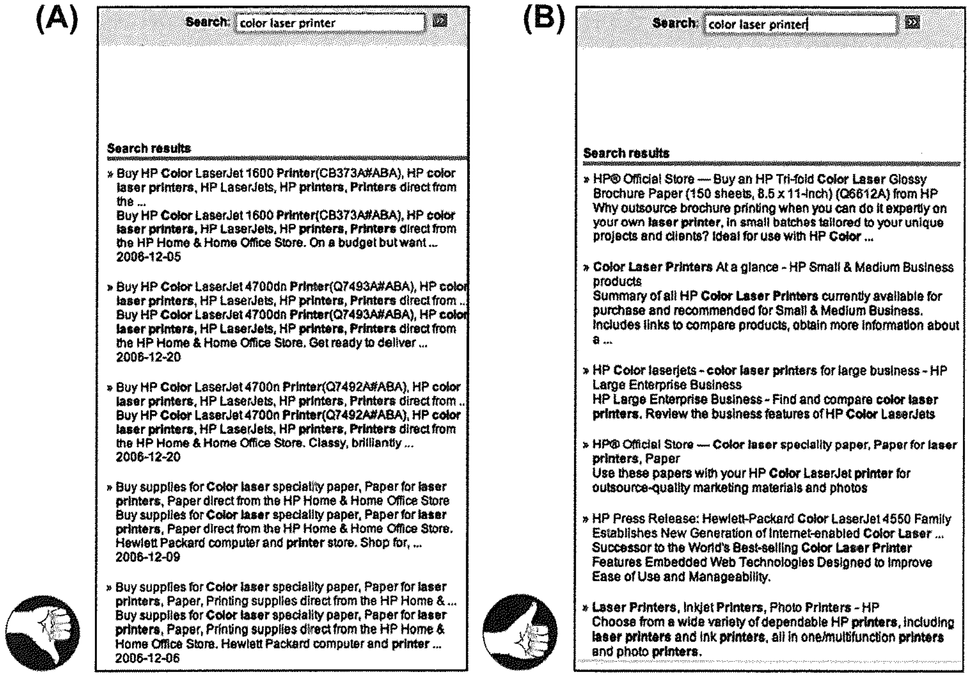


FIGURE 3.4

In 2006, HP.com's site search produced repetitious, "noisy" results (A), but by 2009 was improved (B).

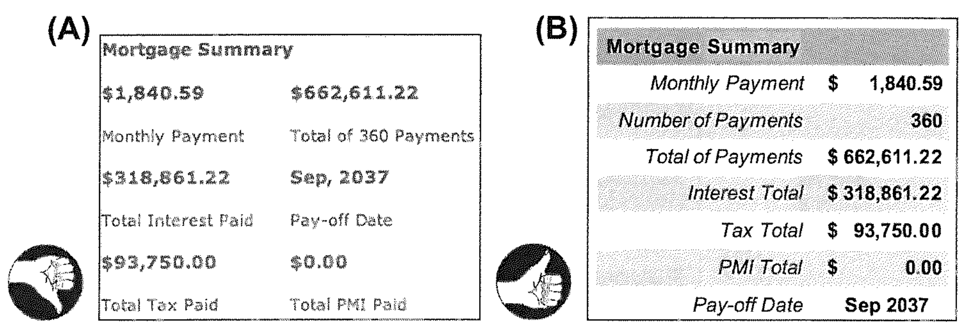


FIGURE 3.5

(A) Mortgage summary presented by a software mortgage calculator; (B) an improved design.

value, so proximity (see Chapter 2) could not be used to perceive that labels were grouped with their values. To understand this mortgage results table, users had to scrutinize it carefully and slowly figure out which labels went with which numbers.

The revised design, in contrast, allows users to perceive the correspondence between labels and values without conscious thought (see Fig. 3.5B).

STRUCTURE ENHANCES PEOPLE'S ABILITY TO SCAN LONG NUMBERS

Even small amounts of information can be made easier to scan if they are structured. Two examples are telephone numbers and credit card numbers (see Fig. 3.6). Traditionally, such numbers were broken into parts to make them easier to scan and remember.

A long number can be broken up in two ways: either the user interface breaks it up explicitly by providing a separate field for each part of the number, or the interface provides a single number field but lets users break the number into parts with spaces or punctuation (see Fig. 3.7A). However, many of today's computer presentations of phone and credit card numbers do not segment the numbers and do not

Easy:	(415) 123-4567
Hard:	4151234567
Easy:	1234 5678 9012 3456
Hard:	1234567890123456

FIGURE 3.6

Telephone and credit card numbers are easier to scan and understand when segmented.

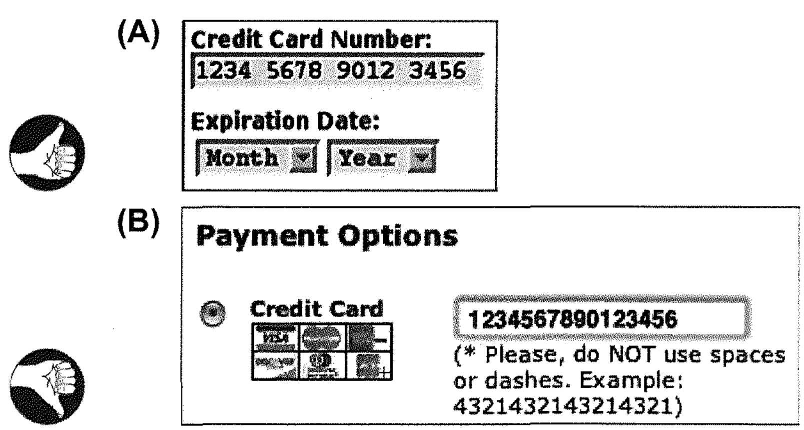


FIGURE 3.7

(A) At Democrats.org, credit card numbers can include spaces. (B) At StuffIt.com, they cannot, making them harder to scan and verify.

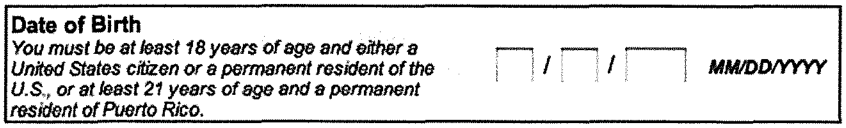


FIGURE 3.8

At BankOfAmerica.com, segmented data fields provide useful structure.

allow users to include spaces or other punctuation (see Fig. 3.7B). This limitation makes it harder for people to scan a number or verify that they typed it correctly, and so is considered a user-interface design blooper (Johnson, 2007). Forms presented in software and Web sites should accept credit card numbers, social security numbers, phone numbers, and so on in a variety of different formats and parse them into the internal format.

Segmenting data fields can provide useful visual structure even when the data to be entered is not, strictly speaking, a number. Dates are an example of a case in which segmented fields can improve readability and help prevent data entry errors, as shown by a date field at Bank of America's Web site (see Fig. 3.8).

DATA-SPECIFIC CONTROLS PROVIDE EVEN MORE STRUCTURE

A step up in structure from segmented data fields are data-specific controls. Instead of using simple text fields—whether segmented or not—designers can use controls that are designed specifically to display (and accept as input) a value of a specific type. For example, dates can be presented (and accepted) in the form of menus combined with pop-up calendar controls (see Fig. 3.9).

It is also possible to provide visual structure by mixing segmented text fields with data-specific controls, as demonstrated by an email address field at Southwest Airlines' Web site (see Fig. 3.10).

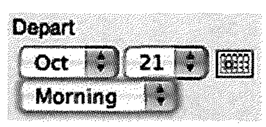


FIGURE 3.9

At NWA.com, dates are displayed and entered using a control that is specifically designed for dates.

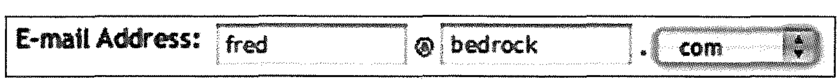


FIGURE 3.10

At SWA.com email addresses are entered into fields structured to accept parts of the address.

VISUAL HIERARCHY LETS PEOPLE FOCUS ON THE RELEVANT INFORMATION

One of the most important goals in structuring information presentations is to provide a visual hierarchy—an arrangement that:

- Breaks the information into distinct sections, and breaks large sections into subsections.
- Labels each section and subsection prominently and in such a way as to clearly identify its content.
- Presents the sections and subsections as a hierarchy, with higher-level sections presented more strongly than lower-level ones.

A visual hierarchy allows people, when scanning information, to instantly separate what is relevant to their goals from what is irrelevant, and to focus their attention on the relevant information. They find what they are looking for more quickly because they can easily skip everything else.

Try it for yourself. Look at the two information displays in Figure 3.11 and find the information about prominence. How much longer does it take you to find it in the nonhierarchical presentation?

(A)

Create a Clear Visual Hierarchy

Organize and prioritize the contents of a page by using size, prominence, and content relationships. Let's look at these relationships more closely. The more important a headline is, the larger its font size should be. Big bold headlines help to grab the user's attention as they scan the Web page. The more important the headline or content, the higher up the page it should be placed. The most important or popular content should always be positioned prominently near the top of the page, so users can view it without having to scroll too far. Group similar content types by displaying the content in a similar visual style, or in a clearly defined area.

(B)

Create a Clear Visual Hierarchy

Organize and prioritize the contents of a page by using size, prominence, and content relationships.

Let's look at these relationships more closely:

- **Size.** The more important a headline is, the larger its font size should be. Big bold headlines help to grab the user's attention as they scan the Web page.
- **Prominence.** The more important the headline or content, the higher up the page it should be placed. The most important or popular content should always be positioned prominently near the top of the page, so users can view it without having to scroll too far.
- **Content Relationships.** Group similar content types by displaying the content in a similar visual style, or in a clearly defined area.

FIGURE 3.11

Find the advice about prominence in each of these displays. Prose text format (A) makes people read everything. Visual hierarchy (B) lets people ignore information irrelevant to their goals.

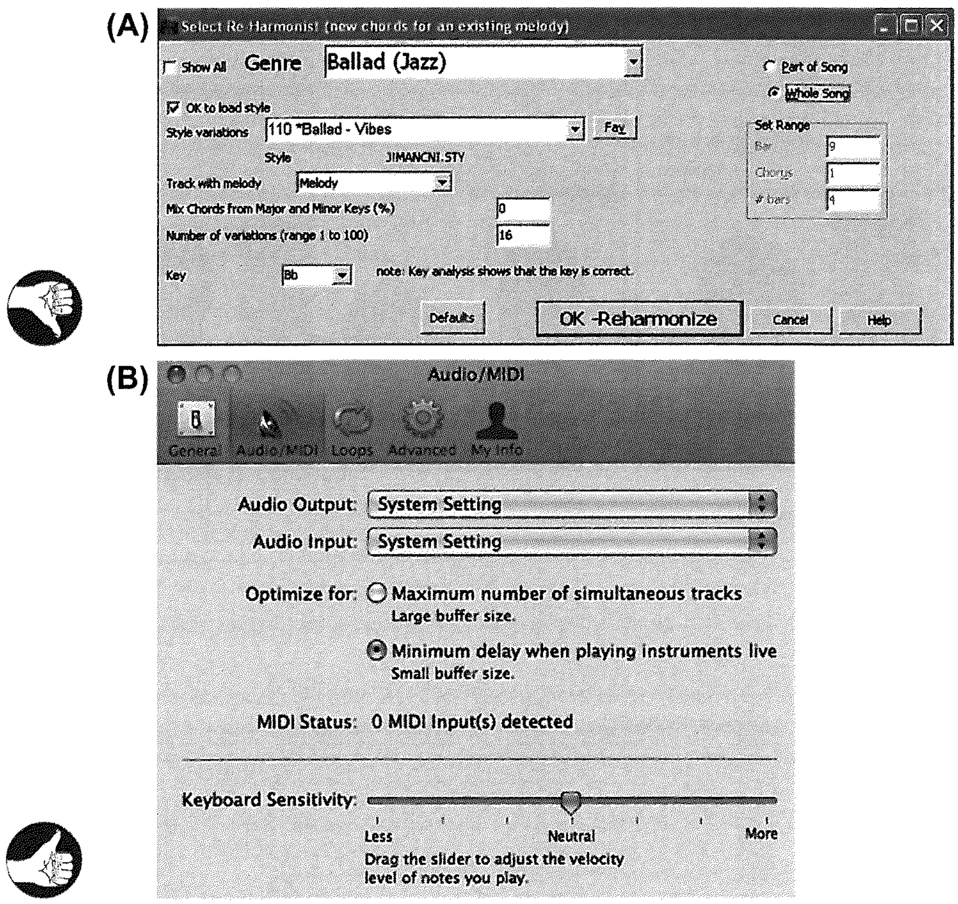
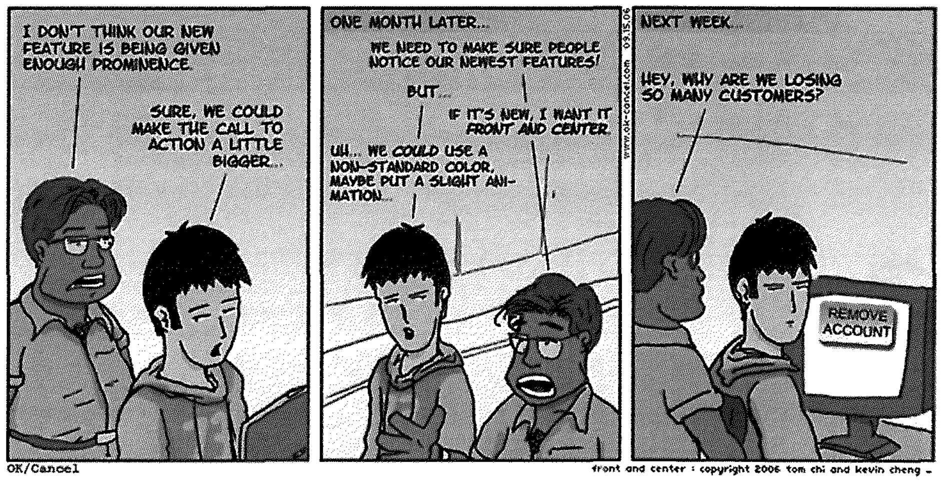


FIGURE 3.12

Visual hierarchy in interactive control panels and forms lets users find settings quickly: (A) Band-in-a-Box (bad) and (B) GarageBand (good).

The examples in Figure 3.11 show the value of visual hierarchy in a textual, read-only information display. Visual hierarchy is equally important in interactive control panels and forms—perhaps even more so. Compare dialog boxes from two different music software products (see Fig. 3.12). The Reharmonize dialog box of Band-in-a-Box has poor visual hierarchy, making it hard for users to find things quickly. In contrast, GarageBand's Audio/MIDI control panel has good visual hierarchy, so users can quickly find the settings they are interested in.



Used by permission, www.OK/Cancel.com.

CHAPTER

Our Peripheral Vision is Poor

5

Chapter 4 explained that the human visual system differs from a digital camera in the way it detects and processes color. Our visual system also differs from a camera in its resolution. On a digital camera's photo sensor, photoreceptive elements are spread uniformly in a tight matrix, so the spatial resolution is constant across the entire image frame. The human visual system is not like that.

This chapter explains why

- Stationary items in muted colors presented in the periphery of people's visual field often will not be noticed.
- Motion in the periphery is usually noticed.

RESOLUTION OF THE FOVEA COMPARED TO THE PERIPHERY

The spatial resolution of the human visual field drops greatly from the center to the edges. There are three reasons for this:

- ***Pixel density.*** Each eye has 6 to 7 million retinal cone cells. They are packed much more tightly in the center of our visual field—a small region called the *fovea*—than they are at the edges of the retina (see Fig. 5.1). The fovea has about 158,000 cone cells in each square millimeter. The rest of the retina has only 9,000 cone cells per square millimeter.
- ***Data compression.*** Cone cells in the fovea connect 1:1 to the ganglial neuron cells that begin the processing and transmission of visual data, while elsewhere on the retina, multiple photoreceptor cells (cones and rods) connect to each ganglion cell. In technical terms, information from the visual periphery is compressed (with data loss) before transmission to the brain, while information from the fovea is not.

Designing with the Mind in Mind. http://dx.doi.org/10.1016/B978-0-12-407914-4.00005-1

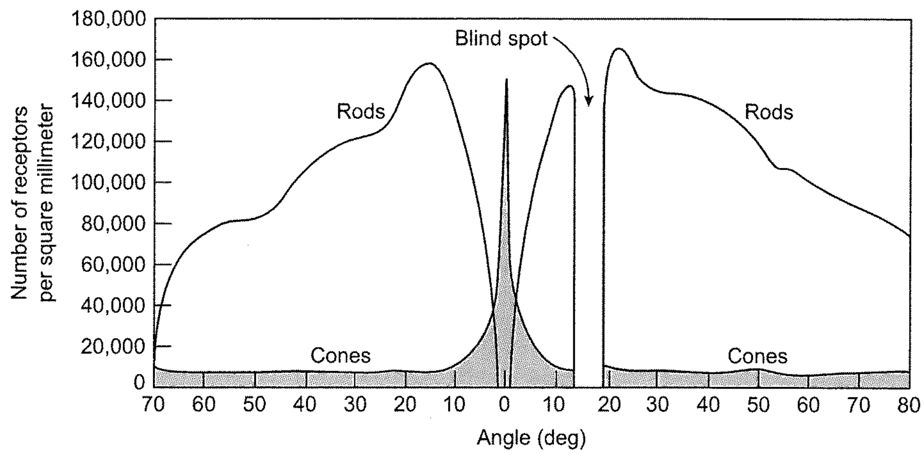


FIGURE 5.1

Distribution of photoreceptor cells (cones and rods) across the retina. *From Lindsay and Norman (1972).*

- ***Processing resources.*** The fovea is only about 1% of the retina, but the brain's visual cortex devotes about 50% of its area to input from the fovea. The other half of the visual cortex processes data from the remaining 99% of the retina.

The result is that our vision has much, much greater resolution in the center of our visual field than elsewhere (Lindsay and Norman, 1972; Waloszek, 2005). Said in developer jargon: in the center 1% of your visual field (i.e., the fovea), you have a high-resolution TIFF, and everywhere else, you have only a low-resolution JPEG. That is *nothing like* a digital camera.

To visualize how small the fovea is compared to your entire visual field, hold your arm straight out and look at your thumb. Your thumbnail, viewed at arm's length, corresponds approximately to the fovea (Ware, 2008). While you have your eyes focused on the thumbnail, everything else in your visual field falls outside of your fovea on your retina.

In the fovea, people with normal vision have very high resolution: they can resolve several thousand dots within that region—better resolution than many of today's pocket digital cameras. Just outside of the fovea, the resolution is already down to a few dozen dots per inch viewed at arm's length. At the edges of our vision, the "pixels" of our visual system are as large as a melon (or human head) at arm's length (see Fig. 5.2).

Even though our eyes have more rods than cones—125 million versus 6-7 million—peripheral vision has much lower resolution than foveal vision. This is because while most of our cone cells are densely packed in the fovea (1% of the retina's area), the rods are spread out over the rest of the retina (99% of the retina's area). In people with normal vision, peripheral vision is about 20/200, which in the United States is considered

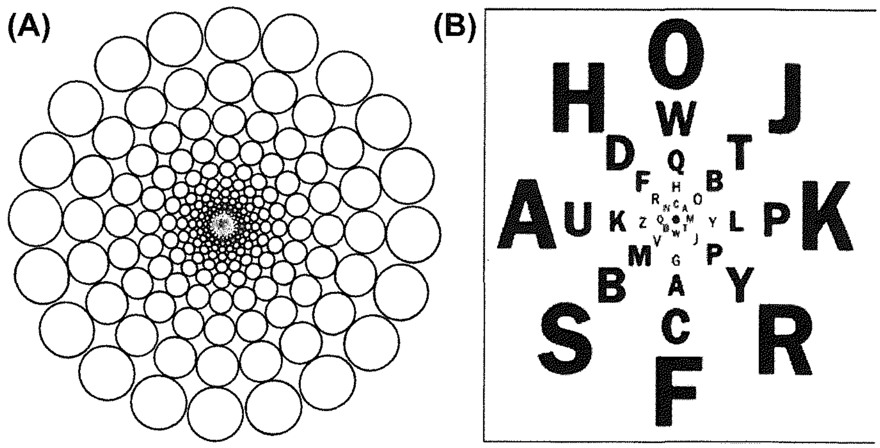


FIGURE 5.2

The resolution of our visual field is high in the center but much lower at the edges. *Right image from* Vision Research, *Vol. 14 (1974), Elsevier.*

legally blind. Think about that: in the periphery of your visual field, you are legally blind. Here is how brain researcher David Eagleman (2012; page 23) describes it:

> The resolution in your peripheral vision is roughly equivalent to looking through a frosted shower door, and yet you enjoy the illusion of seeing the periphery clearly. ... Wherever you cast your eyes appears to be in sharp focus, and therefore you assume the whole visual world is in focus.

If our peripheral vision has such low resolution, one might wonder why we don't see the world in a kind of tunnel vision where everything is out of focus except what we are directly looking at now. Instead, we seem to see our surroundings sharply and clearly all around us. We experience this illusion because our eyes move rapidly and constantly about three times per second even when we don't realize it, focusing our fovea on selected pieces of our environment. Our brain fills in the rest in a gross, impressionistic way based on what we know and expect.[1] Our brain does not have to maintain a high-resolution mental model of our environment because it can order the eyes to sample and resample details in the environment as needed (Clark, 1998).

For example, as you read this page, your eyes dart around, scanning and reading. No matter where on the page your eyes are focused, you have the impression of viewing a complete page of text, because, of course, you are.

[1] Our brains also fill in perceptual gaps that occur during rapid (saccadic) eye movements, when vision is suppressed (see Chapter 14).

+ @

FIGURE 5.3

To "see" the retinal gap, cover your left eye, hold this book near your face, and focus your right eye on the **+**. Move the book slowly away from you, staying focused on the **+**. The @ will disappear at some point.

But now, imagine that you are viewing this page on a computer screen, and the computer is tracking your eye movements and knows where your fovea is on the page. Imagine that wherever you look, the right text for that spot on the page is shown clearly in the small area corresponding to your fovea, but everywhere else on the page, the computer shows random, meaningless text. As your fovea flits around the page, the computer quickly updates each area where your fovea stops to show the correct text there, while the last position of your fovea returns to textual noise. Amazingly, experiments have shown that people *rarely notice* this: not only can they read, they believe that they are viewing a full page of meaningful text (Clark, 1998). However, it does slow people's reading, even if they don't realize it (Larson, 2004).

The fact that retinal cone cells are distributed tightly in and near the fovea, and sparsely in the periphery of the retina, affects not only spatial resolution but color resolution. We can discriminate colors better in the center of our visual field than at the edges.

Another interesting fact about our visual field is that it has a gap—a small area (blind spot) in which we see nothing. The gap corresponds to the spot on our retina where the optic nerve and blood vessels exit the back of the eye (see Fig. 5.1). There are no retinal rod or cone cells at that spot, so when the image of an object in our visual field happens to fall on that part of the retina, we don't see it. We usually don't notice this hole in our vision because our brain fills it in with the surrounding content, like a graphic artist using Photoshop to fill in a blemish on a photograph by copying nearby background pixels.

People sometimes experience the blind spot when they gaze at stars. As you look at one star, a nearby star may disappear briefly into the blind spot until you shift your gaze. You can also observe the gap by trying the exercise in Figure 5.3. Some people have other gaps resulting from imperfections on the retina, retinal damage, or brain strokes that affect the visual cortex,[2] but the optic nerve gap is an imperfection everyone shares.

IS THE VISUAL PERIPHERY GOOD FOR ANYTHING?

It seems that the fovea is better than the periphery at just about everything. One might wonder why we have peripheral vision. What is it good for? Our peripheral vision serves three important functions: it guides fovea, detects motion, and lets us see better in the dark.

[2] See VisionSimulations.com.

Function 1: Guides fovea

First, peripheral vision provides low-resolution cues to guide our eye movements so that our fovea visits all the interesting and crucial parts of our visual field. Our eyes don't scan our environment randomly. They move so as to focus our fovea on important things, the most important ones (usually) first. The fuzzy cues on the outskirts of our visual field provide the data that helps our brain plan where to move our eyes, and in what order.

For example, when we scan a medicine label for a "use by" date, a fuzzy blob in the periphery with the vague form of a date is enough to cause an eye movement that lands the fovea there to allow us to check it. If we are browsing a produce market looking for strawberries, a blurry reddish patch at the edge of our visual field draws our eyes and our attention, even though sometimes it may turn out to be radishes instead of strawberries. If we hear an animal growl nearby, a fuzzy animal-like shape in the corner of our eye will be enough to zip our eyes in that direction, especially if the shape is moving toward us (see Fig. 5.4).

How peripheral vision guides and augments central, foveal vision is discussed more in the "Visual Search Is Linear Unless Targets 'Pop' in the Periphery" section later in this chapter.

Function 2: Detects motion

A related guiding function of peripheral vision is that it is good at detecting motion. Anything that moves in our visual periphery, even slightly, is likely to draw our attention—and hence our fovea—toward it. The reason for this phenomenon is that our ancestors—including prehuman ones—were selected for their ability to spot food and avoid predators. As a result, even though we can move our eyes under conscious, intentional control, some of the mechanisms that control where they look are preconscious, involuntary, and very fast.

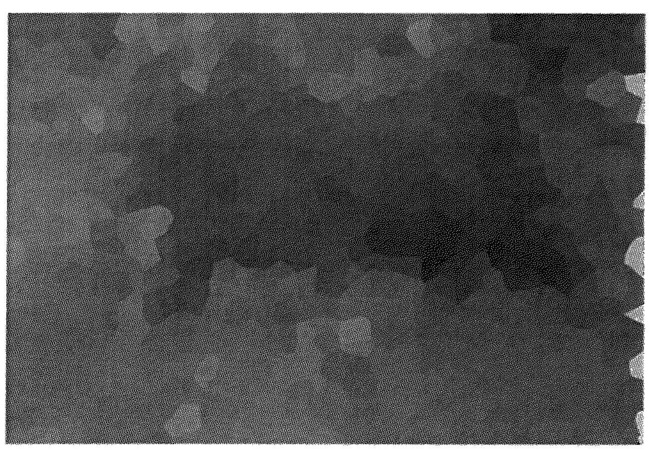


FIGURE 5.4

A moving shape at the edge of our vision draws our eye: it could be food, or it might consider us food.

What if we have no reason to expect that there might be anything interesting in a certain spot in the periphery,[3] and nothing in that spot attracts our attention? Our eyes may never move our fovea to that spot, so we may never see what is there.

Function 3: Lets us see better in the dark

A third function of peripheral vision is to allow us to see in low-light conditions—for example, on starlit nights, in caves, around campfires, etc. These were conditions under which vision evolved, and in which people—like the animals that preceded them on Earth—spent much of their time until the invention of the electric light bulb in the 1800s.

Just as the rods are overloaded in well-lighted conditions (see Chapter 5), the cones don't function very well in low light, so our rods take over. Low-light, rods-only vision is called *scotopic vision*. An interesting fact is that because there are no rods in the fovea, you can see objects better in low-light conditions (e.g., faint stars) if you don't look directly at them.

EXAMPLES FROM COMPUTER USER INTERFACES

The low acuity of our peripheral vision explains why software and website users fail to notice error messages in some applications and websites. When someone clicks a button or a link, that is usually where his or her fovea is positioned. Everything on the screen that is not within 1–2 centimeters of the click location (assuming normal computer viewing distance) is in peripheral vision, where resolution is low. If, after the click, an error message appears in the periphery, it should not be surprising that the person might not notice it.

For example, at InformaWorld.com, the online publications website of Informa Healthcare, if a user enters an incorrect username or password and clicks "Sign In," an error message appears in a "message bar" far away from where the user's eyes are most likely focused (see Fig. 5.5). The red word "Error" might appear in the user's peripheral vision as a small reddish blob, which would help draw the eyes in that direction. However, the red blob could fall into a gap in the viewer's visual field, and so not be noticed at all.

Consider the sequence of events from a user's point of view. The user enters a username and password and then clicks "Sign In." The page redisplays with blank fields. The user thinks "Huh? I gave it my login information and hit 'Sign In,' didn't I? Did I hit the wrong button?" The user reenters the username and password, and clicks "Sign In" again. The page redisplays with empty fields again. Now the user is really confused. The user sighs (or curses), sits back in his chair and lets his eyes scan the screen. Suddenly noticing the error message, the user says "A-ha! Has that error message been there all along?"

[3] See Chapter 1 on how expectations bias our perceptions.

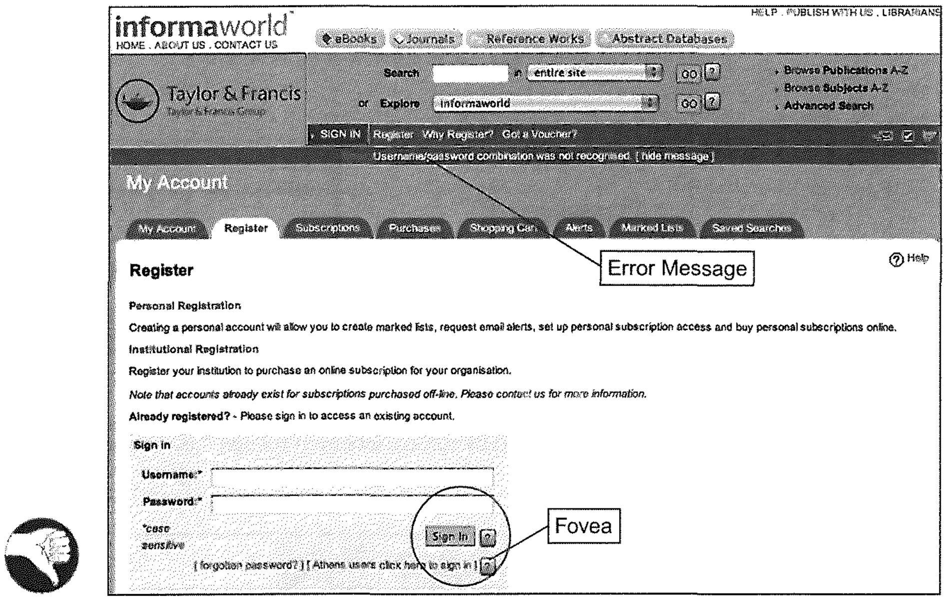


FIGURE 5.5

This error message for a faulty sign-in appears in peripheral vision, where it will probably be missed.

Even when an error message is placed nearer to the center of the viewer's visual field than in the preceding example, other factors can diminish its visibility. For example, until recently the website of Airborne.com signaled a login failure by displaying an error message in red just above the Login ID field (see Fig. 5.6). This error message is entirely in red and fairly near the "Login" button where the user's eyes are probably focused. Nonetheless, some users would not notice this error message when it first appeared. Can you think of any reasons people might not initially see this error message?

One reason is that even though the error message is much closer to where users will be looking when they click the "Login" button, it is still in the periphery, not in the fovea. The fovea is small: just a centimeter or two on a computer screen, assuming the user is the usual distance from the screen.

A second reason is that the error message is not the only thing near the top of the page that is red. The page title is also red. Resolution in the periphery is low, so when the error message appears, the user's visual system may not register any change: there was something red up there before, and there still is (see Fig. 5.7).

If the page title were black or any other color besides red, the red error message would be more likely to be noticed, even though it appears in the periphery of the users' visual field.

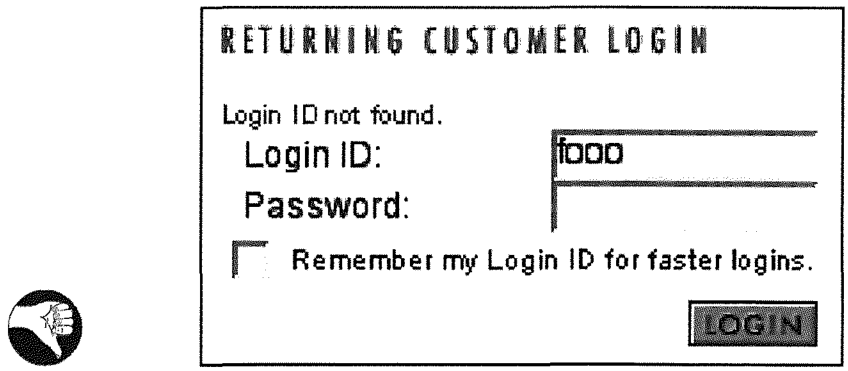


FIGURE 5.6

This error message for a faulty login is missed by some users even though it is not far from the "Login" button.

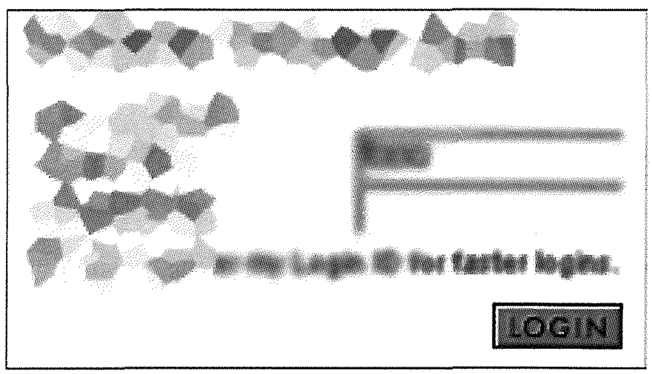


FIGURE 5.7

Simulation of a user's visual field while the fovea is fixed on the "Login" button.

COMMON METHODS OF MAKING MESSAGES VISIBLE

There are several common and well-known methods of ensuring that an error message will be seen:

- ***Put it where users are looking.*** People focus in predictable places when interacting with graphical user interfaces (GUIs). In Western societies, people tend to traverse forms and control panels from upper left to lower right. While moving the screen pointer, people usually look either at where it is or where they are moving it to. When people click a button or link, they can usually be assumed to be looking directly at it, at least for a few moments afterward. Designers can use this predictability to position error messages near where they expect users to be looking.

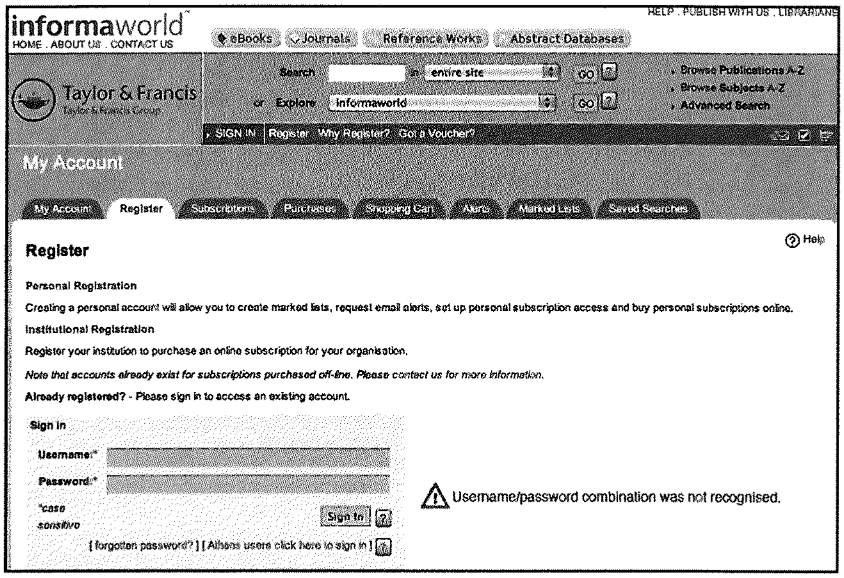


FIGURE 5.8

This error message for faulty sign-in is displayed more prominently, near where users will be looking.

- ***Mark the error.*** Somehow mark the error prominently to indicate clearly that something is wrong. Often this can be done by simply placing the error message near what it refers to, unless that would place the message too far from where users are likely to be looking.
- ***Use an error symbol.*** Make errors or error messages more visible by marking them with an error symbol, such as .
- ***Reserve red for errors.*** By convention, in interactive computer systems the color red connotes *alert*, *danger*, *problem*, *error*, etc. Using red for any other information on a computer display invites misinterpretation. But suppose you are designing a website for Stanford University, which has red as its school color. Or suppose you are designing for a Chinese market, where red is considered an auspicious, positive color. What do you do? Use another color for errors, mark them with error symbols, or use stronger methods (see the next section).

An improved version of the InformaWorld sign-in error screen uses several of these techniques (see Fig. 5.8).

At America Online's website, the form for registering for a new email account follows the guidelines pretty well (see Fig. 5.9). Data fields with errors are marked with red error symbols. Error messages are displayed in red and are near the error. Furthermore, most of the error messages appear as soon as an erroneous entry is made, when

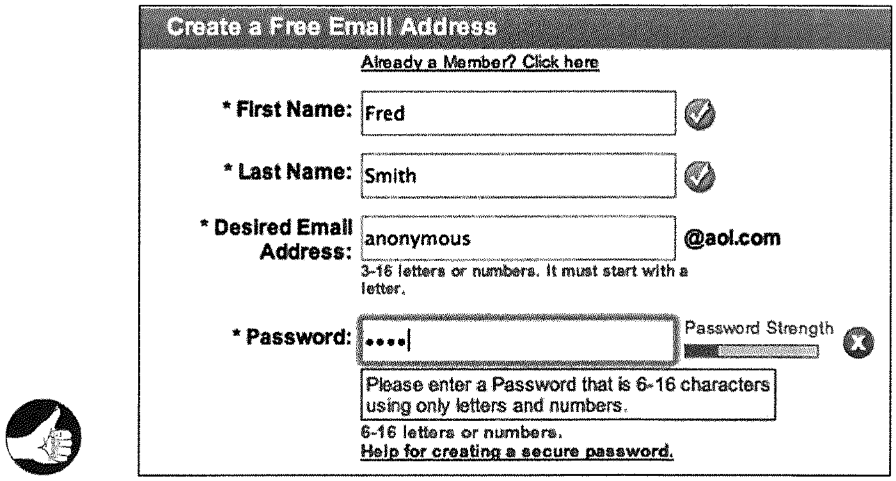


FIGURE 5.9

New member registration at AOL.com displays error messages prominently, near each error.

the user is still focused on that part of the form, rather than only after the user submits the form. It is unlikely that AOL users will miss seeing these error messages.

HEAVY ARTILLERY FOR MAKING USERS NOTICE MESSAGES

If the common, conventional methods of making users notice messages are not enough, three stronger methods are available to user-interface designers: pop-up message in error dialog box, use of sound (e.g., beep), and wiggle or blink briefly. However, these methods, while very effective, have significant negative effects, so they should be used sparingly and with great care.

Method 1: Pop-up message in error dialog box

Displaying an error message in a dialog box sticks it right in the user's face, making it hard to miss. Error dialog boxes interrupt the user's work and demand immediate attention. That is good if the error message signals a critical condition, but it can annoy people if such an approach is used for a minor message, such as confirming the execution of a user-requested action.

The annoyance of pop-up messages rises with the degree of modality. *Nonmodal* pop-ups allow users to ignore them and continue working. *Application-modal* pop-ups block any further work in the application that displayed the error, but allow users to interact with other software on their computer. *System-modal* pop-ups block any user action until the dialog has been dismissed.

Application-modal pop-ups should be used sparingly—for example, only when application data may be lost if the user doesn't attend to the error. System-modal

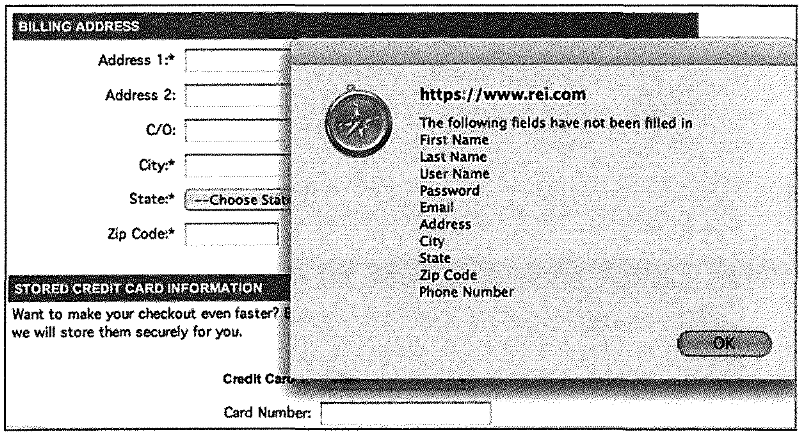


FIGURE 5.10

REI's pop-up dialog box signals required data that was omitted. It is hard to miss, but perhaps overkill.

pop-ups should be used extremely rarely—basically only when the system is about to crash and take hours of work with it, or if people will die if the user misses the error message.

On the Web, an additional reason to avoid pop-up error dialog boxes is that some people set their browsers to block *all* pop-up windows. If your website relies on pop-up error messages, some users may never see them.

REI.com has an example of a pop-up dialog being used to display an error message. The message is displayed when someone who is registering as a new customer omits required fields in the form (see Fig. 5.10). Is this an appropriate use of a pop-up dialog? AOL.com (see Fig. 5.9) shows that missing data errors can be signaled quite well without pop-up dialogs, so REI.com's use of them seems a bit heavy-handed.

Examples of more appropriate use of error dialog boxes come from Microsoft Excel (see Fig. 5.11A) and Adobe InDesign (see Fig. 5.11B). In both cases, loss of data is at stake.

Method 2: Use sound (e.g., beep)

When a computer beeps, that tells its user something has happened that requires attention. The person's eyes reflexively begin scanning the screen for whatever caused the beep. This can allow the user to notice an error message that is someplace other than where the user was just looking, such as in a standard error message box on the display. That is the value of beeping.

However, imagine many people in a cubicle work environment or a classroom, all using an application that signals all errors and warnings by beeping. Such a workplace would be very annoying, to say the least. Worse, people wouldn't be able to tell whether their own computer or someone else's was beeping.

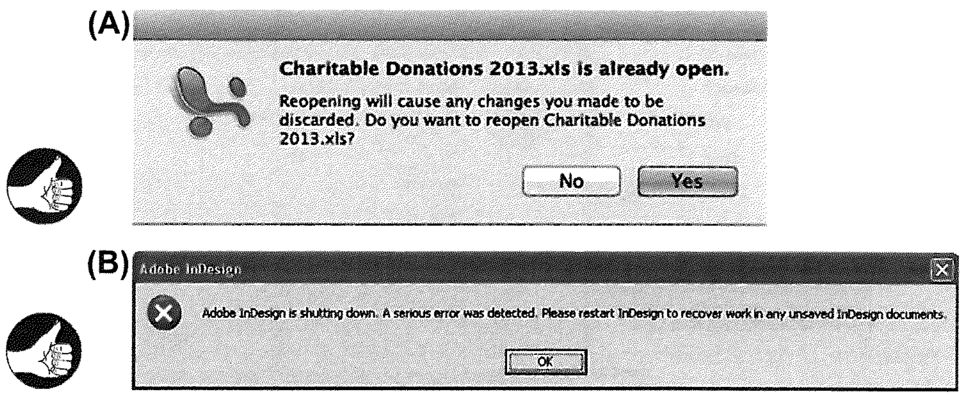


FIGURE 5.11

Appropriate pop-up error dialogs: (A) Microsoft Excel and (B) Adobe InDesign.

The opposite situation is noisy work environments (e.g., factories or computer server rooms), where auditory signals emitted by an application might be masked by ambient noise. Even in non-noisy environments, some computer users simply prefer quiet, and mute the sound on their computers or turn it way down.

For these reasons, signaling errors and other conditions with sound are remedies that can be used only in very special, controlled situations.

Computer games often use sound to signal events and conditions. In games, sound isn't annoying; it is expected. Its use in games is widespread, even in game arcades, where dozens of machines are all banging, roaring, buzzing, clanging, beeping, and playing music at once. (Well, it is annoying to parents who have to go into the arcades and endure all the screeching and booming to retrieve their kids, but the games aren't designed for parents.)

Method 3: Wiggle or blink briefly

As described earlier in this chapter, our peripheral vision is good at detecting motion, and motion in the periphery causes reflexive eye movements that bring the motion into the fovea. User-interface designers can make use of this by wiggling or flashing messages briefly when they want to ensure that users see them. It doesn't take much motion to trigger eye movement toward the motion. Just a tiny bit of motion is enough to make a viewer's eyes zip over in that direction. Millions of years of evolution have had quite an effect.

As an example of using motion to attract users' eye attention, Apple's iCloud online service briefly shakes the entire dialog box horizontally when a user enters an invalid username or password (see Fig. 5.12). In addition to clearly indicating "No" (like a person shaking his head), this attracts the users' eyeballs, guaranteed. (Because, after all, the motion in the corner of your eye might be a leopard.)

The most common use of blinking in computer user interfaces (other than advertisements) is in menu bars. When an action (e.g., Edit or Copy) is selected from a

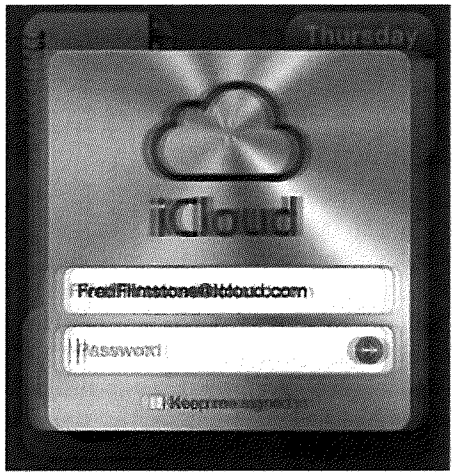


FIGURE 5.12

Apple's iCloud shakes the dialog box briefly on login errors to attract a user's fovea toward it.

menu, it usually blinks once before the menu closes to confirm that the system "got" the command—that is, that the user didn't miss the menu item. This use of blinking is very common. It is so quick that most computer users aren't even aware of it, but if menu items didn't blink once, we would have less confidence that we actually selected them.

Motion and blinking, like pop-up dialog boxes and beeping, must be used sparingly. Most experienced computer users consider wiggling, blinking objects on screen to be annoying. Most of us have learned to ignore displays that blink because many such displays are advertisements. Conversely, a few computer users have attentional impairments that make it difficult for them to ignore something that is blinking or wiggling.

Therefore, if wiggling or blinking is used, it should be brief—it should last about a quarter- to a half-second, no longer. Otherwise, it quickly goes from an unconscious attention-grabber to a conscious annoyance.

Use heavy-artillery methods sparingly to avoid habituating your users

There is one final reason to use the preceding heavy-artillery methods sparingly (i.e., only for critical messages): to avoid *habituating* your users. When pop-ups, sound, motion, and blinking are used too often to attract users' attention, a psychological phenomenon called *habituation* sets in (see Chapter 1). Our brain pays less and less attention to any stimulus that occurs frequently.

It is like the old fable of the boy who cried "Wolf!" too often: eventually, the villagers learned to ignore his cries, so when a wolf actually did come, his cries went unheeded. Overuse of strong attention-getting methods can cause important messages to be blocked by habituation.

VISUAL SEARCH IS LINEAR UNLESS TARGETS "POP" IN THE PERIPHERY

As explained earlier, one function of peripheral vision is to drive our eyes to focus the fovea on important things—things we are seeking or that might be a threat. Objects moving in our peripheral vision fairly reliably "yank" our eyes in that direction.

When we are looking for an object, our entire visual system, including the periphery, primes itself to detect that object. In fact, the periphery is a *crucial* component in visual search, despite its low spatial and color resolution. However, just how helpful the periphery is in aiding visual search depends strongly on what we are looking for.

Look quickly at Figure 5.13 and find the Z.

To find the Z, you had to scan carefully through the characters until your fovea landed on it. In the lingo of vision researchers, the time to find the Z is *linear*: it depends approximately linearly on the number of distracting characters and the position of the Z among them.

Now look quickly at Figure 5.14 and find the bold character.

That was much easier (i.e., faster), wasn't it? You didn't have to scan your fovea carefully through the distracting characters. Your periphery quickly detected the boldness and determined its location, and because that is what you were seeking,

L Q R B T J P L F B M R W S
F R N Q S P D C H K U T
G T H U J L U 9 J V Y I A
E X C F T Y N H T D O L L8
G V N G R Y J G Z S T 6 S
3 L C T V B H U S E M U K
W Q E L F G H U Y I K D 9

FIGURE 5.13

Finding the Z requires scanning carefully through the characters.

G T H U J L U 9 J V Y I A
L Q R B T J P L F B M R W S
3 L C T V B H U S E M U K
F R N Q S P D C H K U T
W Q E L F G H B Y I K D 9
G V N **G** R Y J G Z S T 6 S
E X C F T Y N H T D O L L 8

FIGURE 5.14

Finding the bold letter does **not** require scanning through everything.

L Q R B T J P L F BM R W S
F R N Q S P D C H K U T
G T H U J L U 9 J V Y I A
E X C F T Y N H T D O L L8
3 L C T V B H U S E M U K
G V N G R Y J G Z S T 6 S
W Q E L F G H U Y I K D 9

FIGURE 5.15

Counting L's is hard; character shape doesn't "pop" among characters.

W Q E L F G H U Y I K D 9
F R N Q S P D C H K U T
3 L C T V B H U S E M U K
G T H U J L U 9 J V Y I A
L Q R B T J P L F BM R W S
E X C F T Y N H T D O L L8
G V N G R Y J G Z S T 6 S

FIGURE 5.16

Counting blue characters is easy because color "pops."

your visual system moved your fovea there. Your periphery could not determine exactly *what* was bold—that is beyond its resolution and abilities—but it did locate the boldness. In vision-researcher lingo, the periphery was primed to look for boldness in parallel over its entire area, and boldness is a distinctive feature of the target, so searching for a bold target is nonlinear. In designer lingo, we simply say that boldness "pops out" ("pops" for short) in the periphery, assuming that only the target is bold.

Color "pops" even more strongly. Compare counting the L's in Figure 5.15 with counting the blue characters in Figure 5.16.

What else makes things "pop" in the periphery? As described earlier, the periphery easily detects motion, so motion "pops." Generalizing from *boldness*, we also can say that font *weight* "pops," because if all but one of the characters on a display were bold, the *non*bold character would stand out. Basically, a visual target will pop out in your periphery if it differs from surrounding objects in features the periphery can detect. The more distinctive features of the target, the more it "pops," assuming the periphery can detect those features.

Using peripheral "pop" in design

Designers use peripheral "pop" to focus the attention of a product's users, as well as to allow users to find information faster. Chapter 3 described how visual hierarchy—titles, headings, boldness, bullets, and indenting—can make it easier for users to spot

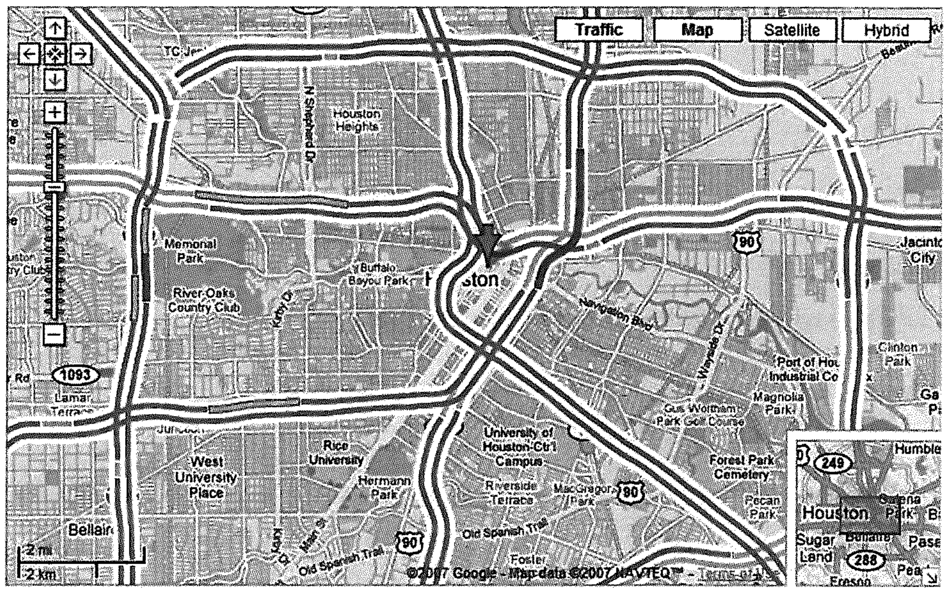


FIGURE 5.17

Google Maps uses color to show traffic conditions. Red indicates traffic jams.

and extract from text the information they need. Glance back at Figure 3.11 in Chapter 3 and see how the headings and bullets make the topics and subtopics "pop" so readers can go right to them.

Many interactive systems use color to indicate status, usually reserving red for problems. Online maps and some vehicle GPS devices mark traffic jams with red so they stand out (see Fig. 5.17). Systems for controlling air traffic mark potential collisions in red (see Fig. 5.18). Applications for monitoring servers and networks use color to show the health status of assets or groups of them (see Fig. 5.19).

These are all uses of peripheral "pop" to make important information stand out and visual search nonlinear.

When there are many possible targets

Sometimes in displays of many items, *any* of them could be what the user wants. Examples include command menus (see Fig. 5.20A) and object pallets (see Fig. 5.20B). Let's assume that the application cannot anticipate which item or items a user is likely to want, and highlight those. That is a fair assumption for today's applications.[4] Are users doomed to have to search linearly through such displays for the item they want?

That depends. Designers can try to make each item so distinctive that when a specific one is the user's target, the user's peripheral vision will be able to spot it among

[4] But in the not-too-distant future it might not be.

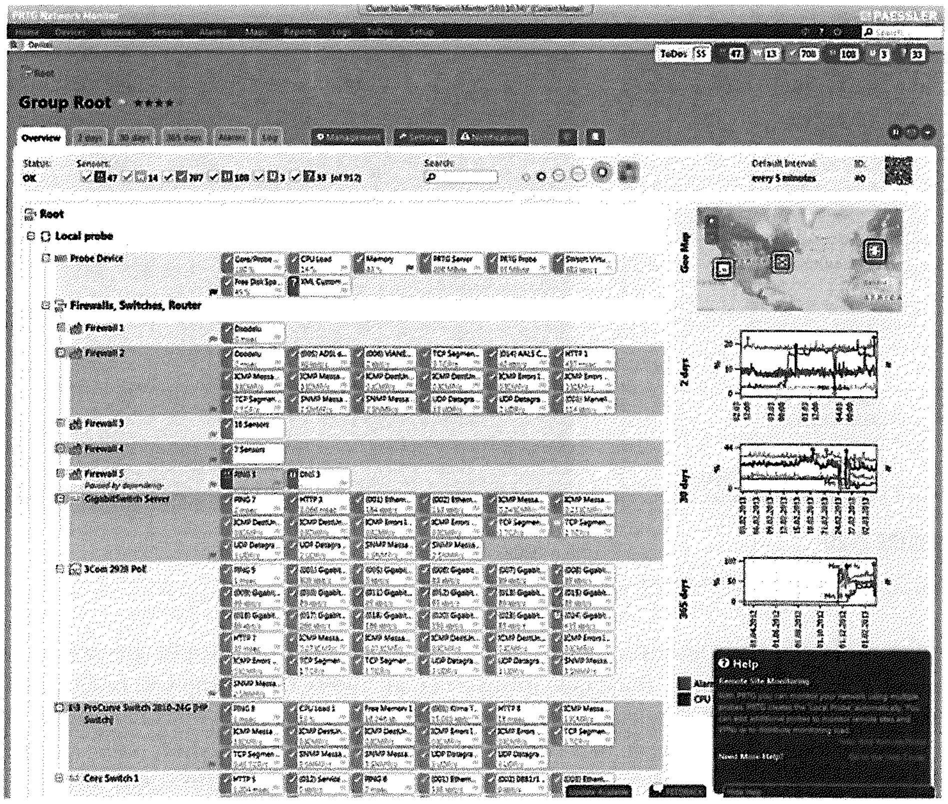


FIGURE 5.18

Air traffic control systems often use red to make potential collisions stand out.

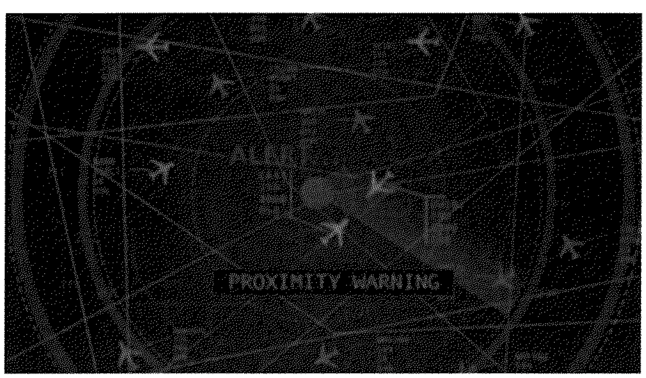


FIGURE 5.19

Paessler's monitoring tool uses color to show the health of network components.

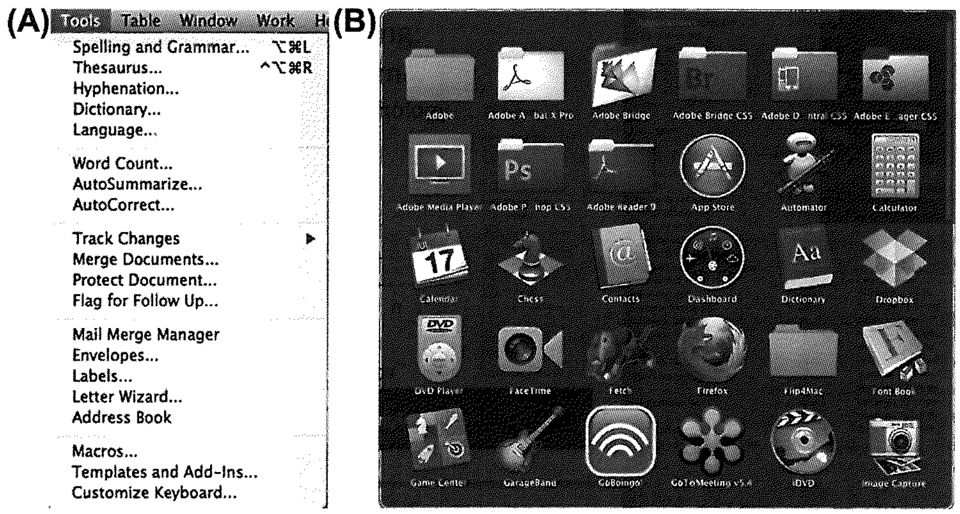


FIGURE 5.20

(A) Microsoft Word Tools menu, and (B) MacOS application pallet.

all the other items. Designing distinctive sets of icons is hard—especially when the set is large—but it can be done (see Johnson et. al, 1989). Designing sets of icons that are so distinctive that they can be distinguished in peripheral vision is *very* hard, but not impossible. For example, if a user goes to the Mac OS application pallet to open his or her calendar, a white rectangular blob in the periphery with something black in the middle is more likely to attract the user's eye than a blue circular blob (see Fig. 5.20B). The trick is not to get too fancy and detailed with the icons—give each one a distinctive color and gross shape.

On the other hand, if the potential targets are all words, as in command menus (see Fig. 20A), visual distinctiveness is not an option. In textual menus and lists, visual search *will* be linear, at least at first. With practice, users learn the positions of frequently used items in menus, lists, and pallets, so searching for particular items is no longer linear.

That is why applications should *never* move items around in menus, lists, or pallets. Doing that prevents users from learning item positions, thereby dooming them to search linearly forever. Therefore, "dynamic menus" is considered a major user-interface design blooper (Johnson, 2007).

CHAPTER

Our Attention is Limited; Our Memory is Imperfect

7

Just as the human visual system has strengths and weaknesses, so do human attention and memory. This chapter describes some of those strengths and weaknesses as background for understanding how we can design interactive systems to support and augment attention and memory rather than burdening or confusing them. We will start with an overview of how memory works, and how it is related to attention.

SHORT- VERSUS LONG-TERM MEMORY

Psychologists historically have distinguished *short-term* memory from *long-term* memory. Short-term memory covers situations in which information is retained for intervals ranging from a fraction of a second to a few minutes. Long-term memory covers situations in which information is retained over longer periods (e.g., hours, days, years, even lifetimes).

It is tempting to think of short- and long-term memory as separate memory stores. Indeed, some theories of memory have considered them separate. After all, in a digital computer, the short-term memory stores (central processing unit [CPU] data registers) are separate from the long-term memory stores (random access memory [RAM], hard disk, flash memory, CD-ROM, etc.). More direct evidence comes from findings that damage to certain parts of the human brain results in short-term memory deficits but not long-term ones, or vice versa. Finally, the speed with which information or plans can disappear from our immediate awareness contrasts sharply with the seeming permanence of our memory of important events in our lives, faces of significant people, activities we have practiced, and information we have studied. These phenomena led many researchers to theorize that short-term memory is a separate store in the brain where information is held temporarily after entering through our perceptual senses (e.g., visual or auditory), or after being retrieved from long-term memory (see Fig. 7.1).

Designing with the Mind in Mind. http://dx.doi.org/10.1016/B978-0-12-407914-4.00007-5

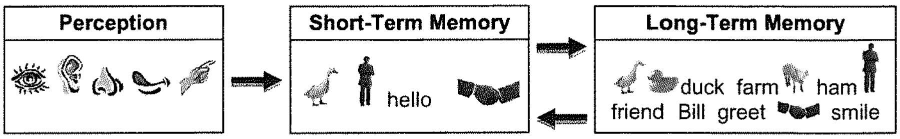


FIGURE 7.1

Traditional (antiquated) view of short-term versus long-term memory.

A MODERN VIEW OF MEMORY

Recent research on memory and brain function indicates that short- and long-term memory are functions of a single memory system—one that is more closely linked with perception than previously thought (Jonides et al., 2008).

Long-term memory

Perceptions enter through the visual, auditory, olfactory, gustatory, or tactile sensory systems and trigger responses starting in areas of the brain dedicated to each sense (e.g., visual cortex, auditory cortex), then spread into other areas of the brain that are *not* specific to any particular sensory modality. The sensory modality-specific areas of the brain detect only simple features of the data, such as a dark-light edge, diagonal line, high-pitched tone, sour taste, red color, or rightward motion. Downstream areas of the brain combine low-level features to detect higher-level features of the input, such as animal, the word "duck," Uncle Kevin, minor key, threat, or fairness.

As described in Chapter 1, the set of neurons activated by a perceived stimulus depends on both the features and context of the stimulus. The context is as important as the features of the stimulus in determining what neural patterns are activated. For example, a dog barking near you when you are walking in your neighborhood activates a different pattern of neural activity in your brain than the same sound heard when you are safely inside your car. The more similar two perceptual stimuli are—that is, the more features and contextual elements they share—the more overlap there is between the sets of neurons that fire in response to them.

The initial strength of a perception depends on how much it is amplified or dampened by other brain activity. All perceptions create some kind of trace, but some are so weak that they can be considered as not registered: the pattern was activated once but never again.

Memory formation consists of changes in the neurons involved in a neural activity pattern, which make the pattern easier to reactivate in the future.[1] Some such changes result from chemicals released near neural endings that boost or inhibit their sensitivity to stimulation. These changes last only until the chemicals dissipate

[1] There is evidence that the long-term neural changes associated with learning occur mainly during sleep, suggesting that separating learning sessions by periods of sleep may facilitate learning (Stafford and Webb, 2005).

or are neutralized by other chemicals. More permanent changes occur when neurons grow and branch, forming new connections with others.

Activating a memory consists of reactivating the same pattern of neural activity that occurred when the memory was formed. Somehow the brain distinguishes initial activations of neural patterns from *re*activations—perhaps by measuring the relative ease with which the pattern was reactivated. New perceptions very similar to the original ones reactivate the same patterns of neurons, resulting in *recognition* if the reactivated perception reaches awareness. In the absence of a similar perception, stimulation from activity in other parts of the brain can also reactivate a pattern of neural activity, which if it reaches awareness results in *recall*.

The more often a neural memory pattern is reactivated, the stronger it becomes—that is, the easier it is to reactivate—which in turn means that the perception it corresponds to is easier to recognize and recall. Neural memory patterns can also be strengthened or weakened by excitatory or inhibitory signals from other parts of the brain.

A particular memory is not located in any specific spot in the brain. The neural activity pattern comprising a memory involves a network of millions of neurons extending over a wide area. Activity patterns for different memories overlap, depending on which features they share. Removing, damaging, or inhibiting neurons in a particular part of the brain typically does not completely wipe out memories that involve those neurons, but rather just reduces the detail or accuracy of the memory by deleting features.[2] However, some areas in a neural activity pattern may be critical pathways, so that removing, damaging, or inhibiting them may prevent most of the pattern from activating, thereby effectively eliminating the corresponding memory.

For example, researchers have long known that the hippocampus, twin seahorse-shaped neural clusters near the base of the brain, plays an important role in storing long-term memories. The modern view is that the hippocampus is a controlling mechanism that directs neural rewiring so as to "burn" memories into the brain's wiring. The amygdala, two jellybean-shaped clusters on the frontal tips of the hippocampus, has a similar role, but it specializes in storing memories of emotionally intense, threatening situations (Eagleman, 2012).

Cognitive psychologists view human long-term memory as consisting of several distinct functions:

- ***Semantic*** long-term memory stores *facts and relationships.*
- ***Episodic*** long-term memory records *past events.*
- ***Procedural*** long-term memory remembers *action sequences.*

These distinctions, while important and interesting, are beyond the scope of this book.

[2]This is similar to the effect of cutting pieces out of a holographic image: it reduces the overall resolution of the image, rather than removing areas of it, as with an ordinary photograph.

Short-term memory

The processes just discussed are about long-term memory. What about short-term memory? What psychologists call short-term memory is actually a *combination* of phenomena involving perception, attention, and retrieval from long-term memory.

One component of short-term memory is perceptual. Each of our perceptual senses has its own very brief short-term "memory" that is the result of residual neural activity after a perceptual stimulus ceases, like a bell that rings briefly after it is struck. Until they fade away, these residual perceptions are available as possible input to our brain's attention and memory-storage mechanisms, which integrate input from our various perceptual systems, focus our awareness on some of that input, and store some of it in long-term memory. These sensory-specific residual perceptions together comprise a minor component of short-term memory. Here, we are only interested in them as potential inputs to working memory.

Also available as potential input to working memory are long-term memories reactivated through recognition or recall. As explained earlier, each long-term memory corresponds to a specific pattern of neural activity distributed across our brain. While activated, a memory pattern is a candidate for our attention and therefore potential input for working memory.

The human brain has multiple attention mechanisms, some voluntary and some involuntary. They focus our awareness on a very small subset of the perceptions and activated long-term memories while ignoring everything else. That tiny subset of all the available information from our perceptual systems and our long-term memories that we are aware of *right now* is the main component of our short-term memory, the part that cognitive scientists often call *working memory*. It integrates information from all of our sensory modalities and our long-term memory. Henceforth, we will restrict our discussion of short-term memory to working memory.

So what is working memory? First, here is what it is *not*: it is not a *store*—it is not a *place* in the brain where memories and perceptions *go* to be worked on. And it is nothing like accumulators or fast random-access memory in digital computers.

Instead, working memory is our combined focus of attention: everything that we are conscious of at a given time. More precisely, it is a few perceptions and long-term memories that are activated enough that we remain aware of them over a short period. Psychologists also view working memory as including an executive function—based mainly in the frontal cerebral cortex—that manipulates items we are attending to and, if needed, refreshes their activation so they remain in our awareness (Baddeley, 2012).

A useful—if oversimplified—analogy for memory is a huge, dark, musty warehouse. The warehouse is full of long-term memories, piled haphazardly (not stacked neatly), intermingled and tangled, and mostly covered with dust and cobwebs. Doors along the walls represent our perceptual senses: sight, hearing, smell, taste, touch. They open briefly to let perceptions in. As perceptions enter, they are briefly illuminated by light coming in from outside, but they quickly are pushed (by more entering perceptions) into the dark tangled piles of old memories.

In the ceiling of the warehouse are a small fixed number of searchlights, controlled by the attention mechanism's executive function (Baddeley, 2012). They swing around and focus on items in the memory piles, illuminating them for a while until they swing away to focus elsewhere. Sometimes one or two searchlights focus on new items after they enter through the doors. When a searchlight moves to focus on something new, whatever it had been focusing on is plunged into darkness.

The small fixed number of searchlights represents the limited capacity of working memory. What is illuminated by them (and briefly through the open doors) represents the contents of working memory: out of the vast warehouse's entire contents, the few items we are attending to at any moment. See Figure 7.2 for a visual.

The warehouse analogy is too simple and should not be taken too seriously. As Chapter 1 explained, our senses are *not* just passive doorways into our brains, through which our environment "pushes" perceptions. Rather, our brain actively and continually *seeks out* important events and features in our environment and "pulls" perceptions in as needed (Ware, 2008). Furthermore, the brain is buzzing with activity most of the time and its internal activity is only modulated—not determined—by sensory input (Eagleman, 2012). Also, as described earlier, memories are embodied as networks of neurons distributed around the brain, not as objects in a specific location. Finally, activating a memory in the brain can activate related ones; our warehouse-with-searchlights analogy doesn't represent that.

Nonetheless, the analogy—especially the part about the searchlights—illustrates that working memory is a *combination* of several *foci of attention*—the currently

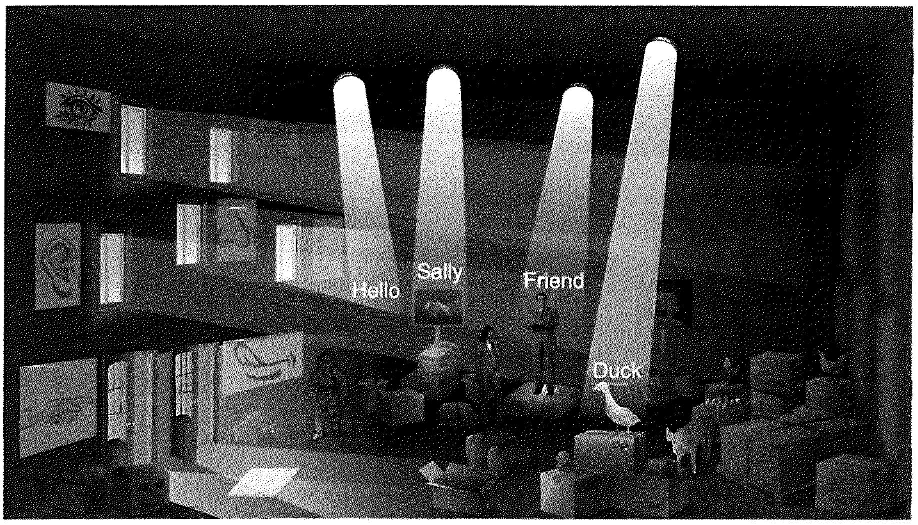


FIGURE 7.2

Modern view of memory: a dark warehouse full of stuff (long-term memory), with searchlights focused on a few items (short-term memory).

activated neural patterns of which we are aware—and that the capacity of working memory is extremely limited, and the content at any given moment is very volatile.

What about the earlier finding that damage to some parts of the brain causes short-term memory deficits, while other types of brain damage cause long-term memory deficits? The current interpretation is that some types of damage decrease or eliminate the brain's ability to focus attention on specific objects and events, while other types of damage harm the brain's ability to store or retrieve long-term memories.

CHARACTERISTICS OF ATTENTION AND WORKING MEMORY

As noted, working memory is equal to the focus of our attention. Whatever is in that focus is what we are conscious of at any moment. But what determines what we attend to and how much we can attend to at any given time?

Attention is highly focused and selective

Most of what is going on around you at this moment you are unaware of. Your perceptual system and brain sample very selectively from your surroundings, because they don't have the capacity to process everything.

Right now you are conscious of the last few words and ideas you've read, but probably not the color of the wall in front of you. But now that I've shifted your attention, you *are* conscious of the wall's color, and may have forgotten some of the ideas you read on the previous page.

Chapter 1 described how our perception is filtered and biased by our goals. For example, if you are looking for your friend in a crowded shopping mall, your visual system "primes" itself to notice people who look like your friend (including how he or she is dressed), and barely notice everything else. Simultaneously, your auditory system primes itself to notice voices that sound like your friend's voice, and even footsteps that sound like those of your friend. Human-shaped blobs in your peripheral vision and sounds localized by your auditory system that match your friend snap your eyes and head toward them. While you look, anyone looking or sounding similar to your friend attracts your attention, and you won't notice other people or events that would normally have interested you.

Besides focusing on objects and events related to our current goals, our attention is drawn to:

- ***Movement, especially movement near or toward us.*** For example, something jumps at you while you walk on a street, or something swings toward your head in a haunted house ride at an amusement park, or a car in an adjacent lane suddenly swerves toward your lane (see the discussion of the *flinch reflex* in Chapter 14).
- ***Threats.*** Anything that signals or portends danger to us or people in our care.

- ***Faces of other people.*** We are primed from birth to notice faces more than other objects in our environment.
- ***Sex and food.*** Even if we are happily married and well fed, these things attract our attention. Even the mere words probably quickly got your attention.

These things, along with our current goals, draw our attention involuntarily. We don't become aware of something in our environment and then orient ourselves toward it. It's the other way around: our perceptual system detects something attention-worthy and orients us toward it preconsciously, and only afterwards do we become aware of it.[3]

Capacity of attention (a.k.a. working memory)

The primary characteristics of working memory are its low capacity and volatility. But what is the capacity? In terms of the warehouse analogy presented earlier, what is *the small fixed number* of searchlights?

Many college-educated people have read about "the magical number seven, plus or minus two," proposed by cognitive psychologist George Miller in 1956 as the limit on the number of simultaneous unrelated items in human working memory (Miller, 1956).

Miller's characterization of the working memory limit naturally raises several questions:

- ***What are the items in working memory?*** They are current perceptions and retrieved memories. They are goals, numbers, words, names, sounds, images, odors—anything one can be aware of. In the brain, they are patterns of neural activity.
- ***Why must items be unrelated?*** Because if two items are related, they correspond to one big neural activity pattern—one set of features—and hence one item, not two.
- ***Why the fudge-factor of plus or minus two?*** Because researchers cannot measure with perfect accuracy how much people can keep track of, and because of differences between individuals in working memory capacity.

Later research in the 1960s and 1970s found Miller's estimate to be too high. In the experiments Miller considered, some of the items presented to people to remember could be "chunked" (i.e., considered related), making it appear that people's working memory was holding more items than it actually was. Furthermore, all the subjects in Miller's experiments were college students. Working memory capacity varies in the general population. When the experiments were revised to disallow unintended chunking and include noncollege students as subjects, the average capacity of working memory was shown to be more like four plus or minus one—that is, three to five items (Broadbent, 1975; Mastin, 2010). Thus, in our warehouse analogy, there would be only four searchlights.

[3] Exactly how long afterwards is discussed in Chapter 14.

More recent research has cast doubt on the idea that the capacity of working memory should be measured in whole items or "chunks." It turns out that in early working memory experiments, people were asked to briefly remember items (e.g., words or images) that were quite different from each other—that is, they had very few features in common. In such a situation, people don't have to remember every feature of an item to recall it a few seconds later; remembering some of its features is enough. So people appeared to recall items as a whole, and therefore working memory capacity seemed measurable in whole items.

Recent experiments have given people items to remember that are similar—that is, they share many features. In that situation, to recall an item and not confuse it with other items, people must remember more of its features. In these experiments, researchers found that people remember more details (i.e., features) of some items than of others, and the items they remember in greater detail are the ones they paid more attention to (Bays and Husain, 2008). This suggests that the unit of attention—and therefore the capacity of working memory—is best measured in item features rather than whole items or "chunks" (Cowan et al., 2004). This jibes with the modern view of the brain as a feature-recognition device, but it is controversial among memory researchers, some of whom argue that the basic capacity of human working memory is three to five whole items, but that is reduced if people attend to a large number of details (i.e., features) of the items (Alvarez and Cavanagh, 2004).

Bottom line: The true capacity of human working memory is still a research topic.

The second important characteristic of working memory is how volatile it is. Cognitive psychologists used to say that new items arriving in working memory often bump old ones out, but that way of describing the volatility is based on the view of working memory as a temporary storage place for information. The modern view of working memory as the current focus of attention makes it even clearer: focusing attention on new information turns it away from some of what it was focusing on. That is why the searchlight analogy is useful.

However we describe it, information can easily be lost from working memory. If items in working memory don't get combined or rehearsed, they are at risk of having the focus shifted away from them. This volatility applies to goals as well as to the details of objects. Losing items from working memory corresponds to forgetting or losing track of something you were doing. We have all had such experiences, for example:

- Going to another room for something, but once there we can't remember why we came.
- Taking a phone call, and afterward not remembering what we were doing before the call.
- Something yanks our attention away from a conversation, and then we can't remember what we were talking about.
- In the middle of adding a long list of numbers, something distracts us, so we have to start over.

WORKING MEMORY TEST

To test your working memory, get a pen or pencil and two blank sheets of paper and follow these instructions:

1. Place one blank sheet of paper after this page in the book and use it to cover the next page.
2. Flip to the next page for three seconds, pull the paper cover down and read the **black numbers** at the top, and flip back to this page. Don't peek at other numbers on that page unless you want to ruin the test.
3. Say your phone number backward, out loud.
4. Now write down the black numbers from memory. ... Did you get all of them?
5. Flip back to the next page for three seconds, read the **red numbers** (under the black ones), and flip back.
6. Write down the numbers from memory. These would be easier to recall than the first ones if you noticed that they are the first seven digits of π (3.141592), because then they would be only one number, not seven.
7. Flip back to the next page for 3 seconds, read the **green numbers**, and flip back.
8. Write down the numbers from memory. If you noticed that they are odd numbers from 1 to 13, they would be easier to recall, because they would be three chunks ("odd, 1, 13" or "odd, seven from 1"), not seven.
9. Flip back to the next page for three seconds, read the **orange words**, and flip back.
10. Write down the words from memory. ... Could you recall them all?
11. Flip back to the next page for three seconds, read the **blue words**, and flip back.
12. Write down the words from memory. ... It was certainly a lot easier to recall them all because they form a sentence, so they could be memorized as one sentence rather than seven words.

3 8 4 7 5 3 9

3 1 4 1 5 9 2

1 3 5 7 9 11 13

town river corn string car shovel

what is the meaning of life

IMPLICATIONS OF WORKING MEMORY CHARACTERISTICS FOR USER-INTERFACE DESIGN

The capacity and volatility of working memory have many implications for the design of interactive computer systems. The basic implication is that user interfaces should help people remember essential information from one moment to the next. Don't require people to remember system status or what they have done, because their attention is focused on their primary goal and progress toward it. Specific examples follow.

Modes

The limited capacity and volatility of working memory is one reason why user-interface design guidelines often say to either avoid designs that have *modes* or provide adequate mode feedback. In a moded user interface, some user actions have different effects depending on what mode the system is in. For example:

- In a car, pressing the accelerator pedal can move the car either forwards, backwards, or not at all, depending on whether the transmission is in drive, reverse, or neutral.The transmission sets a mode in the car's user interface.
- In many digital cameras, pressing the shutter button can either snap a photo or start a video recording, depending on which mode is selected.

- In a drawing program, clicking and dragging normally selects one or more graphic objects on the drawing, but when the software is in "draw rectangle" mode, clicking and dragging adds a rectangle to the drawing and stretches it to the desired size.

Moded user interfaces have advantages; that is why many interactive systems have them. Modes allow a device to have more functions than controls: the same control provides different functions in different modes. Modes allow an interactive system to assign different meanings to the same gestures to reduce the number of gestures users must learn.

However, one well-known *dis*advantage of modes is that people often make *mode errors*: they forget what mode the system is in and do the wrong thing by mistake (Johnson, 1990). This is especially true in systems that give poor feedback about what the current mode is. Because of the problem of mode errors, many user-interface design guidelines say to either avoid modes or provide strong feedback about which mode the system is in. Human working memory is too unreliable for designers to assume that users can, without clear, continuous feedback, keep track of what mode the system is in, even when the users are the ones changing the system from one mode to another.

Search results

When people use a search function on a computer to find information, they enter the search terms, start the search, and then review the results. Evaluating the results often requires knowing what the search terms were. If working memory were less limited, people would always remember, when browsing the results, what they had entered as search terms just a few seconds earlier. But as we have seen, working memory is very limited. When the results appear, a person's attention naturally turns away from what he or she entered and toward the results. Therefore, it should be no surprise that people viewing search results often do not remember the search terms they just typed.

Unfortunately, some designers of online search functions don't understand that. Search results sometimes don't show the search terms that generated the results. For example, in 2006, the search results page at Slate.com provided search fields so users could search again, but didn't show what a user had searched for (see Fig. 7.3A). A recent version of the site shows the user's search terms (see Fig. 7.3B), reducing the burden on users' working memory.

Calls to action

A well-known "netiquette" guideline for writing email messages, especially messages that require responses or ask the recipients to do something, is to restrict each message to one topic. If a message contains multiple topics or requests, its recipients may focus on one of them (usually the first one), get engrossed in responding to that, and forget to respond to the rest of the email. The guideline to put different topics or requests into separate emails is a direct result of the limited capacity of human attention.

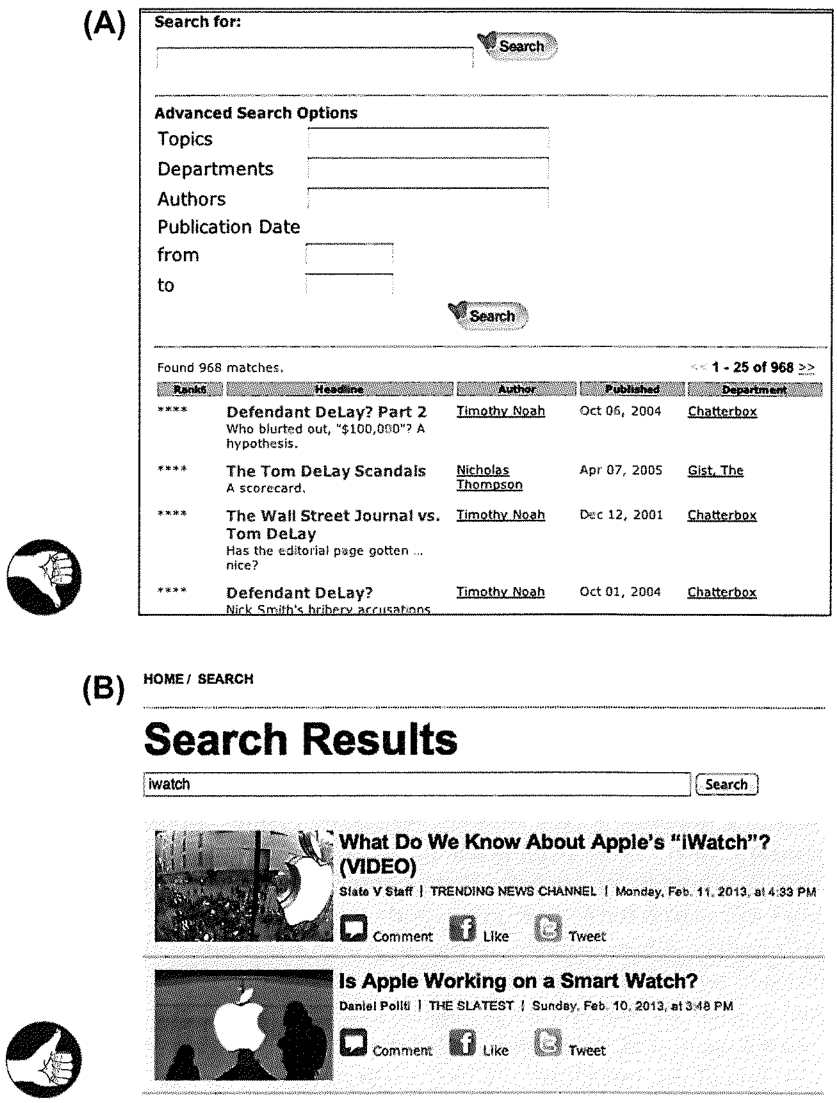


FIGURE 7.3

Slate.com search results: (A) in 2007, users' search terms were not shown, but (B) in 2013, search terms are shown.

Web designers are familiar with a similar guideline: Avoid putting competing calls to action on a page. Each page should have only *one* dominant call to action—or one for each possible user goal—to not overwhelm users' attention capacity and cause them go down paths that don't achieve their (or the site owner's) goals.

A related guideline: Once users have specified their goal, don't distract them from accomplishing it by displaying extraneous links and calls to action. Instead, guide them to the goal by using a design pattern called the *process funnel* (van Duyne et al., 2002; see also Johnson, 2007).

Instructions

If you asked a friend for a recipe or for directions to her home, and she gave you a long sequence of steps, you probably would not try to remember it all. You would know that you could not reliably keep all of the instructions in your working memory, so you would write them down or ask your friend to send them to you by email. Later, while following the instructions, you would put them where you could refer to them until you reached the goal.

Similarly, interactive systems that display instructions for multistep operations should allow people to refer to the instructions while executing them until completing all the steps. Most interactive systems do this (see Fig. 7.4), but some do not (see Fig. 7.5).

Navigation depth

Using a software product, digital device, phone menu system, or Web site often involves navigating to the user's desired information or goal. It is well established that navigation hierarchies that are broad and shallow are easier for most people—especially those who are nontechnical—to find their way around in than narrow, deep hierarchies (Cooper, 1999). This applies to hierarchies of application windows and dialog boxes, as well as to menu hierarchies (Johnson, 2007).

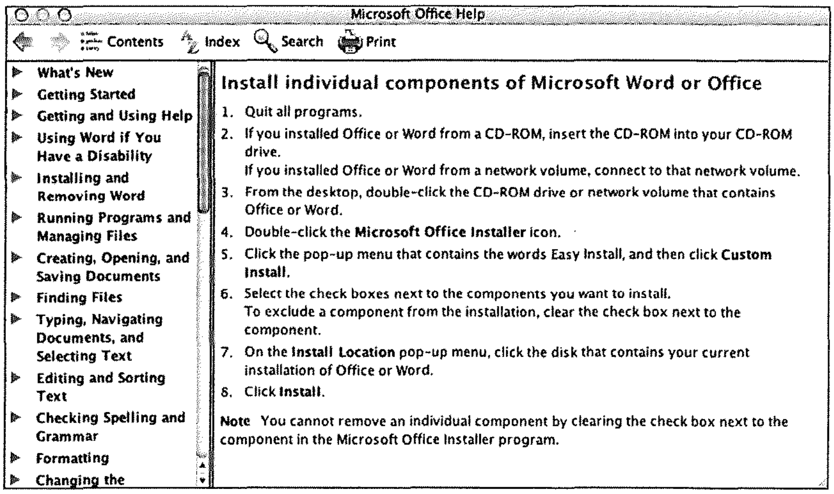


FIGURE 7.4

Instructions in Windows Help files remain displayed while users follow them.

FIGURE 7.5

Instructions for Windows XP wireless setup start by telling users to close the instructions.

A related guideline: In hierarchies deeper than two levels, provide navigation "breadcrumb" paths to constantly remind users where they are (Nielsen, 1999; van Duyne et al., 2002).

These guidelines, like the others mentioned earlier, are based on the limited capacity of human working memory. Requiring a user to drill down through eight levels of dialog boxes, web pages, menus, or tables—especially with no visible reminders of their location—will probably exceed the user's working memory capacity, thereby causing him or her to forget where he or she came from or what his or her overall goals were.

CHARACTERISTICS OF LONG-TERM MEMORY

Long-term memory differs from working memory in many respects. Unlike working memory, it actually *is* a memory store.

However, specific memories are not stored in any one neuron or location in the brain. As described earlier, memories, like perceptions, consist of patterns of activation of large sets of neurons. Related memories correspond to overlapping patterns of activated neurons. This means that every memory is stored in a distributed fashion, spread among many parts of the brain. In this way, long-term memory in the brain is similar to holographic light images.

Long-term memory evolved to serve our ancestors and us very well in getting around in our world. However, it has many weaknesses: it is error-prone, impressionist, free-associative, idiosyncratic, retroactively alterable, and easily biased by a variety of factors at the time of recording or retrieval. Let's examine some of these weaknesses.

Error-prone

Nearly everything we've ever experienced is stored in our long-term memory. Unlike working memory, the capacity of human long-term memory seems almost unlimited. Adult human brains each contain about 86 billion neurons (Herculano-Houzel, 2009). As described earlier, individual neurons do not store memories; memories are encoded by networks of neurons acting together. Even if only some of the brain's neurons are involved in memory, the large number of neurons allows for a great many different combinations of them, each capable of representing a different memory. Still, no one has yet measured or even estimated the maximum information capacity of the human brain.[4] Whatever the capacity is, it's a lot.

However, what is in long-term memory is not an accurate, high-resolution recording of our experiences. In terms familiar to computer engineers, one could characterize long-term memory as using heavy compression methods that drop a great deal of information. Images, concepts, events, sensations, actions—all are reduced to combinations of abstract features. Different memories are stored at different levels of detail—that is, with more or fewer features.

For example, the face of a man you met briefly who is not important to you might be stored simply as an average Caucasian male face with a beard, with no other details—a whole face reduced to three features. If you were asked later to describe the man in his absence, the most you could honestly say was that he was a "white guy with a beard." You would not be able to pick him out of a police lineup of other Caucasian men with beards. In contrast, your memory of your best friend's face includes many more features, allowing you to give a more detailed description and pick your friend out of any police lineup. Nonetheless, it is still a set of features, not anything like a bitmap image.

As another example, I have a vivid childhood memory of being run over by a plow and badly cut, but my father says it happened to my brother. One of us is wrong.

In the realm of human-computer interaction, a Microsoft Word user may remember that there is a command to insert a page number, but may not remember which *menu* the command is in. That specific feature may not have been recorded when the user learned how to insert page numbers. Alternatively, perhaps the menu-location feature *was* recorded, but just does not reactivate with the rest of the memory pattern when the user tries to recall how to insert a page number.

Weighted by emotions

Chapter 1 described a dog that remembered seeing a cat in his front yard every time he returned home in the family car. The dog was excited when he first saw the cat, so his memory of it was strong and vivid.

A comparable human example would be an adult could easily have strong memories of her first day at nursery school, but probably not of her tenth. On the first day, she was probably upset about being left at the school by her parents, whereas by the tenth day, being left there was nothing unusual.

Retroactively alterable

Suppose that while you are on an ocean cruise with your family, you see a whale-shark. Years later, when you and your family are discussing the trip, you might remember

[4]The closest researchers have come is Landauer's (1986) use of the average human learning rate to calculate the amount of information a person can learn in a lifetime: 10^9 bits, or a few hundred megabytes.

A LONG-TERM MEMORY TEST

Test your long-term memory by answering the following questions:

1. Was there a roll of tape in the toolbox in Chapter 1?
2. What was your *previous* phone number?
3. Which of these words were *not* in the list presented in the working memory test earlier in this chapter: city, stream, corn, auto, twine, spade?
4. What was your first-grade teacher's name? Second grade? Third grade? ...
5. What Web site was presented earlier that does not show search terms when it displays search results?

Regarding question 3: When words are memorized, often what is retained is the *concept*, rather than the exact word that was presented. For example, one could hear the word "town" and later recall it as "city."

seeing a whale, and one of your relatives might recall seeing a shark. For both of you, some details in long-term memory were dropped because they did not fit a common concept.

A true example comes from 1983, when the late President Ronald Reagan was speaking with Jewish leaders during his first term as president. He spoke about being in Europe during World War II and helping to liberate Jews from the Nazi concentration camps. The trouble was, he was never in Europe during World War II. When he was an actor, he was in a *movie* about World War II, made entirely in Hollywood. That important detail was missing from his memory.

IMPLICATIONS OF LONG-TERM MEMORY CHARACTERISTICS FOR USER-INTERFACE DESIGN

The main thing that the characteristics of long-term memory imply is that people need tools to augment it. Since prehistoric times, people have invented technologies to help them remember things over long periods: notched sticks, knotted ropes, mnemonics, verbal stories and histories retold around campfires, writing, scrolls, books, number systems, shopping lists, checklists, phone directories, datebooks, accounting ledgers, oven timers, computers, portable digital assistants (PDAs), online shared calendars, etc.

Given that humankind has a need for technologies that *augment* memory, it seems clear that software designers should try to provide software that fulfills that need. At the very least, designers should avoid developing systems that *burden* long-term memory. Yet that is exactly what many interactive systems do.

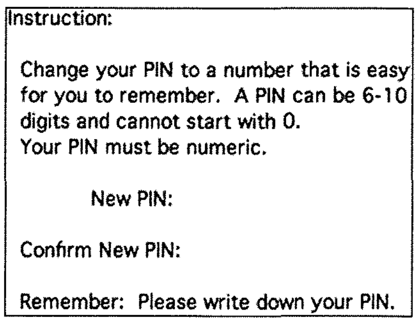
Authentication is one functional area in which many software systems place burdensome demands on users' long-term memory. For example, a web application developed a few years ago told users to change their personal identification number (PIN) "to a number that is easy to remember," but then imposed restrictions that made it impossible to do so (see Fig. 7.6). Whoever wrote those instructions seems to have realized that the PIN requirements were unreasonable, because the instructions end by advising users to write down their PIN! Nevermind that writing a PIN down creates a security risk and adds yet *another* memory task: users must remember where they hid their written-down PIN.

A contrasting example of burdening people's long-term memory for the sake of security comes from Intuit.com. To purchase software, visitors must register. The site requires users to select a security question from a menu (see Fig. 7.7). What if you can't answer *any* of the questions? What if you don't recall your first pet's name, your high school mascot, or any of the answers to the other questions?

But that isn't where the memory burden ends. Some questions could have several possible answers. Many people had several elementary schools, childhood friends, or heroes. To register, they must choose a question and then *remember* which answer they gave to Intuit.com. How? Probably by writing it down somewhere. Then, when Intuit.com asks them the security question, they have to *remember* where they put the answer. Why burden people's memory, when it would be easy to let users make up a security question for which they can easily recall the one possible answer?

Such unreasonable demands on people's long-term memory counteract the security and productivity that computer-based applications supposedly provide (Schrage, 2005), as users:

- Place sticky notes on or near computers or "hide" them in desk drawers.
- Contact customer support to recover passwords they cannot recall.


Instruction:

Change your PIN to a number that is easy for you to remember. A PIN can be 6-10 digits and cannot start with 0.
Your PIN must be numeric.

New PIN:

Confirm New PIN:

Remember: Please write down your PIN.

FIGURE 7.6

Instructions tell users to create an easy-to-remember PIN, but the restrictions make that impossible.

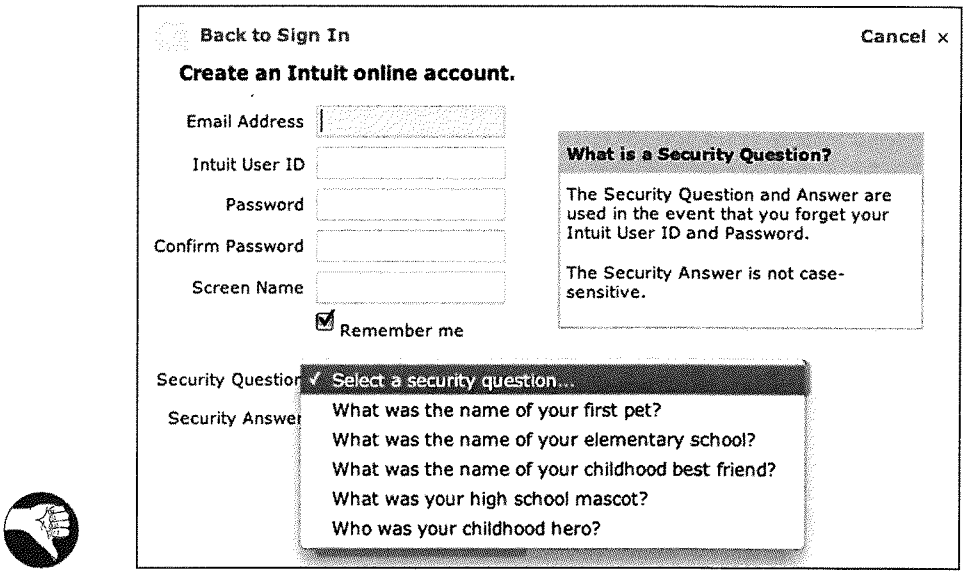


FIGURE 7.7

Intuit.com's registration burdens long-term memory: users may have no unique, memorable answer for any of the questions.

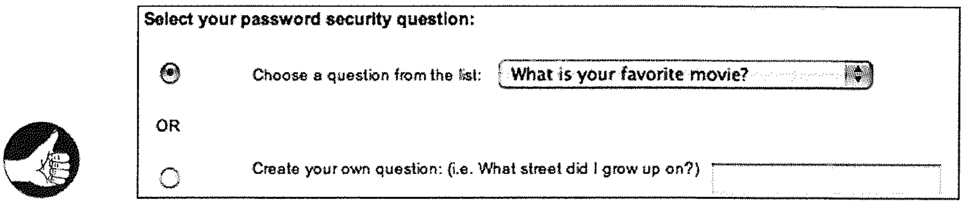


FIGURE 7.8

NetworkSolutions.com lets users create a security question if none on the menu works for them.

- Use passwords that are easy for others to guess.
- Set up systems with no login requirements at all, or with one shared login and password.

The registration form at NetworkSolutions.com represents a small step toward more usable security. Like Intuit.com, it offers a choice of security questions, but it also allows users to create their own security question—one for which they can more easily remember the answer (see Fig. 7.8).

Another implication of long-term memory characteristics for interactive systems is that learning and long-term retention are enhanced by user-interface consistency.

The more consistent the operation of different functions, or the more consistent the actions on different types of objects, the less users have to learn. User interfaces that have many exceptions and little consistency from one function or object to another require users to store in their long-term memory many features about each function or object and its correct usage context. The need to encode more features makes such user interfaces harder to learn. It also makes it more likely that a user's memory will drop essential features during storage or retrieval, increasing the chances that the user will fail to remember, misremember, or make other memory errors.

Even though some have criticized the concept of consistency as ill-defined and easy to apply badly (Grudin, 1989), the fact is that consistency in a user interface greatly reduces the burden on users' long-term memory. Mark Twain once wrote: "If you tell the truth, you never have to remember anything." One could also say, "If everything worked the same way, you would not have to remember much." We will return to the issue of user-interface consistency in Chapter 11.

This page intentionally left blank

CHAPTER

Limits on Attention Shape Our Thought and Action

8

When people interact purposefully with the world around them, including computer systems, some aspects of their behavior follow predictable patterns, some of which result from the limited capacity of attention and short-term memory. When interactive systems are designed to recognize and support those patterns, they fit better with the way people operate. Some user-interface design rules, then, are based directly on the patterns and thus indirectly on the limits of short-term memory and attention. This chapter describes seven important patterns.

WE FOCUS ON OUR GOALS AND PAY LITTLE ATTENTION TO OUR TOOLS

As Chapter 7 explained, our attention has very limited capacity. When people are doing a task—trying to accomplish a goal—most of their attention is focused on the goals and data related to that task. Normally, people devote very little attention to the tools they are using to perform a task, whether they are using computer applications, online services, or interactive appliances. Instead, people think about their tools only superficially, and then only when necessary.

We are, of course, *capable* of attending to our tools. However, attention (i.e., short-term memory) is limited in capacity. When people refocus their attention on their tools, it is pulled away from the details of the task. This shift increases the chances of users losing track of what they were doing or exactly where they were in doing it.

For example, if your lawn mower stops running while you are mowing your lawn, you will immediately stop and focus on the mower. Restarting the mower becomes your primary task, with the mower itself as the focus. You pay scant attention to any tools you use to restart the mower, just as you paid scant attention to the mower when your primary focus was the lawn. After you restart the mower and

Designing with the Mind in Mind. http://dx.doi.org/10.1016/B978-0-12-407914-4.00008-7

resume mowing the lawn, you probably wouldn't remember where you were in mowing the lawn, except that the lawn itself shows you.

Other tasks—for example, reading a document, measuring a table, counting goldfish in a fish tank—might not provide such a clear reminder of your interrupted task and your position in it. You might have to start over from the beginning. You might even forget what you were doing altogether and go off to do something else.

That is why most software design guidelines state that software applications and most Web sites should not call attention to themselves; they should fade into the background and allow users to focus on their own goals. That design guideline is even the title of a popular web design book: *Don't Make Me Think* (Krug, 2005). That is, if your software or Web site makes me think about *it*, rather than what I am trying to do, you've lost me.

WE NOTICE THINGS MORE WHEN THEY ARE RELATED TO OUR GOALS

Chapter 1 described how our immediate goals filter and bias our perception. Chapter 7 discussed the connections between attention and memory. This chapter provides examples showing that perceptual filtering and biasing, attention, and memory are all closely related.

Our environment is full of perceptual details and events. Obviously, we cannot notice and keep track of everything that happens around us. Nonetheless, it is surprising to most people how little we notice of what goes on around us. Because of our extremely limited short-term memory and attention, we don't waste those resources. When an event happens, the few details about it that we notice and remember later are usually those that were important for our goals at the time of the event. This is demonstrated by two related psychological phenomena: *inattentional blindness* and *change blindness.*

Inattentional blindness

When our mind is intensely occupied with a task, goal, or emotion, we sometimes fail to notice objects and events in our environment that we otherwise would have noticed and remembered. This phenomenon has been heavily studied by psychologists and labeled *inattentional blindness* (Simons and Chabris, 1999; Simons, 2007).

A clear demonstration of inattentional blindness is an experiment in which human subjects watched a video of two basketball teams passing a ball from one player to another. Subjects were told to count the number of times the white-suited team passed the ball. While they watched the video and counted the ball-passes, a person in a gorilla suit sauntered onto the basketball court, thumped his chest, then walked out of view (see Fig. 8.1). Afterwards, subjects were asked what they remembered from the video. Surprisingly, about half of them did not notice the gorilla. Their attention was fully occupied with the task (Simons and Chabris, 1999).

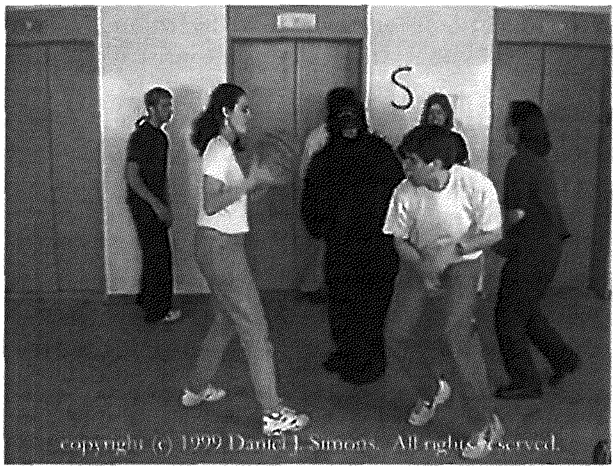


FIGURE 8.1

Scene from video used in "invisible gorilla" study. *Figure provided by Daniel Simons.* For more information about the study and to see the video, go to www.dansimons.com or www.theinvisiblegorilla.com.

Change blindness

Another way researchers have shown that our goals strongly focus our attention and memory is through the following: show people a picture, then show them a second version of the same picture and ask them how the two pictures differ. Surprisingly, the second picture can differ from the first in many ways without people noticing. To explore further, researchers gave people questions to answer about the first picture, affecting their goals in looking at it, and therefore what features of the picture they paid attention to. The result was people don't notice differences in features other than those their goals made them pay attention to. This is called *change blindness* (Angier, 2008).

A particularly striking example of how our goals focus our attention and affect what we remember comes from experiments in which experimenters holding city maps posed as lost tourists and asked local people walking by for directions. When the local person focused on the tourist's map to figure out the best route, two workmen—actually, more experimenters—walked between the tourist and the advice-giver carrying a large door, and in that moment the tourist was replaced by another experimenter-tourist. Astoundingly, after the door passed, over half of the local people continued helping the tourist without noticing any change, even when the two tourists differed in hair color or in whether they had a beard (Simons and Levin, 1998).[1] Some people even failed to notice changes in gender. In summary,

[1] For demonstrations of change blindness, search YouTube for those words, or for "door study" and "person swap."

people focus on the tourist only long enough to determine if he or she is a threat or worth helping, record only that the person is a tourist who needs help, and then focus on the map and the task of giving directions.

When people interact with software, electronic appliances, or online services, it is not uncommon for them to fail to notice important changes in what is displayed. For example, in a study of seniors using travel websites, changes in price resulting from user actions were often not obvious. Even after spotting the price information, participants were prone to change blindness: when they changed trip options (e.g., departure city, additional excursions, cabin class), they often would not notice that the price changed (Finn and Johnson, 2013). The study only used older adults, so we don't know whether younger participants would have had the same trouble noticing price changes.

The user-interface design guideline that follows from such findings is to make changes obvious—that is, highly salient—and take steps to draw users' attention to the change. For example, a way to draw users' attention to a new error message is to vibrate it briefly when it first appears (see Chapter 6), or highlight it briefly before it reverts to a "normal" appearance.

What happens in our brains

Using functional magnetic resonance imagery (fMRI) and electrical encephalography (EEG), researchers have studied the effect of attention on how our brains respond to objects displayed on a computer screen.

When people passively watch a computer display with objects appearing, moving around, and disappearing, the visual cortex of their brains registers a certain activity level. When people are told to look for (i.e., pay attention to) certain objects, the activity level of their visual cortex increases significantly. When they are told to ignore certain objects, the neural activity level in their visual cortex actually drops when those objects appear. Later, their memory for which objects they saw and didn't see corresponds to the degree of attention they paid to them and to the level of brain activity (Gazzaley, 2009).

WE USE EXTERNAL AIDS TO KEEP TRACK OF WHAT WE ARE DOING

Because our short-term memory and attention are so limited, we learn not to rely on them. Instead, we mark up our environment to show us where we are in a task. Examples include:

- ***Counting objects.*** If possible, we move already counted objects into a different pile to indicate which objects have already been counted. If we cannot move an object, we point to the last object counted. To keep track of the number we are on, we count on our fingers, draw marks, or write numbers.

- ***Reading books.*** When we stop reading, we insert bookmarks to show what page we were on.
- ***Arithmetic.*** We learn methods of doing arithmetic on paper, or we use a calculator.
- ***Checklists.*** We use checklists to aid both our long- and short-term memory. In critical or rarely performed tasks, checklists help us remember everything that needs to be done. In that way, they augment our faulty long-term memory. While doing the task, we check off items as we complete them. That is a short-term memory aid. A checklist that we can't mark up is hard to use, so we copy it and mark the copy.
- ***Editing documents.*** People often keep to-be-edited documents, documents that are currently being edited, and already edited documents in separate folders.

One implication of this pattern is that interactive systems should indicate what users have done versus what they have not yet done. Most email applications do this by marking already-read versus unread messages, most Web sites do it by marking visited versus unvisited links, and many applications do it by marking completed steps of a multipart task (see Fig. 8.2).

A second design implication is that interactive systems should allow users to mark or move objects to indicate which ones they have worked on versus which ones they have not worked on. Mac OS lets users assign colors to files. Like moving files between folders, this technique can be used to keep track of where one is in a task (see Fig. 8.3).

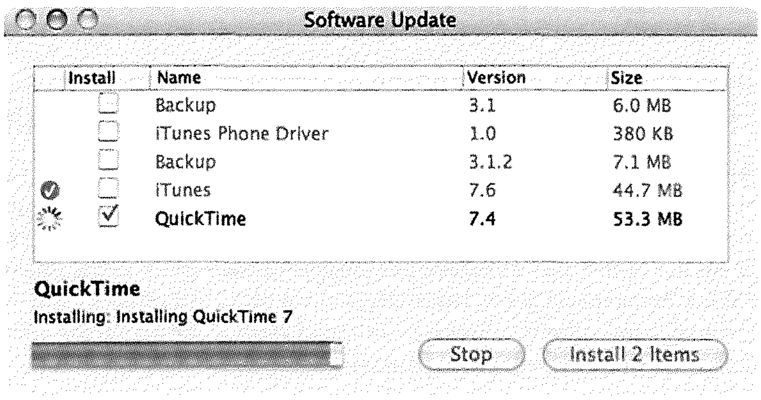


FIGURE 8.2

The Mac OS Software Update shows which updates are done (green check) versus which are in progress (rotating circle).

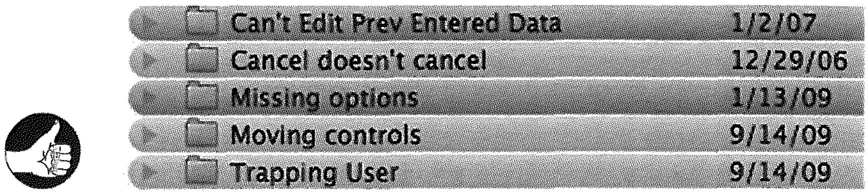
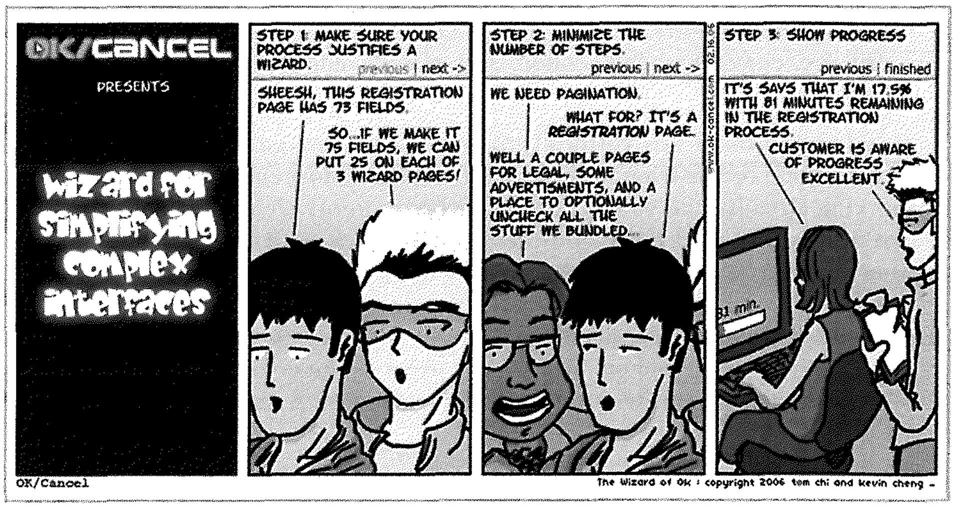


FIGURE 8.3

Mac OS lets users assign colors to files or folders; users can use the colors to track their work.

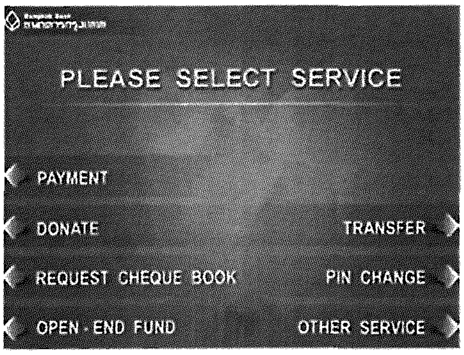


Used by permission, www.OK/Cancel.com.

WE FOLLOW THE INFORMATION "SCENT" TOWARD OUR GOAL

Focusing our attention on our goals makes us interpret what we see on a display or hear in a telephone menu in a very *literal* way. People don't think deeply about instructions, command names, option labels, icons, navigation bar items, or any other aspect of the user interface of computer-based tools. If the goal in their head is to make a flight reservation, their attention will be attracted by anything displaying the words "buy," "flight," "ticket," or "reservation." Other items that a designer or marketer might think will attract customers, such as "bargain hotels," will not attract the attention of people who are trying to book a flight, although they might be noticed by people who are looking for bargains.

This tendency of people to notice only things on a computer display that match their goal, and the literal thinking that they exhibit when performing a task on a computer, has been called "following the *scent* of information toward the goal" (Chi



For each goal below, what on the screen would attract your attention?

- Pay a bill
- Transfer money to your savings account
- Pay your dentist by funds transfer
- Change your PIN
- Open a new account
- Purchase travelers' cheque

FIGURE 8.4

ATM screen—our attention is drawn initially toward items that match our goal literally.

et al., 2001; Nielsen, 2003). Consider the ATM machine display shown in Figure 8.4. What is the first thing on the screen that gets your attention when you are given each of the goals listed?

You probably noticed that some of the listed goals direct your attention initially to the wrong option. Is "Pay your dentist by funds transfer" under "Payment" or "Transfer"? "Open a new account" probably sent your eyes briefly to "Open-End Fund," even though it is actually under "Other Service." Did the goal "Purchase traveler's cheques" make you glance at "Request Cheque Book" because of the word they share?

The goal-seeking strategy of following the information scent, observed across a wide variety of situations and systems, suggests that interactive systems should be designed so that the scent is strong and really leads users to their goals. To do that, designers need to understand the goals that users are likely to have at each decision point in a task, and ensure that each choice point in the software provides options for every important user goal and clearly indicates which option leads to which goal.

For example, imagine that you want to cancel a reservation you made or a payment you scheduled. You tell the system to cancel it, and a confirmation dialog box appears asking if you really want to do that. How should the options be labeled? Given that people interpret words literally in following the information scent toward their goal, the standard confirmation button labels "OK" (for yes) and "Cancel" (for no) would give a misleading scent. If we compare a cancellation confirmation dialog box from Marriott.com to one from WellsFargo.com, we see that Marriott.com's labeling provides a clearer scent than Quicken.com's (see Fig. 8.5).

As a second example, imagine that you forgot that a certain document was already open, and you tried to open it again. The designers of Microsoft Excel did a better job than the designers of Microsoft Word in anticipating this situation, understanding the goals you might have at this point, and presenting you with instructions and options that make it clear what to do (see Fig. 8.6).

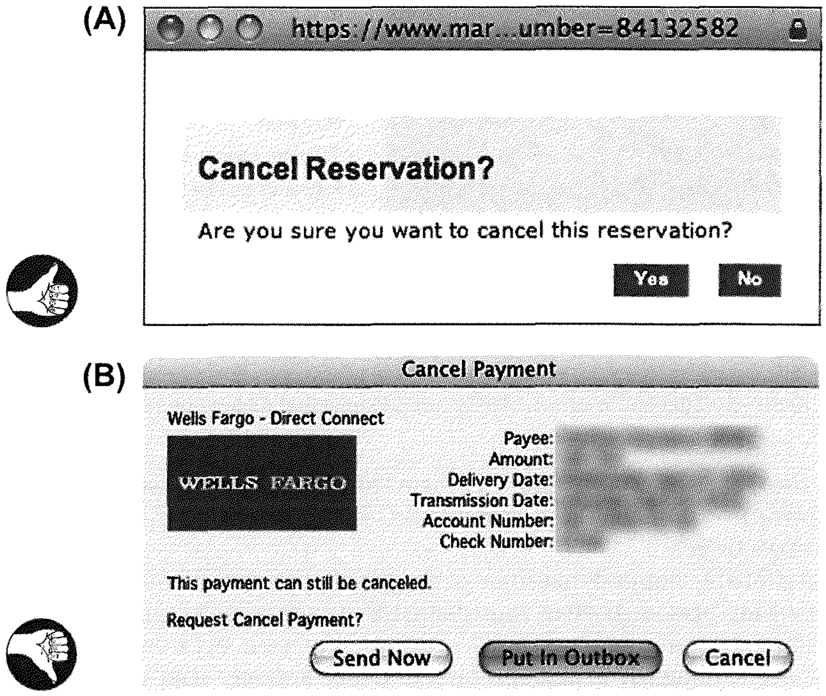


FIGURE 8.5

Marriott's cancellation confirmation (A) provides a clearer scent than Wells Fargo's (B).

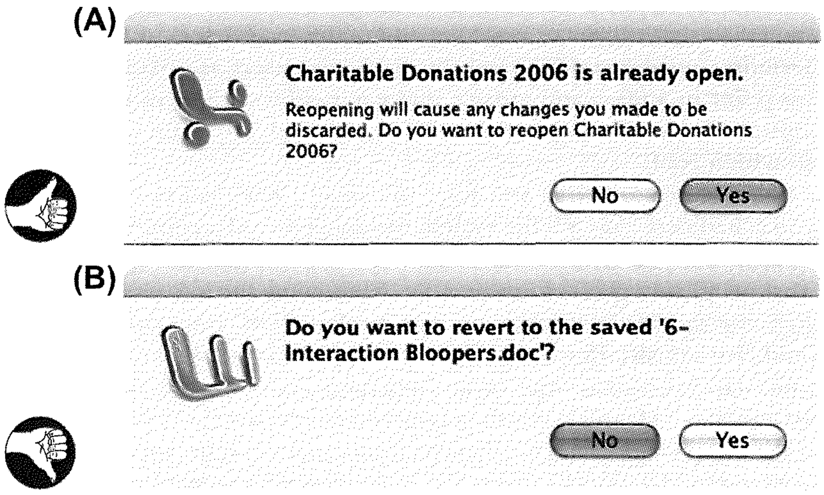


FIGURE 8.6

Microsoft Excel's (A) warning when users try to open an already-open file is clearer than Microsoft Word's (B).

WE PREFER FAMILIAR PATHS

People know that their attention is limited, and they act accordingly. While pursuing a goal, they take *familiar* paths whenever possible rather than exploring *new* ones, especially when working under deadlines. As explained more fully in Chapter 10, exploring new paths is problem solving, which severely taxes our attention and short-term memory. In contrast, taking familiar, well-learned routes can be done fairly automatically and does not consume much attention and short-term memory.

Years ago, in a usability test session, a test participant in the middle of a task said to me:

> I'm in a hurry, so I'll do it the long way.

He knew there probably was a more efficient way to do what he was doing, but he also knew that learning the shorter way would require time and thought, which he was unwilling to spend.

Once we learn one way to perform a certain task using a software application, we may continue to do it that way and *never* discover a more efficient way. Even if we discover or are told that there is a "better" way, we may stick with the old way because it is familiar, comfortable, and, most important, requires little thought. Avoiding thought when using computers is important. People are willing to type *more* to think *less*.

Why is that? Are we mentally lazy? Usually, yes. Conscious thought is slow, strains working memory, and consumes much energy. We essentially run on batteries—that is, the food we eat—so energy conservation is an important feature that is part of our makeup. Operating via automatic processes is fast, doesn't strain working memory, and conserves energy. So the brain tries to run on automatic as much as possible (Kahneman, 2011; Eagleman, 2012).

The human preference for familiar, mindless paths and "no-brainer" decisions has several design implications for interactive systems:

- ***Sometimes mindlessness trumps keystrokes.*** With software intended for casual or infrequent use, such as bank ATM machines or household accounting applications, allowing users to become productive quickly and reducing their need to problem-solve while working is more important than saving keystrokes. Such software simply isn't used enough for the number of keystrokes per task to matter much. On the other hand, in software that is used all day by highly trained users in intensive work environments, such as airline telephone reservation operators, unnecessary extra keystrokes needed to perform a task are very costly.
- ***Guide users to the best paths.*** From its first screen or homepage, software should show users the way to their goals. This is basically the guideline that software should provide a clear information scent.

- ***Help experienced users speed up.*** Make it easy for users to switch to faster paths after they have gained experience. The slower paths for newcomers should show users faster paths if there are any. This is why most applications show the keyboard accelerators for frequently used functions in the menu-bar menus.

OUR THOUGHT CYCLE: GOAL, EXECUTE, EVALUATE

Over many decades, scientists studying human behavior have found a cyclical pattern that seems to hold across a wide variety of activities:

- Form a ***goal*** (e.g., open a bank account, eat a peach, or delete a word from a document).
- Choose and ***execute*** actions to try to make progress toward the goal.
- ***Evaluate*** whether the actions worked—that is, whether the goal has been reached or is nearer than before.
- Repeat until the goal is reached (or appears unreachable).

People cycle through this pattern constantly (Card et al., 1983). In fact, we run through it at many different levels simultaneously. For example, we might be trying to insert a picture into a document, which is part of a higher-level task of writing a term paper, which is part of a higher-level task of passing a history course, which is part of a higher-level task of completing college, which is part of a higher-level goal of getting a good job, which we want to achieve our top-level goal of having a comfortable life.

As an example, let's run through the cycle for a typical computer task: buying an airline ticket online. The person first forms the primary goal of the task and then begins to break that down into actions that appear to lead toward the goal. Promising actions are selected for execution, executed, and then evaluated to determine if they have moved the person closer to the goal.

- ***Goal:*** Buy airline ticket to Berlin, using your favorite travel Web site.
- ***Step 1:*** Go to travel Web site. You are still far from the goal.
- ***Step 2:*** Search for suitable flights. This is a very normal, predictable step at travel Web sites.
- ***Step 3:*** Look at search results. Choose a flight from those listed. If no flights on the results list are suitable, return to Step 2 with new search criteria. You are not at the goal yet, but you feel confident of getting there.
- ***Step 4:*** Go to checkout. Now you are getting so close to your goal that you can almost smell it.

- ***Step 5:*** Confirm flight details. Check it—all correct? If no, back up; otherwise proceed. Almost done.
- ***Step 6:*** Purchase ticket with credit card. Check credit card information. Everything look okay?
- ***Step 7:*** Print e-ticket. Goal achieved.

In the airline ticket example, to keep the example short, we didn't get down into the details of each step. If we had, we would have seen substeps that followed the same *goal-execute-evaluate* cycle.

Let's try another example, this time examining the details of some of the higher-level steps. This time the task is sending flowers to a friend. If we simply look at the top level, we see the task like this:

Send flowers to friend.

If we want to examine the goal-execute-evaluate cycle for this task, we must break down this task a bit. We must ask: *How* do we send flowers to a friend? To do that, we break the top-level task down into subtasks:

Send flowers to friend.
  Find flower delivery Web site.
  Order flowers to be delivered to friend.

For many purposes, the two steps we have identified are enough detail. After we execute each step, we evaluate whether we are closer to our goal. But *how* is each step executed? To see that, we have to treat each major step as a subgoal, and break it down into substeps:

Send flowers to friend.
  Find flower delivery Web site.
    Open web browser.
    Go to Google web search page.
    Type "flower delivery" into Google.
    Scan the first page of search results.
    Visit some of the listed links.
    Choose a flower delivery service.
  Order flowers to be delivered to friend.
    Review service's flower selection.
    Choose flowers.
    Specify delivery address and date.
    Pay for flowers and delivery.

After each substep is executed, we evaluate to see if it is getting us closer to the subgoal of which it is part. If we want to examine how a substep is executed and evaluated, we have to treat it as a sub-subgoal and break it into its component steps:

Send flowers to friend.
- Find flower delivery Web site.
  - Open web browser.
    - Click browser icon on taskbar, startup menu, or desktop.
  - Go to Google web search page.
    - If Google isn't browser's starting page, choose "Google" from favorites list.
    - If Google is not on favorites list, type "Google.com" into browser's address box.
  - Type "flower delivery" into Google.
    - Set text-insertion point in search box.
    - Type the text.
    - Correct typo: "floowers" to "flowers."
  - Visit some of the resulting links.
    - Move screen pointer to link.
    - Click on link.
    - Look at resulting web page.
  - Choose a flower delivery service.
    - Enter chosen service's URL into browser.

...

You get the idea. We could keep expanding, down to the level of individual keystrokes and individual mouse movements, but we rarely need that level of detail to be able to understand the task well enough to design software to fit its steps and the goal-execute-evaluate cycle that is applied to each step.

How can software support users in carrying out the goal-execute-evaluate cycle? Any of these ways:

- ***Goal.*** Provide clear paths—including initial steps—for the user goals that the software is intended to support.
- ***Execute.*** Software concepts (objects and actions) should be based on the task rather than the implementation (see Chapter 11). Don't force users to figure out how the software's objects and actions map to those of the task. Provide a clear information scent at choice points to guide users to their goals. Don't make them choose actions that seem to take them away from their goal to achieve it.
- ***Evaluate.*** Provide feedback and status information to show users their progress toward the goal. Allow users to back out of tasks that didn't take them toward their goal.

An example of the evaluate guideline—clear feedback about the user's progress through a series of steps—is provided by ITN.com's flight reservation system (see Fig. 8.7). By the way, does the figure seem familiar? If so, it is because you saw it in Chapter 4 (see Fig. 4.16B), and your brain recognized it.

FIGURE 8.7

ITN.com's flight reservation system clearly shows users' progress toward making a reservation.

AFTER WE ACHIEVE A TASK'S PRIMARY GOAL, WE OFTEN FORGET CLEANUP STEPS

The goal-execute-evaluate cycle interacts strongly with short-term memory. This interaction makes perfect sense: short-term memory is really just what the focus of attention is at any given moment. Part of the focus of attention is our current goal. The rest of our attentional resources are directed toward obtaining the information needed to achieve our current goal. The focus shifts as tasks are executed and the current goal shifts from high-level goals to lower-level ones, then back to the next high-level goal.

Attention is a very scarce resource. Our brain does not waste it by keeping it focused on anything that is no longer important. Therefore, when we complete a task, the attentional resources focused on that task's main goal are freed to be refocused on other things that are now more important. The impression we get is that once we achieve a goal, everything related to it often immediately "falls out" of our short-term memory—that is, we forget about it.

One result is that people often forget loose ends of tasks. For example, people often forget to do these things:

- Turn car headlights off after arrival.
- Remove the last pages of documents from copiers and scanners.
- Turn stove burners and ovens off after use.
- Add closing parentheses and quotation marks after typing text passages.
- Turn off turn signals after completing turns.
- Take books they were reading on a flight with them when they exit the plane.
- Log out of public computers when finished using them.
- Set devices and software back into normal mode after putting them into a special mode.

These end-of-task short-term memory lapses are completely predictable and avoidable. When they happen to us, we call ourselves "absent-minded," but they are the result of how the brain works (or doesn't), combined with a lack of support from our devices.

To avoid such lapses, interactive systems can and should be designed to remind people that loose-end steps remain. In some cases, it may even be possible for the system to complete the task itself. For example:

- Cars already turn off turn signals after a turn.
- Cars should (and now do) turn off headlights automatically when the car is no longer in use, or at least remind drivers that the lights are still on.
- Copiers and scanners should automatically eject all documents when tasks are finished, or at least signal that a page has been left behind.
- Stoves should signal when a burner is left on with no pot present for longer than some suitable interval, and ovens should do likewise when left on with nothing in them.
- Computers should issue warnings if users try to power them down or put them to sleep before the computer has finished a background task (e.g., saving files or sending a document to a printer).
- Special software modes should revert to "normal" automatically, either by timing out—as some appliances do—or through the use of spring-loaded mode controls, which must be physically held in the non-normal state and revert to normal when released (Johnson, 1990).

Software designers should consider whether the tasks supported by a system they are designing have cleanup steps that users are likely to forget, and if so, they should design the system either to help users remember, or eliminate the need for users to remember.

CHAPTER

Recognition is Easy; Recall is Hard

9

Chapter 7 described the strengths and limitations of long-term memory and their implications for the design of interactive systems. This chapter extends that discussion by describing important differences between two functions of long-term memory: recognition and recall.

RECOGNITION IS EASY

The human brain was "designed," through millions of years of natural selection and evolution, to recognize things quickly. By contrast, recalling memories—that is, retrieving them without perceptual support—must not have been as crucial for survival, because our brains are much worse at that.

Remember how our long-term memory works (see Chapter 7): Perceptions enter through our sensory systems, and their signals, when they reach the brain, cause complex patterns of neural activity. The neural pattern resulting from a perception is determined not only by the features of the perception, but also by the context in which it occurs. Similar perceptions in similar contexts cause similar patterns of neural activity. Repeated activation of a particular neural pattern makes that pattern easier to reactivate in the future. Over time, connections between neural patterns develop in such a way that activating one pattern activates the other. Roughly speaking, each pattern of neural activity constitutes a different memory.

Patterns of neural activity, which is what memories are, can be activated in two different ways:

1. By more perceptions coming in from the senses.
2. By other brain activity.

If a perception comes in that is similar to an earlier one and the context is close enough, it easily stimulates a similar pattern of neural activity, resulting in a sense of

Designing with the Mind in Mind. http://dx.doi.org/10.1016/B978-0-12-407914-4.00009-9

recognition. Recognition is essentially perception and long-term memory working in concert.

As a result, we assess situations very quickly. Our distant ancestors on the East African savannah had only a second or two to decide whether an animal emerging from the tall grasses was something they would regard as food or something that would regard *them* as food (see Fig. 9.1). Their survival depended on it.

Similarly, people recognize human faces very quickly—usually in a fraction a second (see Fig. 9.2). Until recently, the workings of this process were considered a mystery. However, that was when scientists assumed that recognition was a process

FIGURE 9.1

Early hominids had to recognize quickly whether animals they spotted were prey or predators.

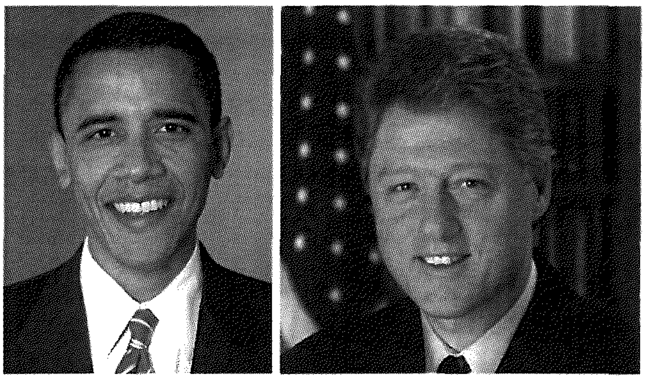


FIGURE 9.2

How long did it take you to recognize these faces?[1]

[1] U.S. Presidents Barack Obama and Bill Clinton.

in which perceived faces were stored in a separate short-term memory and compared with those in long-term memory. Because of the speed with which the brain recognizes faces, cognitive scientists assumed that the brain must search many parts of long-term memory simultaneously, via what computer scientists call *parallel processing*. However, even a massively parallel search process could not account for the astounding rapidity of facial recognition.

Nowadays, perception and long-term memory are considered closely linked, which demystifies the speed of facial recognition somewhat. A perceived face stimulates activity in millions of neurons in distinct patterns. Individual neurons and groups of neurons that make up the pattern respond to specific features of the face and the context in which the face is perceived. Different faces stimulate different patterns of neural response. If a face was perceived previously, its corresponding neural pattern will already have been activated. The same face perceived again reactivates the same pattern of neural activity, only more easily than before. That *is* the recognition. There is no need to search long-term memory: the new perception reactivates the same pattern of neural activity, more or less, as the previous one. Reactivation of a pattern *is* the reactivation of the corresponding long-term memory.

In computer jargon, we could say that the information in human long-term memory is *addressed by its content*, but the word "addressed" wrongly suggests that each memory is located at a specific spot in the brain. In fact, each memory corresponds to a pattern of neural activity extending over a wide area of the brain.

That explains why, when presented with faces we have not seen before and asked if they are familiar, we don't spend a long time searching through our memories to try to see if that face is stored in there somewhere (see Fig. 9.3). There is no search.

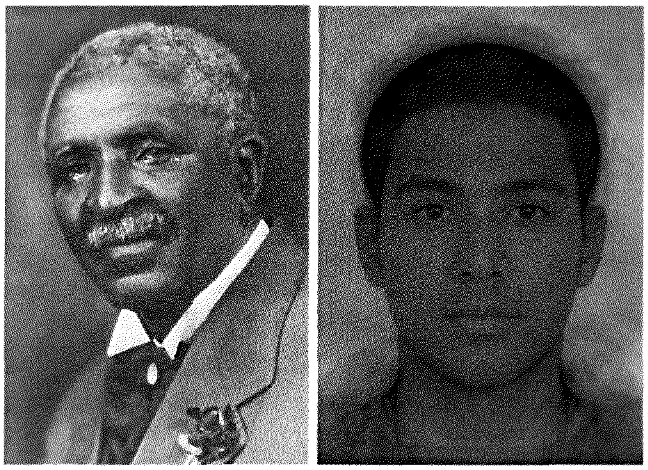


FIGURE 9.3

How long did it take you to realize that you do not recognize these faces?[2]

[2]George Washington Carver (American scientist, educator, and inventor) and average male face (FaceResearch.org).

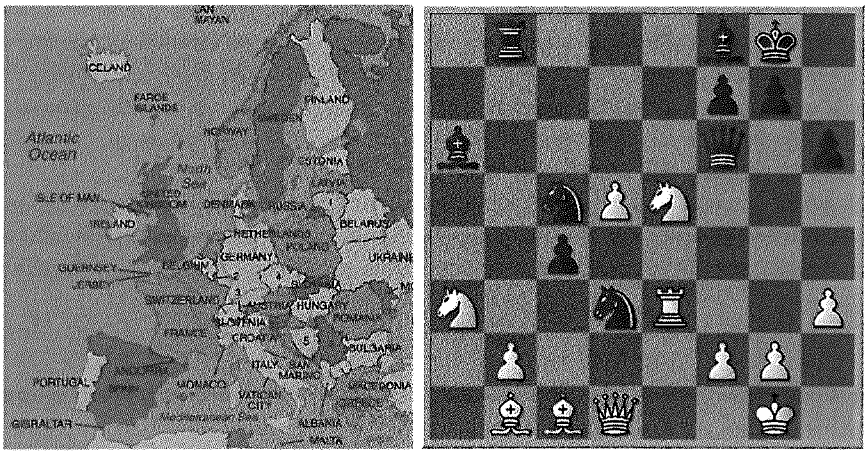


FIGURE 9.4

We can recognize complex patterns quickly.

A new face stimulates a pattern of neural activity that has not been activated before, so no sense of recognition results. Of course, a new face may be so similar to a face we have seen that it triggers a *misrecognition*, or it may be just similar enough that the neural pattern it activates triggers a familiar pattern, causing a feeling that the new face reminds us of someone we know.

An interesting aside is that face recognition is a special type of recognition: it has its own dedicated mechanisms in our brains, hardwired in by evolution; we do not have to learn to recognize human faces (Eagleman, 2012).

Similar mechanisms make our visual system fast at recognizing complex patterns, although unlike face recognition, they develop largely through experience rather than being wired in from birth. Anyone with at least a high school education quickly and easily recognizes a map of Europe and a chessboard (see Fig. 9.4). Chess masters who have studied chess history may even recognize the chess position as Kasparov versus Karpov 1986.

RECALL IS HARD

In contrast, *recall* is long-term memory reactivating old neural patterns without immediate similar perceptual input. That is much harder than reactivating a neural pattern with the same or similar perceptions. People *can* recall memories, so it obviously *is* possible for activity in other neural patterns or input from other areas of the brain to reactivate a pattern of neural activity corresponding to a memory. However, the coordination and timing required to recall a memory increase the likelihood that the wrong pattern or only a subset of the right pattern will be activated, resulting in a failure to recall.

Whatever the evolutionary reasons, our brain did not evolve to recall facts. Many schoolchildren dislike history class because it demands that they remember facts, such as the year the English Magna Carta was signed, the capital city of Argentina, and the names of all 50 U.S. states. Their dislike is not surprising; the human brain is not well suited for that sort of task.

Because people are bad at recall, they develop methods and technologies to help them remember facts and procedures (see Chapter 7). Orators in ancient Greece used the *method of loci* to memorize the main points of long speeches. They imagined a large building or plaza and mentally placed their talking points in spots around it. When presenting the speech, they mentally "walked" through the site, picking up their talking points as they passed.

Today we rely more on external recall aids than on internal methods. Modern-day speakers remember their talking points by writing them down on paper or displaying them in overhead slides or presentation software. Businesses keep track of how much money they have, owe, or are owed by keeping account books. To remember contact information of friends and relatives, we use address books. To remember appointments, birthdays, anniversaries, and other events, we use calendars and alarm clocks. Electronic calendars are best for remembering appointments, because they actively remind us; we don't have to remember to look at them.

RECOGNITION VERSUS RECALL: IMPLICATIONS FOR USER-INTERFACE DESIGN

The relative ease with which we recognize things rather than recall them is the basis of the graphical user interface (GUI) (Johnson et al., 1989). The GUI is based on two well-known user interface design rules:

- ***See and choose is easier than recall and type.*** Show users their options and let them choose among them, rather than force users to recall their options and tell the computer what they want. This rule is the reason GUIs have almost replaced command-line user interfaces (CLIs) in personal computers (see Fig. 9.5).

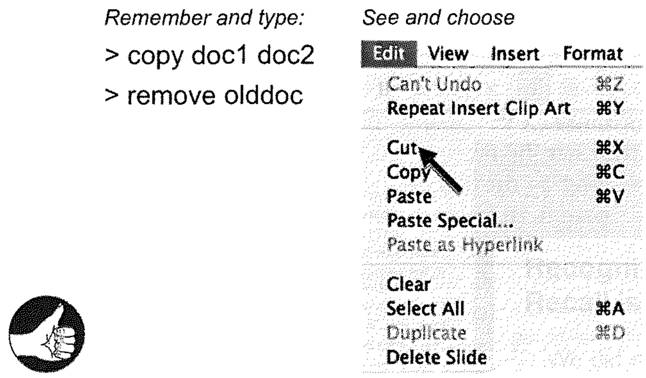


FIGURE 9.5

The main design rule behind today's GUI: "See and choose is easier than remember and type."

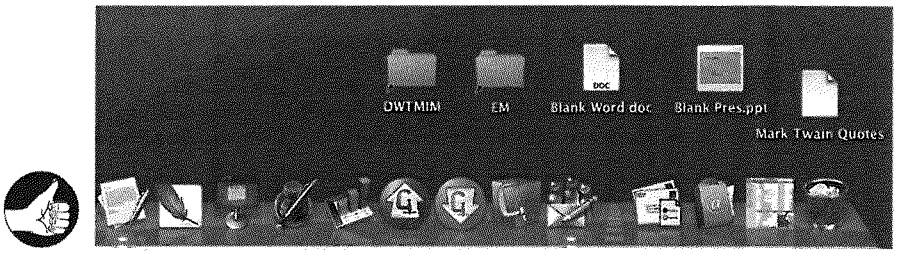


FIGURE 9.6

Desktop icons convey function via recognition—by analogy with physical objects or by experience.

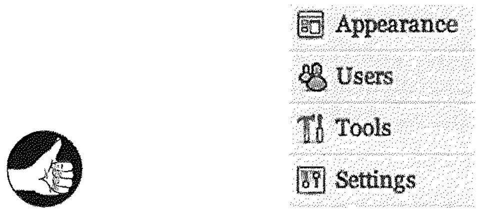


FIGURE 9.7

Wordpress.com uses symbols plus text to label functional pages on the Dashboard.

"Recognition rather than recall" is one of Nielsen and Molich's (1990) widely used heuristics for evaluating user interfaces. By contrast, using language to control a software application sometimes allows more expressiveness and efficiency than a GUI would. Thus, *recall and type* remains a useful approach, especially in cases where users can easily recall what to type, such as when entering target keywords into a search box.

- ***Use pictures where possible to convey function.*** People recognize pictures very quickly, which also stimulates the recall of associated information. For this reason, today's user interfaces often use pictures to convey function (see Figs. 9.6 and 9.7), such as desktop or toolbar icons, error symbols, and graphically depicted choices. Pictures that people recognize from the physical world are useful because they can be recognized without needing to be taught. This recognition is good as long as the familiar meaning matches the intended meaning in the computer system (Johnson, 1987). However, using familiar pictures from the physical world is not absolutely crucial. Computer users can learn to associate new icons and symbols with their intended meaning if these graphics are well designed. Memorable icons and symbols hint at their meaning, are distinguishable from others, and consistently mean the same thing, even across applications.

The GUI originated in the mid-1970s and became widespread in the 1980s and 1990s. Since then, additional design rules have arisen that are based on human

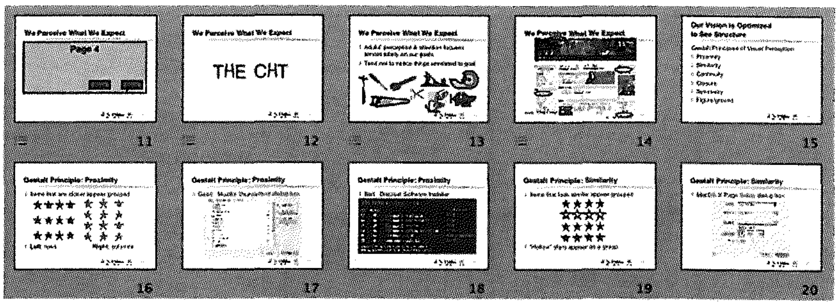


FIGURE 9.8

Microsoft PowerPoint can show slides as thumbnails, providing an overview based on recognition.

perception in general and on recognition and recall in particular. The following sections outline a few of these newer rules.

Use thumbnail images to depict full-sized images compactly

Recognition is fairly insensitive to the size in which objects and events are displayed. After all, we have to be able to recognize things independently of their distance from us. What is important are features: as long as most of the same features are present in the new picture that were in the original one, the new perception stimulates the same neural pattern, resulting in recognition.

Therefore, a great way to display pictures people have already seen is to present them as small thumbnail images. The more familiar a picture, the smaller the thumbnails of it can be and still be recognizable. Displaying small thumbnails instead of full-sized images allows people to see more of their options, their data, their history, etc., at once.

Photo management and presentation applications use thumbnail images to give users an overview of their images or slides (see Fig. 9.8). Web browsers use thumbnails to show pages a user has recently visited (see Fig. 9.9).

The larger the number of people who will use a function, the more visible the function should be

For the reasons described before, recall often fails. If a software application hides its functionality and requires its users to recall what to do, some percentage of users will fail. If the software has a lot of users, that percentage who fail to recall—even if it is small—adds up to a significant number. Software designers obviously don't want a significant number of users to fail in using their product.

The solution is to make functions that many people need highly visible, so users see and *recognize* their options rather than having to *recall* them. By contrast, functionality that few people will use—especially when those few people are highly trained—can be hidden, for example, behind "Details" panels, in right-click menus, or via special key combinations.

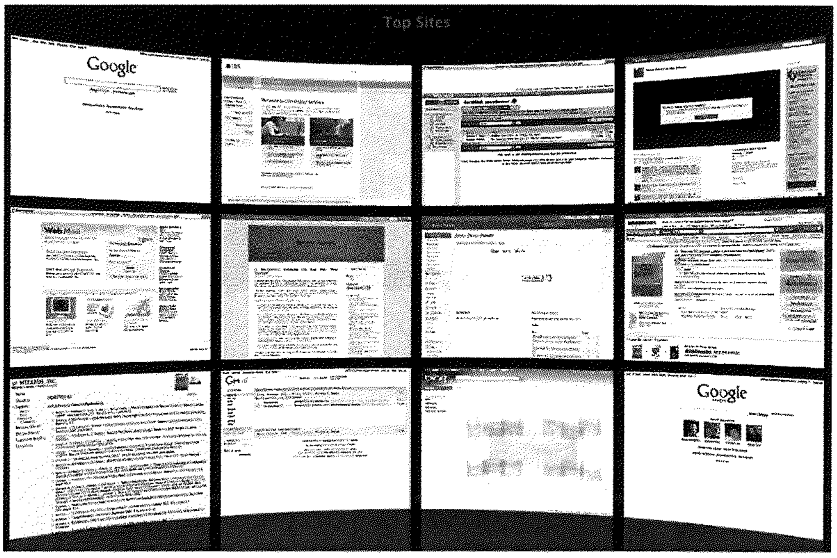


FIGURE 9.9

Apple Safari can show recently visited pages as thumbnail images, for quick recognition and choice.

Use visual cues to let users recognize where they are

Visual recognition is fast and reliable, so designers can use visual cues to show users instantly where they are. For example, it is a well-known Web design rule that all pages in a Web site should have a common distinctive visual style so people can easily tell whether they are still on the site or have gone to a different one. Slight but systematic variations on a site's visual style can show users which section of the site they are in.

Some desktop operating systems allow users to set up multiple desktops ("rooms" or "workspaces") as locations for different categories of work. Each has its own background graphic to allow easy recognition.

Some corporate Web sites use pictures to assure users that they are on a secure site. Users choose a picture as a personal account logo, and the site displays the logo whenever it recognizes the user from cookies or after the user has entered a valid login name but not yet a password (see Fig. 9.10). This lets users know they are at the real company site and not a fake site hosted by someone running a phishing scam.

Make authentication information easy to recall

People know that it is hard to recall arbitrary facts, words, and sequences of letters or digits. That is why they often write passwords and challenge-question answers down and keep the information in places that are easy to reach and thus insecure. Or they base passwords on their children's initials, their birthdates, their street

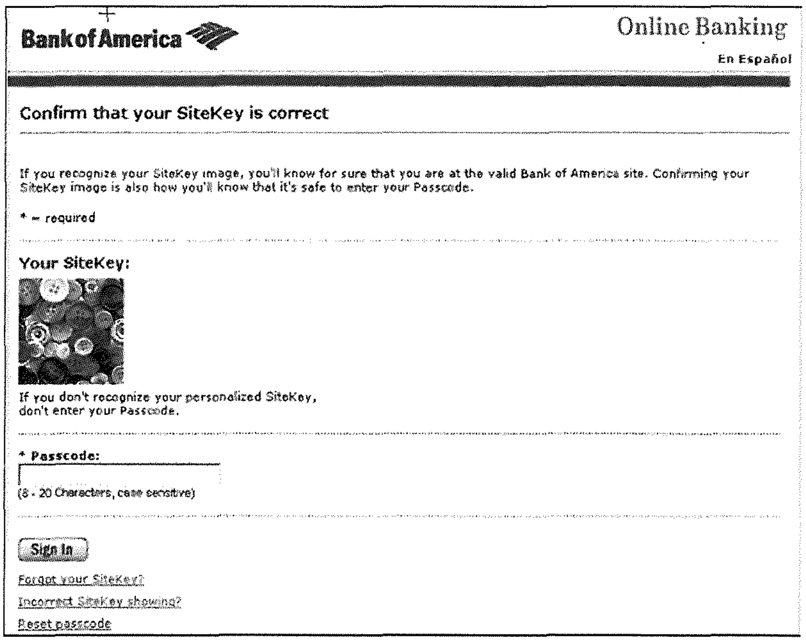


FIGURE 9.10

BankOfAmerica.com shows recognized customers their self-selected account logo (SiteKey) to assure them that it is the real bank's site.

address, and other information they know they can recall. Unfortunately, such passwords are too often easy for other people to guess (Schrage, 2005). How can designers help users avoid such unsafe behavior?

For starters, we can at least not make it hard for people to recall their login information, like the systems cited in Chapter 7 that impose burdensome password restrictions or offer a limited choice of challenge questions.

Instead, we can give users the freedom to create passwords they can remember and challenge questions for which they can remember the correct response. We can also let users supply password *hints* that the system can present to them, under the assumption that users can devise hints that will serve as a recall probe for them but not identify the password to third parties.

Authentication methods that do not rely on users to recall the authentication data would seem to be a solution. Biometric authentication methods, such as iris scans, digital fingerprint scans, and voice identification, fall into this category. However, many people regard these methods as privacy threats because they require the collection and storage of individuals' biometric data, creating the potential for information leaks and abuse. Therefore, while biometric authentication does not burden users' memory, it would have to be implemented in a way that meets stringent privacy requirements to be widely accepted.

This page intentionally left blank

Welcome

Perception: Audition

Abbas Moallem, Ph.D.

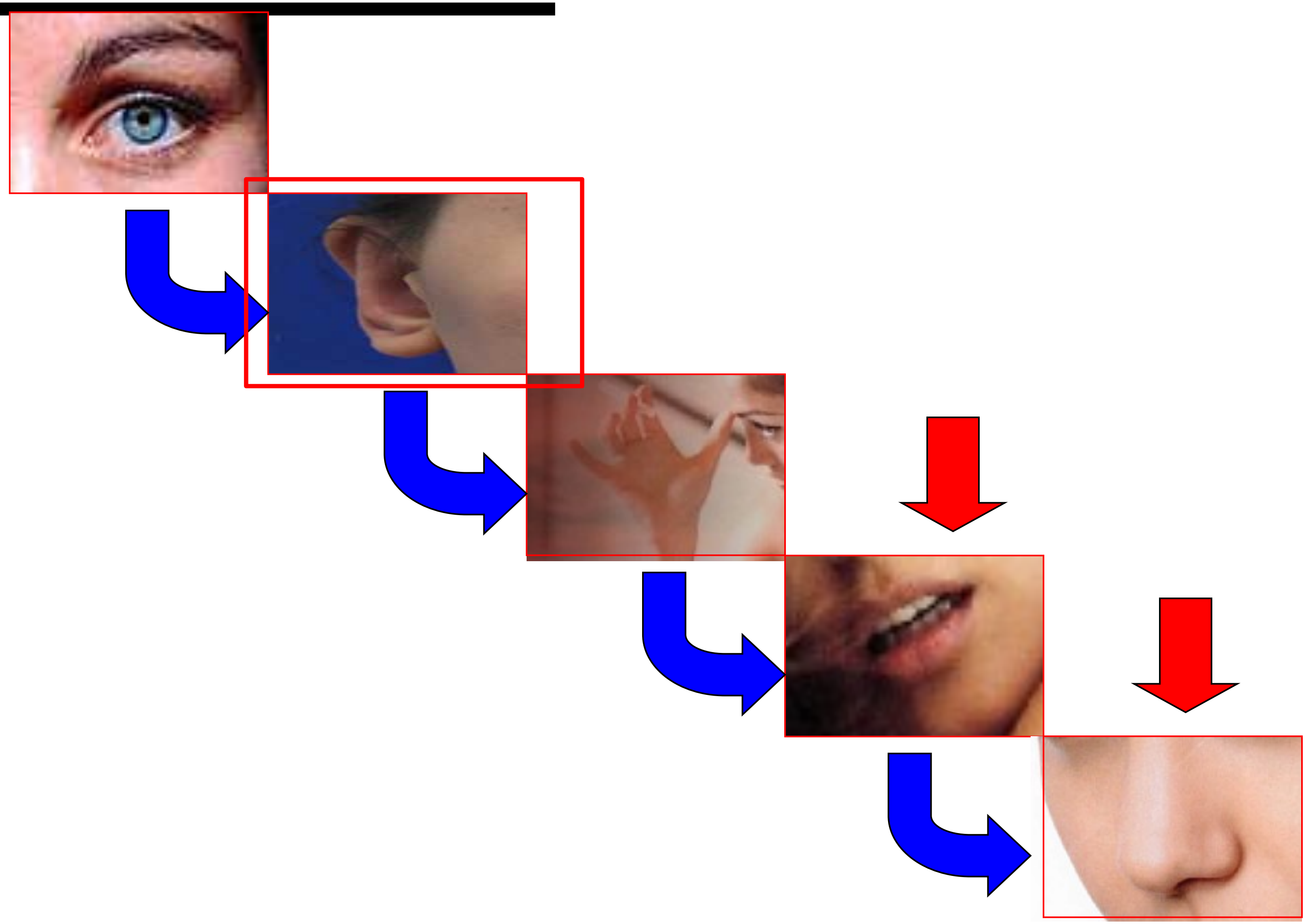
The only true wisdom is in knowing you know nothing.

Socrates

Overview

- Introduction
- Ears
- Audition
- Hearing Function

Input-Output Channels



Sound in the Interface

- The vast majority of computer-based user interfaces on the visual medium.
- Another medium that is frequently used in interface design is sound
- In this presentation we will review human auditive system.

Hearing and Human Ear

Sound

- Any sudden mechanical movement sets up fluctuation in air.
- Pressure which spread our as waves, just like water in stirred produce sound.
- The frequency of a sound is the number of fluctuations or vibration per second, expressed in Hertz (HZ)
- Intensity of sound is measured by logarithmic unit called decibel (dB)

Hearing

- **Provides information about environment:**
  distances, directions, objects etc.
- **Different Parts of Ears:**
  - **outer ear – protects inner and amplifies sound**
  - **middle ear– transmits sound waves as vibrations to inner ear**
  - **inner ear – chemical transmitters are released and cause impulses in auditory nerve**
- **Sound**
  - **pitch – sound frequency**
  - **loudness – amplitude**
  - **timbre – type or quality**

Terminology

- Frequency
  - In audio, the number of repeating cycles of change in air pressure or oscillations in voltage, that occur in one unit of time usually a second. Complex sounds are made up of many pure tones of different frequencies
- Wavelength
  - The wavelength of sound is the distance between analogous points of two successive waves.
- Hertz
  - The SI unit of frequency, in particular the number of times something occurs in one second, abbreviated Hz. Named after Heinrich Rudolph Hertz (1857-94). The unit is sometimes alternatively expressed as CPS.
- Decibel
  - A ratio based measure of the comparative amount of some quality, usually sound level, power or voltage, relative to some reference amount.

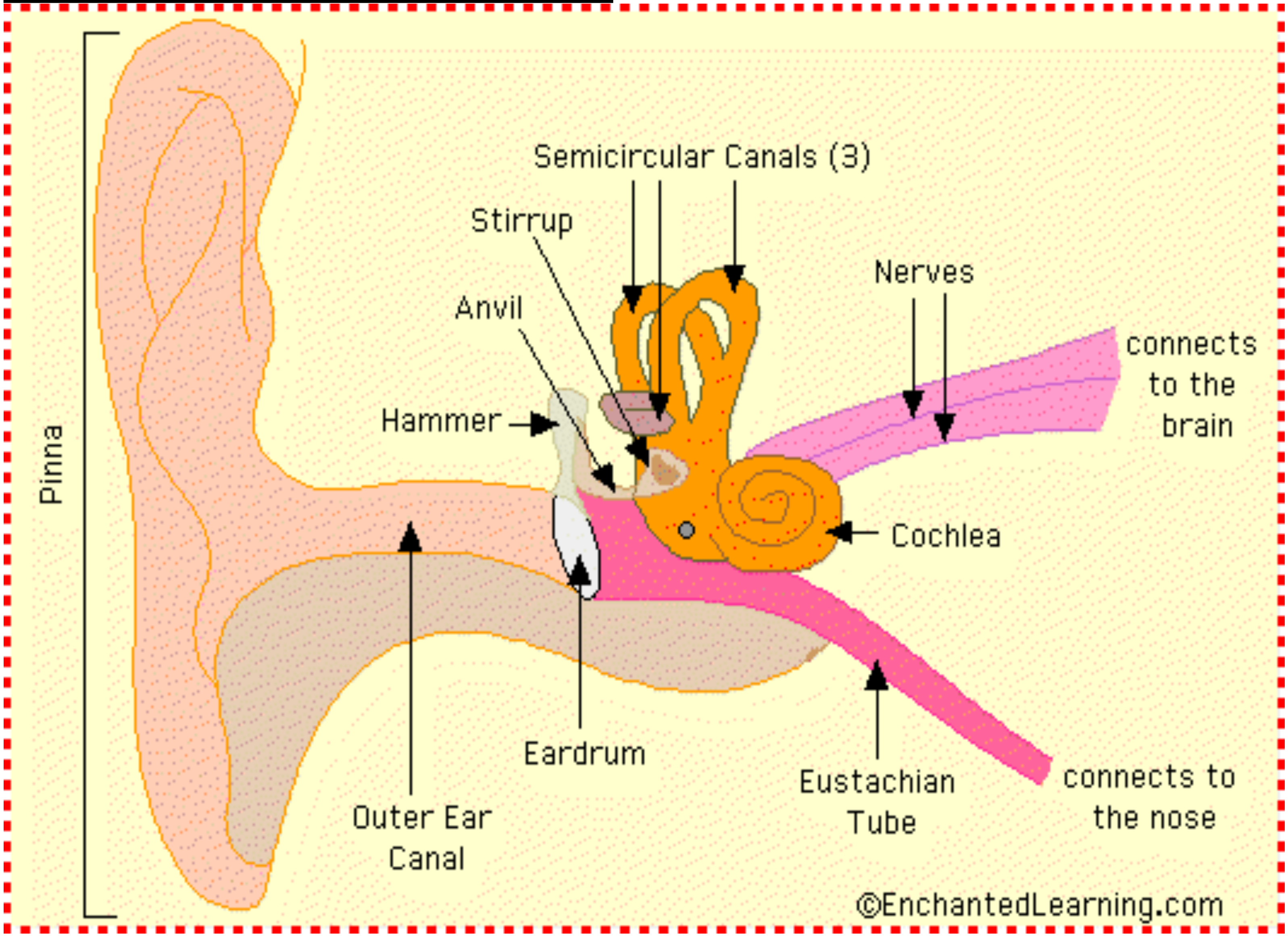
Frequency

- Human ear sensitive to sounds in the frequency range of 16 to 20,000 Hertz
- Below 16 Hz Infrasonic
- Above 20,000 Ultrasonic
- Greatest sensitivity lies in the range of 2000 –5000 Hertz

What is a decibel?

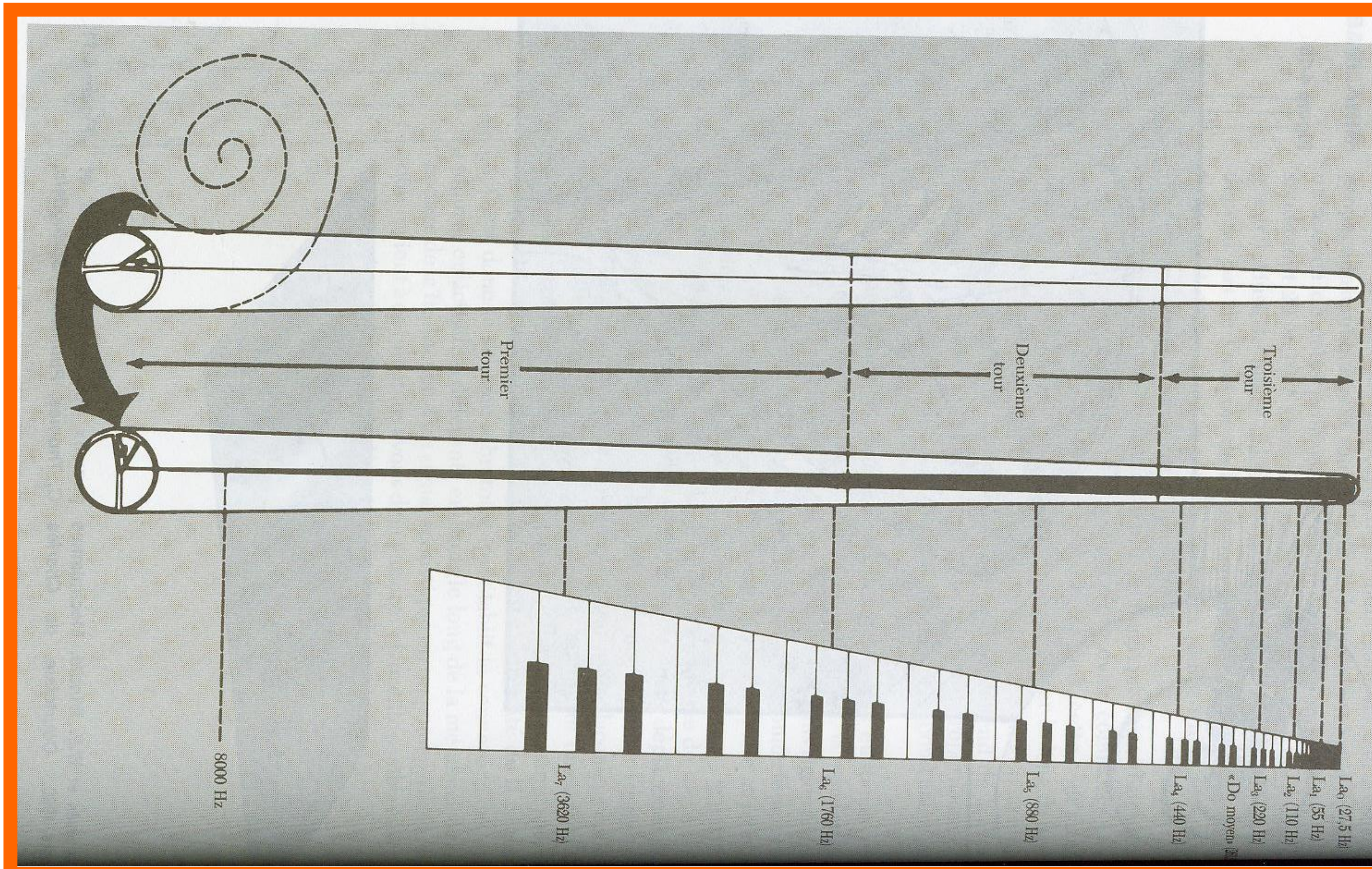
- The decibel (dB) is one tenth of a Bel, which is a unit of measure that was developed by engineers at Bell Telephone Laboratories and named for Alexander Graham Bell.
- The dB is a logarithmic unit that describes a ratio of two measurements.

Human Ear



Pinna
Semicircular Canals (3)
Stirrup
Anvil
Hammer
Nerves
connects to the brain
Cochlea
Eardrum
Outer Ear Canal
Eustachian Tube
connects to the nose
©EnchantedLearning.com

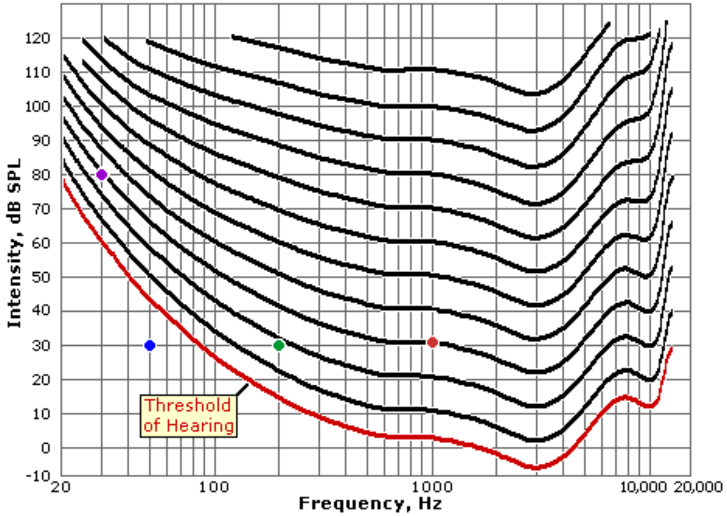
Cochlea



Premier tour
Deuxième tour
Troisième tour
8000 Hz
La_7 (3620 Hz)
La_6 (1760 Hz)
La_5 (880 Hz)
La_4 (440 Hz)
«Do moyen»
La_3 (220 Hz)
La_2 (110 Hz)
La_1 (55 Hz)
La_0 (27,5 Hz)

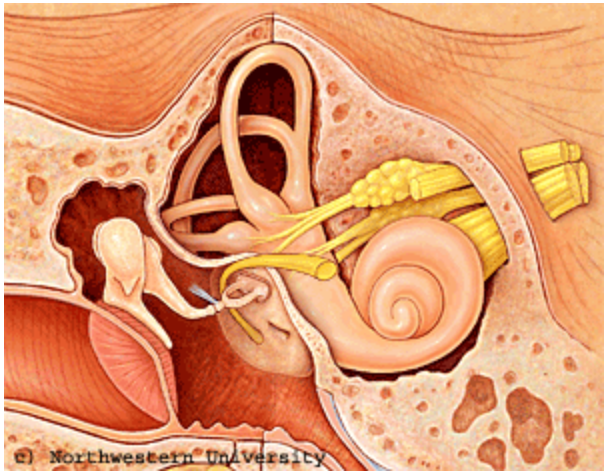
Hearing

- **Humans can hear frequencies from 20Hz to 15kHz**
  - **less accurate distinguishing high frequencies than low.**
- **Auditory system filters sounds**
- **Human ear can distinguish frequency changes of less than 1.5 Hz at low frequency but less accurate at high frequency.**



Processing Sound

- **Hearing Frequencies**
  - **20Hz to 15Hz**
- **Human ear can distinguish frequency changes of less than 1.5 Hz at low frequency but less accurate at high frequency.**



Cocktail Party Phenomenon

Can attend to sounds over background noise?

Hearing Functions

- Convey specific information, as basis for communication between individuals.
- As an alarm system: by activating secondary pathways leading to the brain it plays an essential part in waking increased alertness, and finally Alarm
- Understanding of Speech

Sound in User Interface

- Attention: to attract the user's attention to the critical situation or to the end of process.
- Status information: continuous background sounds can be used to convey status information.
- Confirmation: A sound associated with an action to confirm that the action has been carried out.
- Navigation: using changing sound where the user is in a system.

Sound In User Interface

- Audio Alert
- Information Source
- System Output
- System Input
- Voice Interface

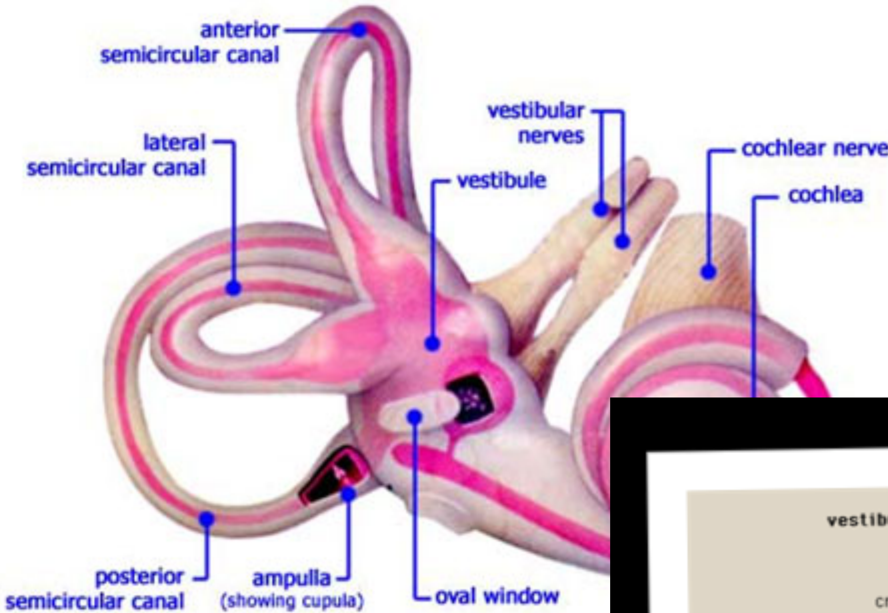
Alarms

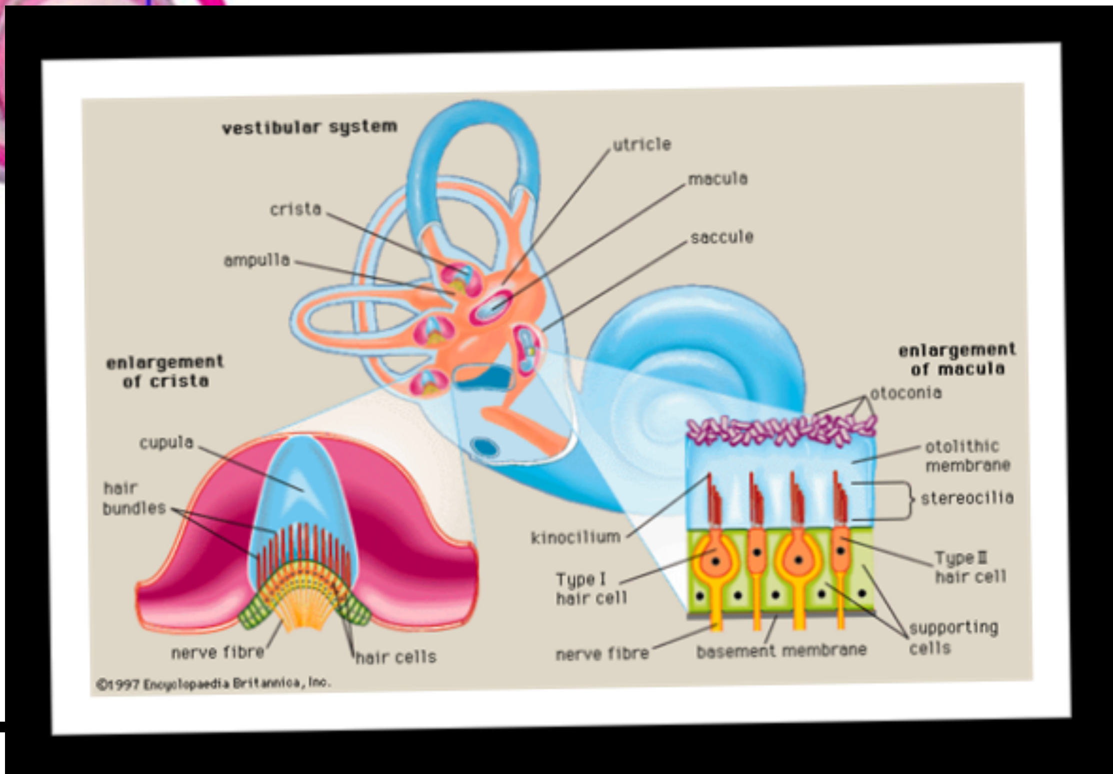
- **Voice Alarms and Meaningful Sounds**
- **The alarm must be**
  - **Heard**
  - **Not be above the danger level**
  - **Not be overlay starling or abrupt**
  - **Not disrupt perceptual understanding of the other signals**
  - **Be informative, signaling the nature of emergency**

Steps in Designing Alarms

- Environmental Factors
- Guarantee Informativeness and minimize confusion
- Alarm sequence
  - Identification
  - Diminishment

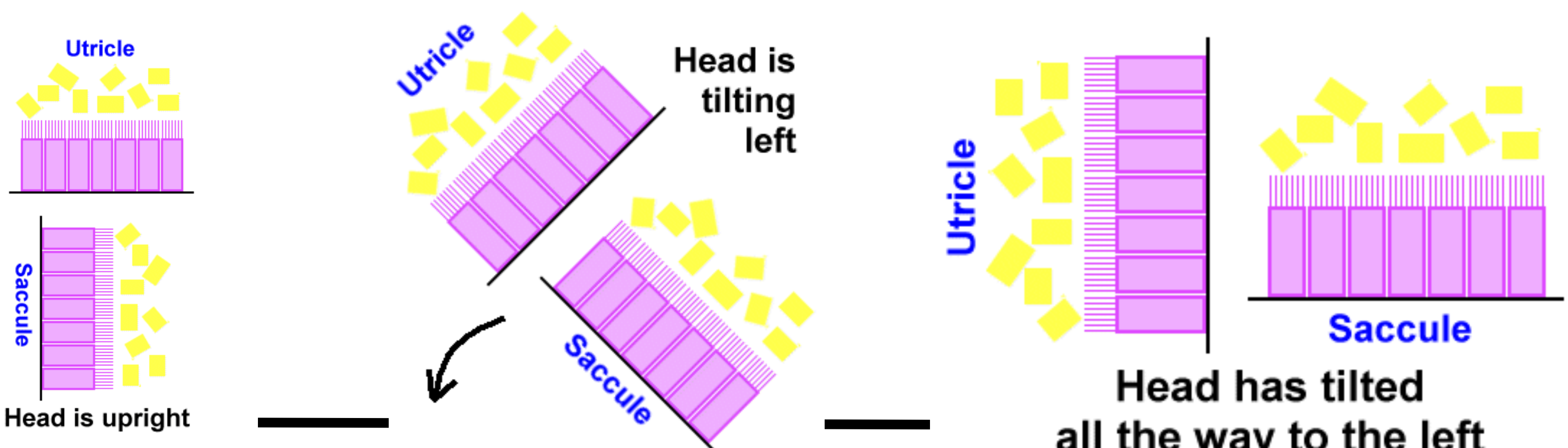
Vestibular Senses





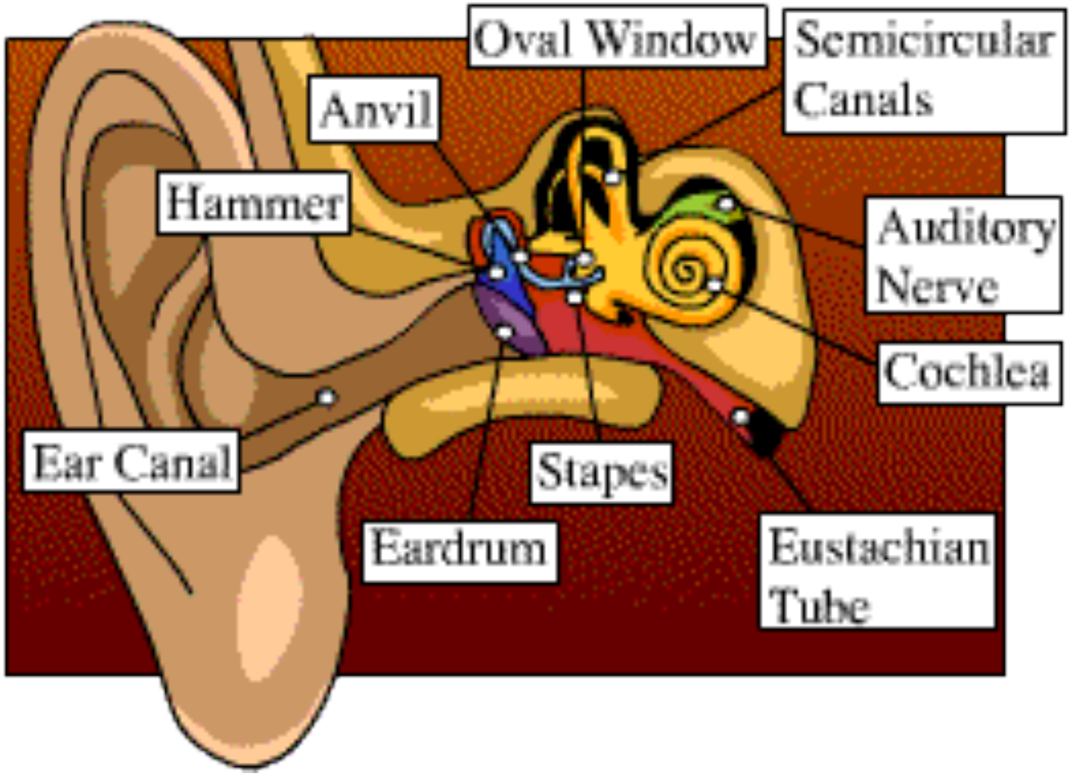
Vestibular Senses

- Awareness of body balance and movement
- Body rotation and of gravitation and movement.
- Static Position & Our Vestibule
- The vestibule contains two sensory areas: the utricle and the saccule.

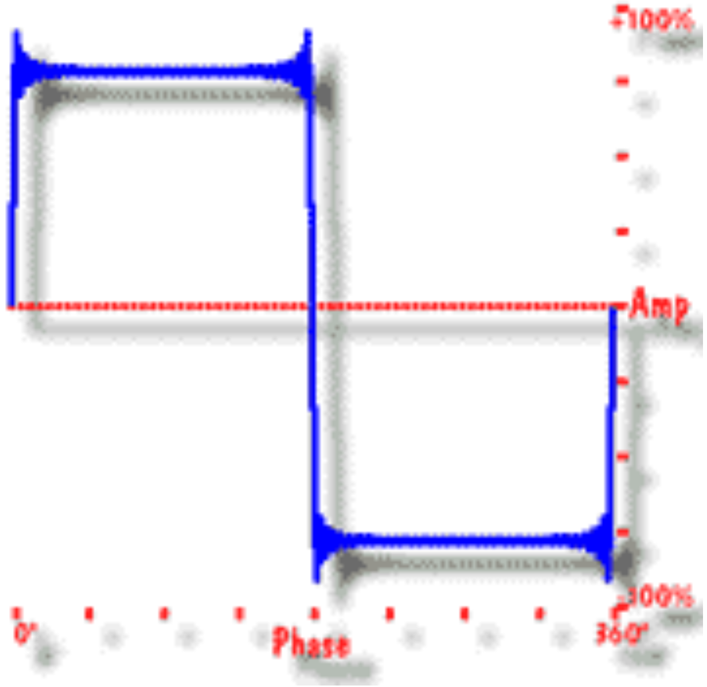


Thank You For Your Participation

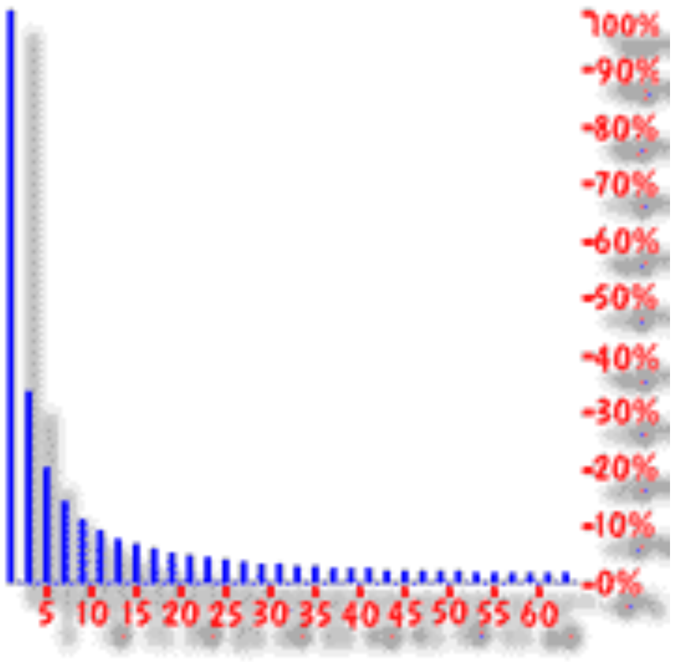
Human Ear


Oval Window
Semicircular Canals
Anvil
Hammer
Auditory Nerve
Cochlea
Ear Canal
Stapes
Eardrum
Eustachian Tube

Wave



A geometrical waveform typically generated by an oscillator.



Harmonic Structure

Sound Transmission

- Speech Signal
- Mask Effect
- Speech Communication
- Speech Distortions

Speech Distortion

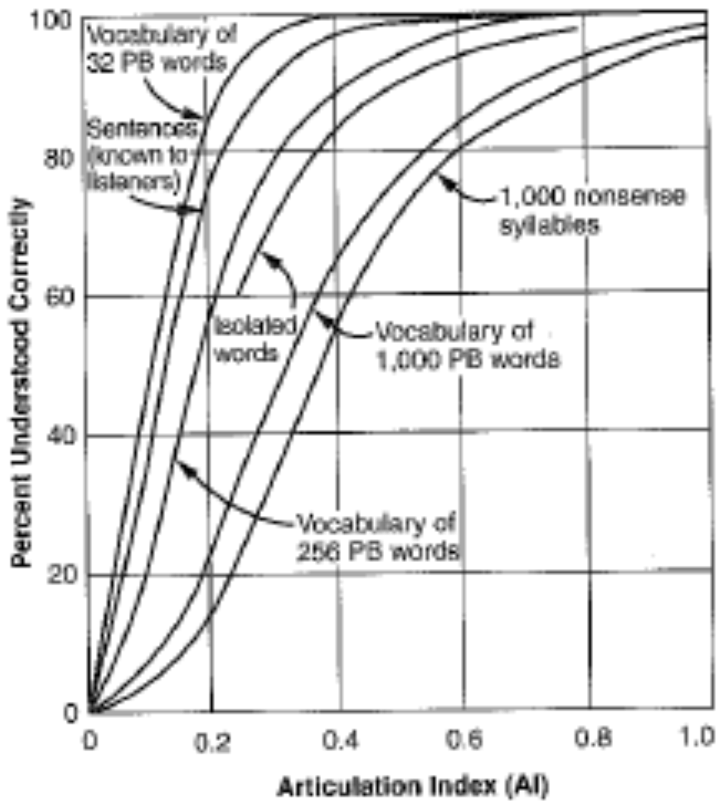


FIGURE 5.9

Relationship between the AI and the intelligibility of various types of speech test materials. Note that at any given AI, a greater percentage of items can be understood if the vocabulary is smaller or if the word strings form coherent sentences. (*Source:* Adapted from Kryter, K., 1972. Speech Communications. In *Human Engineering Guide to System Design*, H. P. Van Cott and R. G. Kinkade, eds., Washington, DC: U.S. Government Printing Office.)

Audio Frequency

- **Audio Frequency**
  - **The range of frequencies which can be experienced by an average human being. The range is defined as 20 Hz to 20 kHz for convenience but in practice, is realistically closer to 20 Hz to 20.000 kHz. Dolphins are believed to hear up to 70 kHz.**

  - **For convenience, the human frequency range is divided into three rough areas or bands. High frequencies (between about 5 kHz and 20 kHz), mid frequencies (between about 200 Hz and 5 kHz) and low frequencies (between about 20 Hz and 200 Hz**

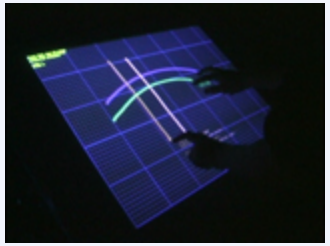
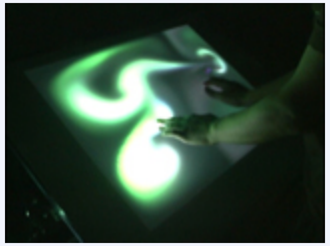
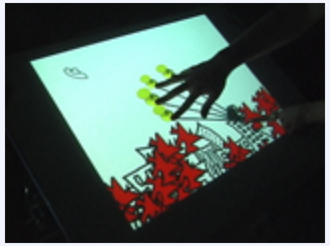
Welcome

- Touch and Movements
- Fitts Law

Touch and Movements

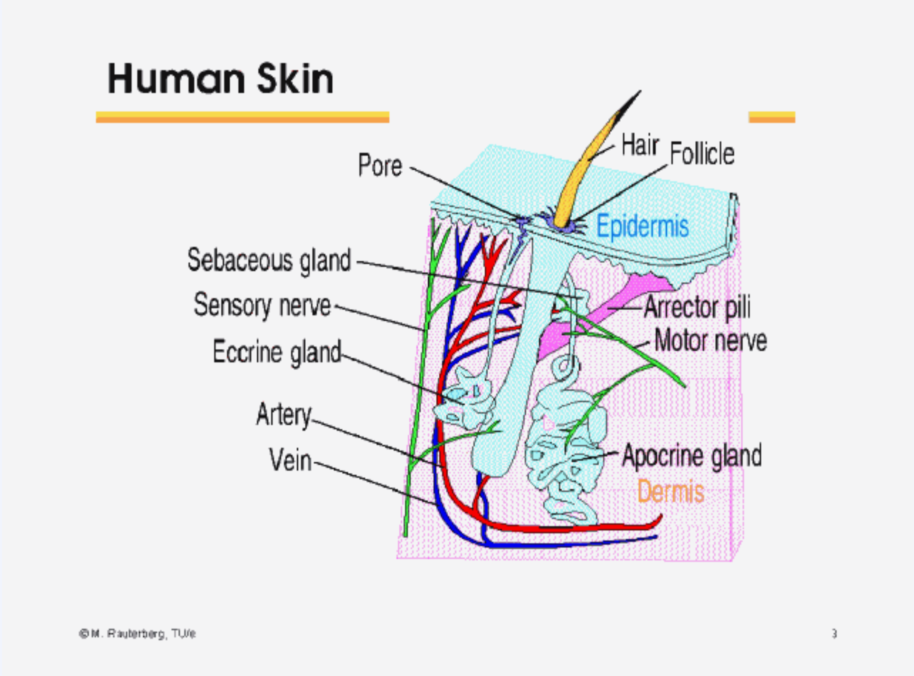
Touch

- Provides important feedback about environment.
- May be key sense for someone who is visually impaired.
- Stimulus received via receptors in the skin:
  - thermoreceptors – heat and cold
  - nociceptors – pain
  - mechanoreceptors – pressure (some instant, some continuous)
- Some areas more sensitive than others e.g. fingers.
- Kinethesis - awareness of body position
  - affects comfort and performance.


Tactile Perception

- Thermo receptors
- Nociceptors
- Mechanoreceptors
  - Rapidly Adapting Mechanoreceptors
  - Slowly Adapting Mechanoreceptors

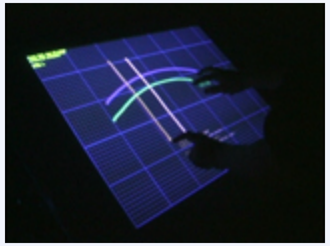


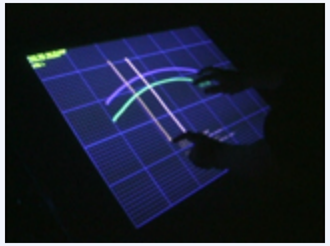
Receptors

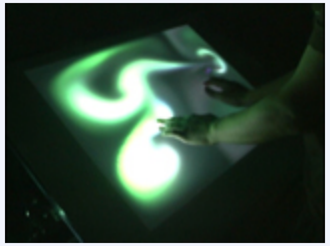
Three Types of Sensations

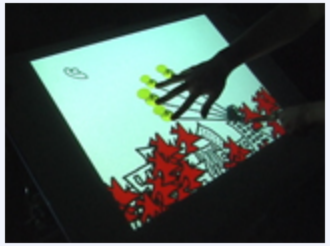
- Pressure
  - light
  - deep
- Temperature
  - cold
  - warm (not hot)
- Pain
  - sharp
  - dull

Touch and User Interface

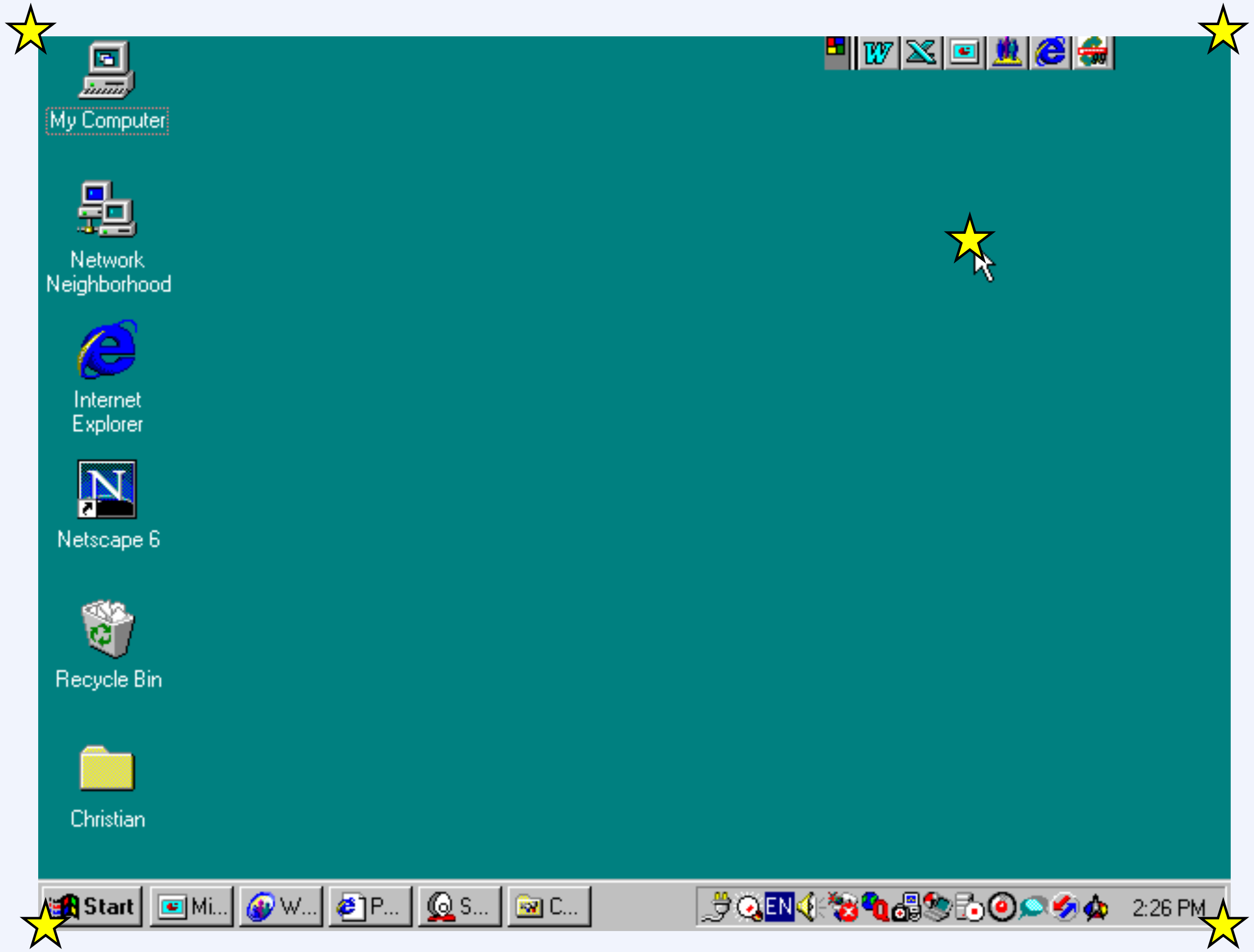








5 Easiest Pixels to Hit?



Movements

- Movement Time
- Reaction Time
  - Varies according Sensory Channel
    - Reaction to
      - an auditory signal 150 ms
      - Visual Signal 200 ms
      - Pain 700 ms
- Accuracy
- Aimed Movement Speed-Accuracy
- Fitt's Law

Fitts' Law

- **What five pixels (a point on the screen) are the easiest to precisely click with the mouse?**
- **Fitts' Law Says:**
  - **An object is easier to precisely click with the mouse if it is closer and larger**
- **Application:**
  - **Make targets larger and closer**

Movements

- Movement Time
- Reaction Time
  - Varies according Sensory Channel
    - Reaction to
      - an auditory signal 150 ms
      - Visual Signal 200 ms
      - Pain 700 ms
- Accuracy
- Aimed Movement Speed-Accuracy
- Fitt’s Law

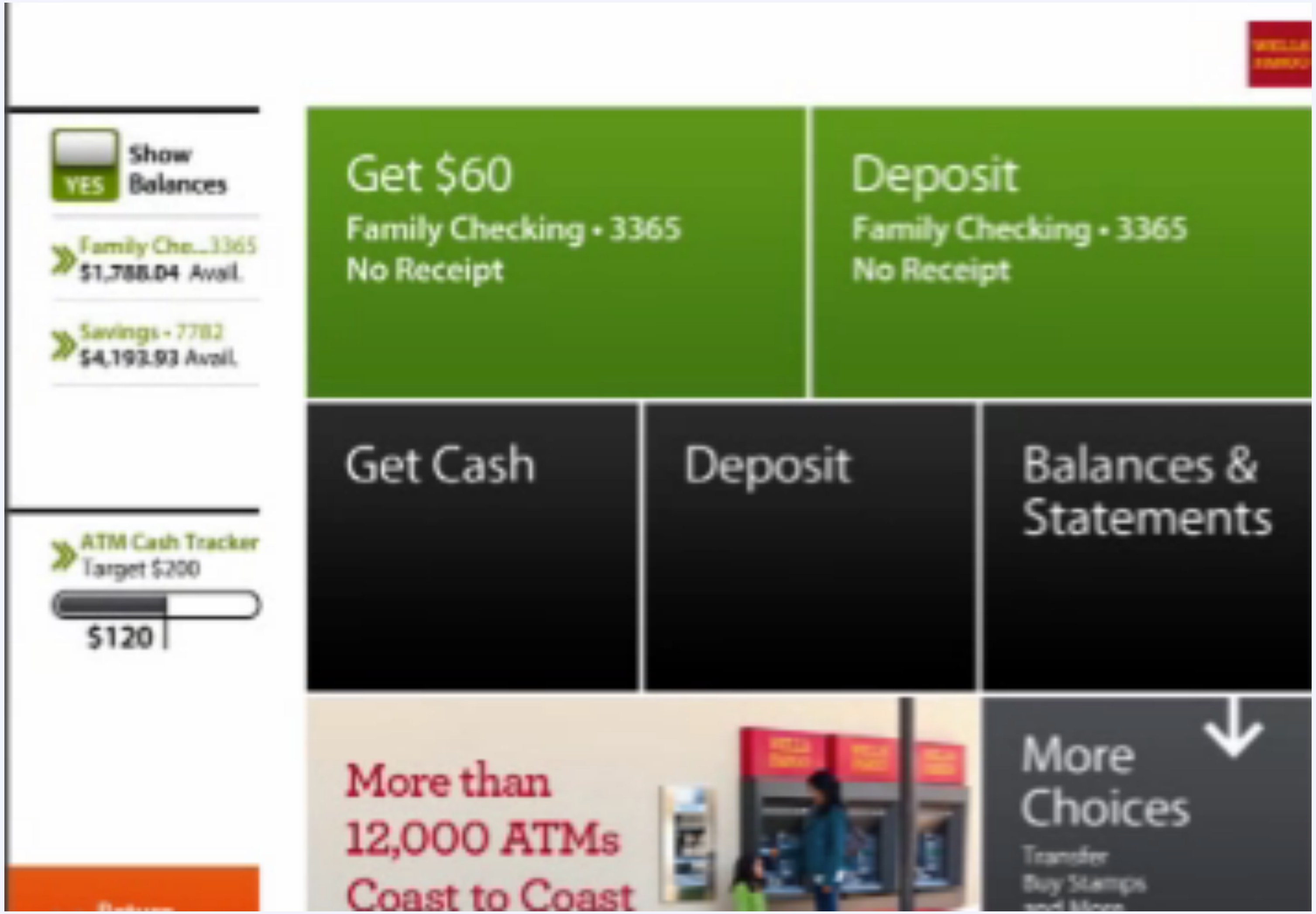
What is Fitts's Law?

- Fitts' Law is a model of human motor response developed by Paul Fitts in 1954.
- The model is based on time and distance.
- Helps predict human movement and human motion based on rapid, aimed movement, not drawing or writing.

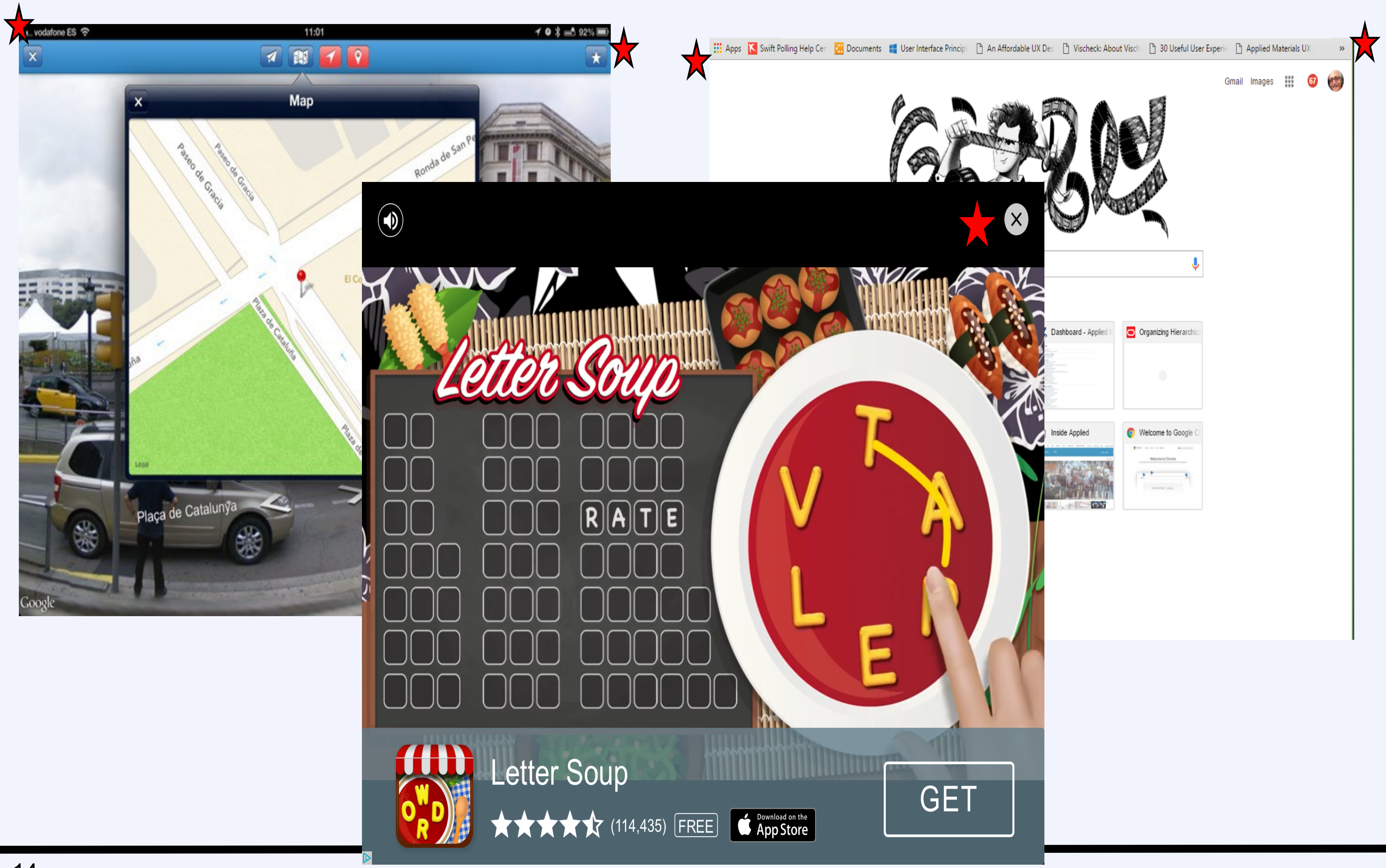
What is Fitts' Law?

- Fitts discovered that movement time was a logarithmic function of distance when target size was held constant, and that movement time was also a logarithmic function of target size when distance was held constant. Mathematically, Fitts' law is stated as follows:
- MT = a + b log2(2A/W)
- where
  - MT = movement time
  - a,b = regression coefficients
  - A = distance of movement from start to target center
  - W = width of the target

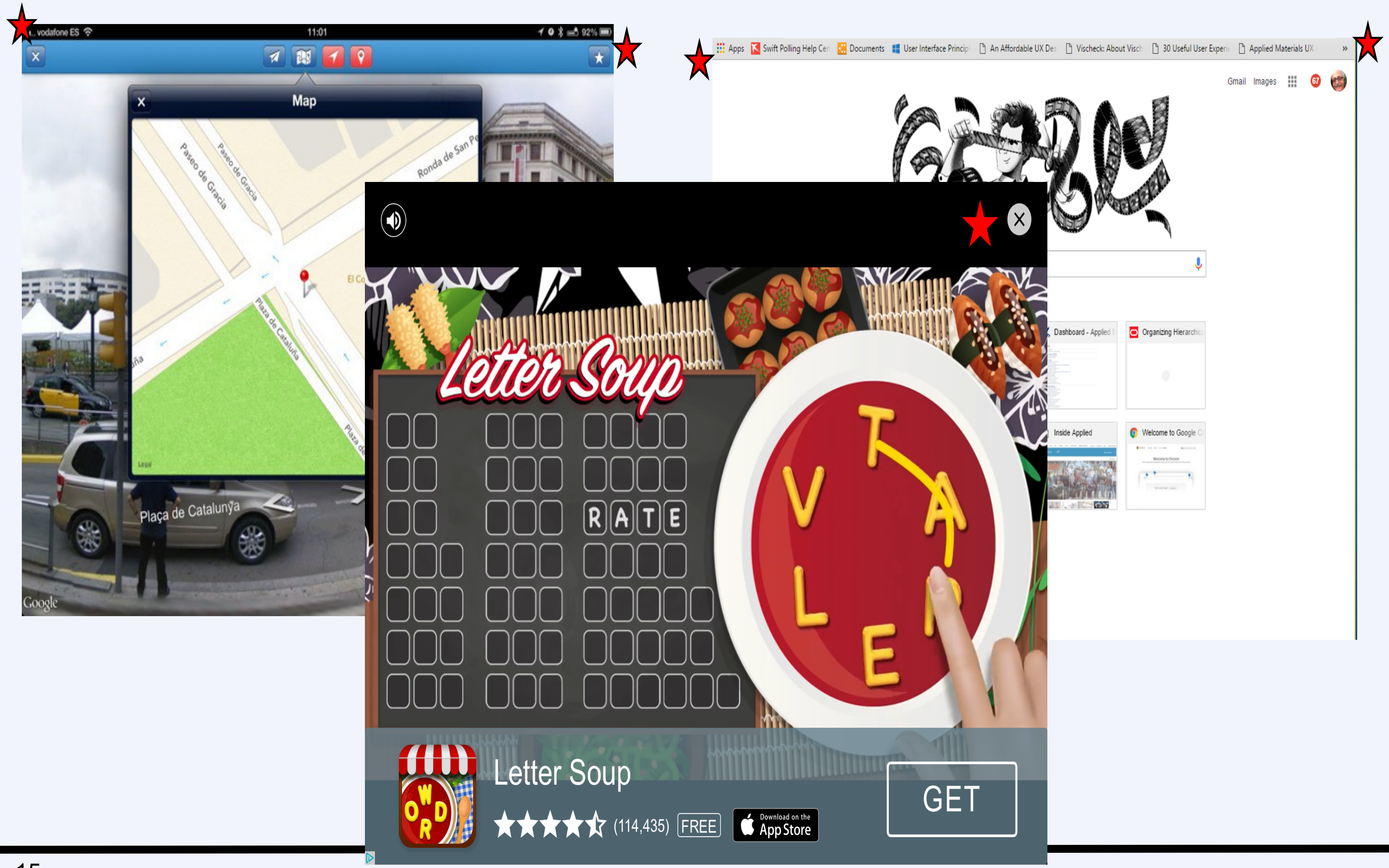
Example of Fitts's Law Application in User Interface



Corners are heavily used on screen design



Corners are heavily used on screen design



Questions


See You Next Week
Thank You For Your Participation

Welcome

- **Human Information Processing**

Human Information Processing

Abbas Moallem, Ph.D.

Information Processing

- Perception
- Information Processing
  - Memory
  - Representation,
  - Reasoning

Information Processing

- The information processing theory study cognitive development of human being.
- Information processing includes attention, working memory, and long term memory
- This theory addresses how as children grow, their brains likewise mature, leading to advances in their ability to process and respond to the information they received through their senses.

Information Processing & HCI

- Information processing helps to better how our cognitive activity process information.

- Predicts the cognitive processes used when a user interacts with a computer.

- Evaluate how long will take to carry out a task a user based on each cognitive approach.

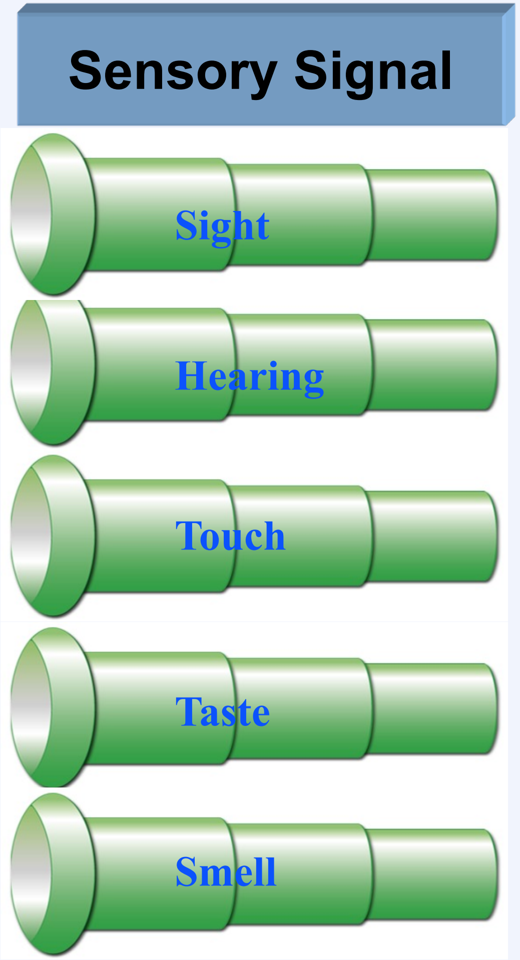
Perception

How information is acquired from the world and transformed into experiences

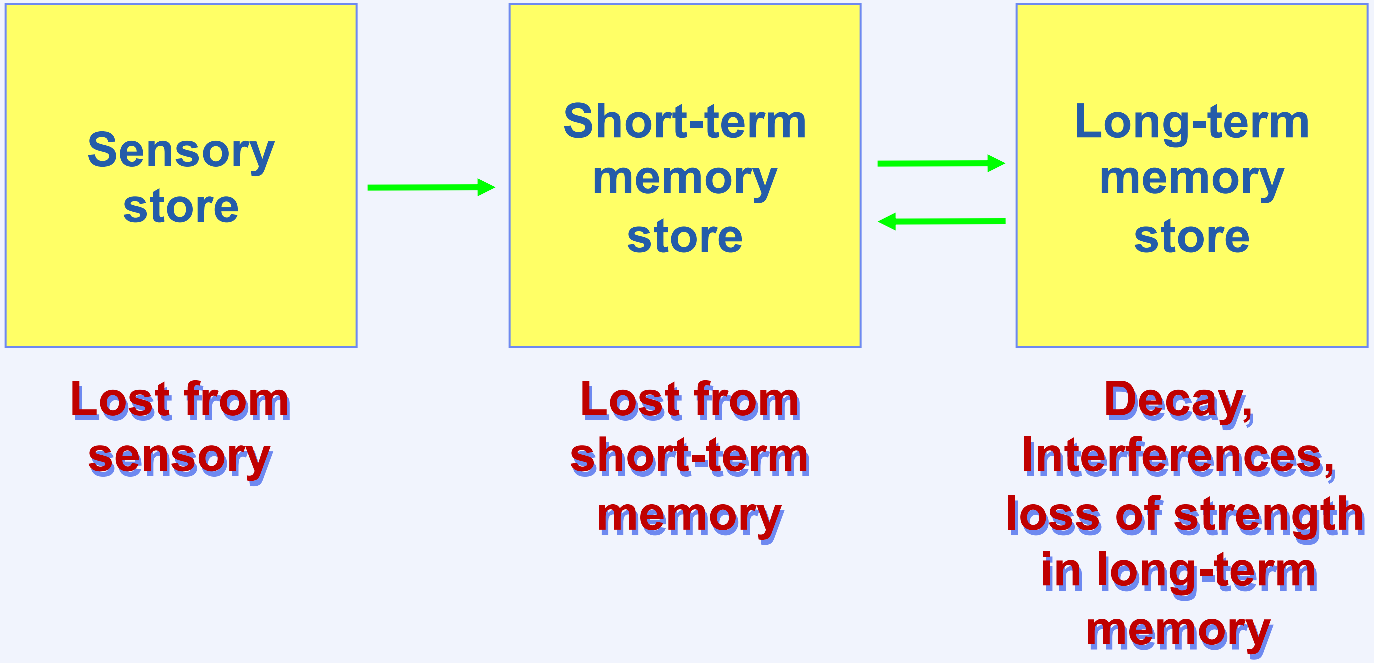
Stimulus Identification

- **Stimulus Contrast**
  - As stimulus contrast, or intensity increases, reaction time decreases until reaching and asymptote.

- **Feature Extraction**
  - Stimulus quality, word recognition
  - Stimulus discriminability affect the feature extraction process.

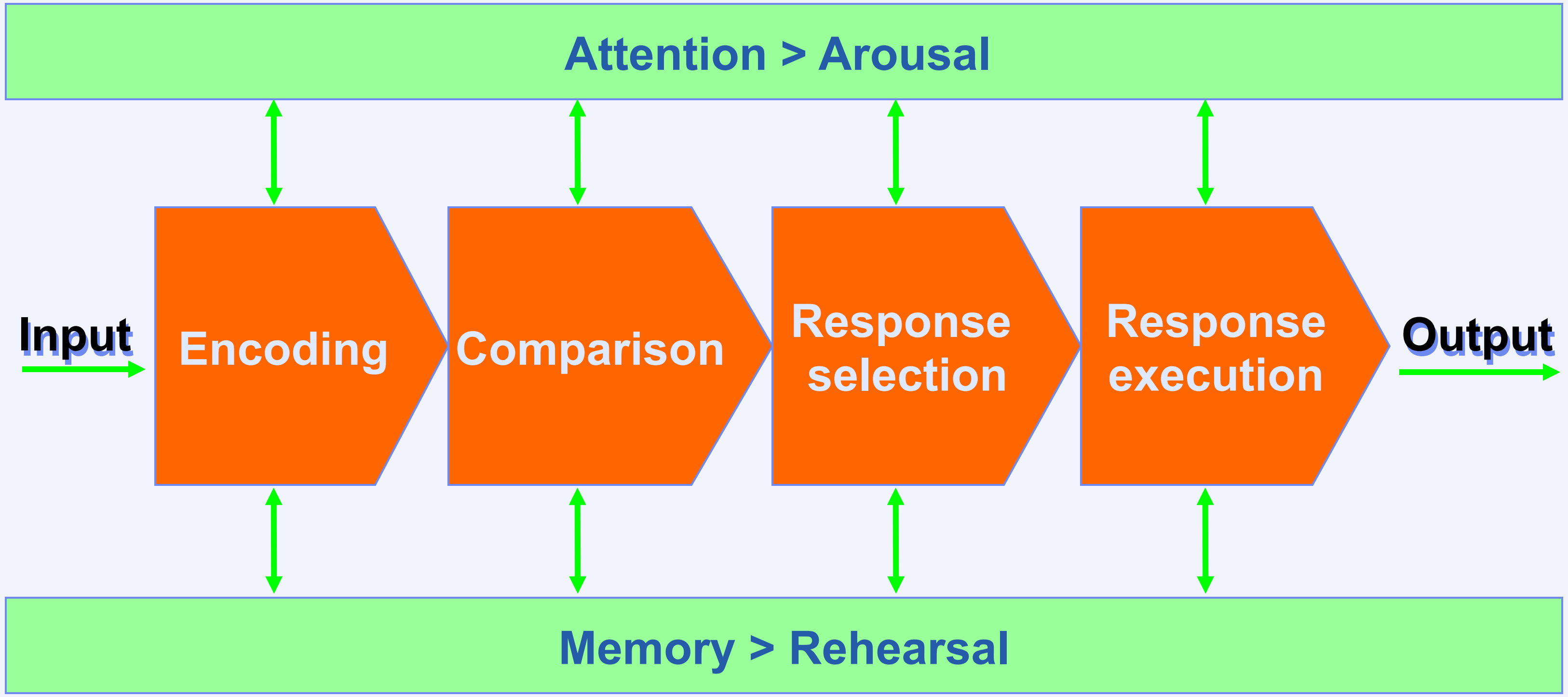


Multi-store Model of Memory



The Multi-store Model of Memory
(Atkinson and Shiffrin, 1968)

Human Information Processing

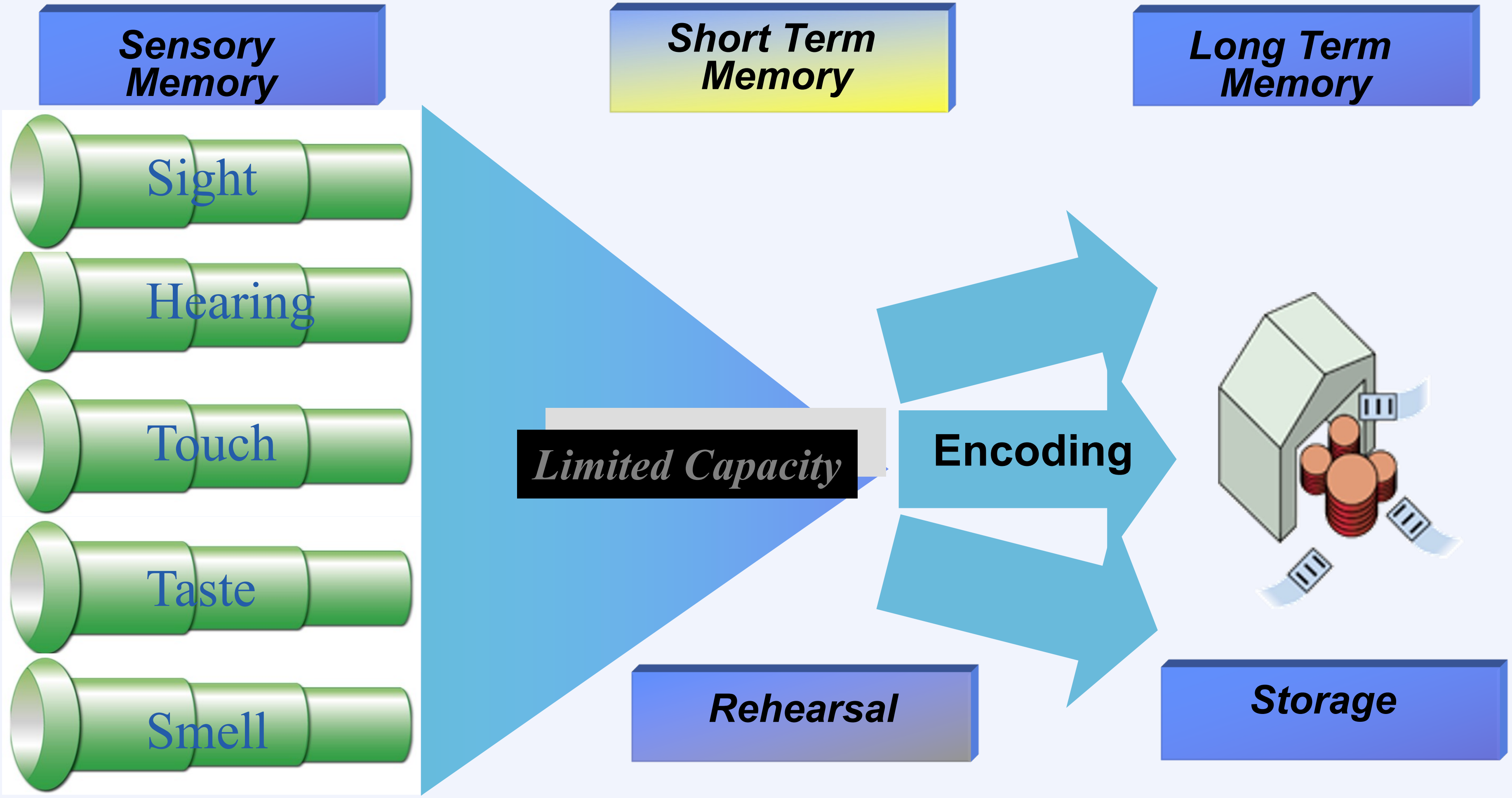


Extended Stages of Information Processing
(Barber, 1988)

Sensory Memory

- Sensory receptors (eyes, ears, skin…) receive information and stored in sensory memory for a very short period of time while transferring to short-term memory.
- Several types of sensory memories:
  - Iconic Memory: Visual,
    - Stored for a very brief period of time <1000 ms
  - Echoic Memory: Aural,
    - Stored for slightly longer periods of time than iconic memories
  - Haptic Memory: Tactile
    - Also stored for short periods of time and decay after approximately two seconds

Memory and Attention



Human Information Processing Stage (Leahey, 1989)

Exercise

- **Read the following string of digits.**

7 9 1 4 0

2 6 5 8 3 1 4 7 0 5 3

Examples

212348278493202

0121 414 2626

HEC ATR ANU PTH ETR EET

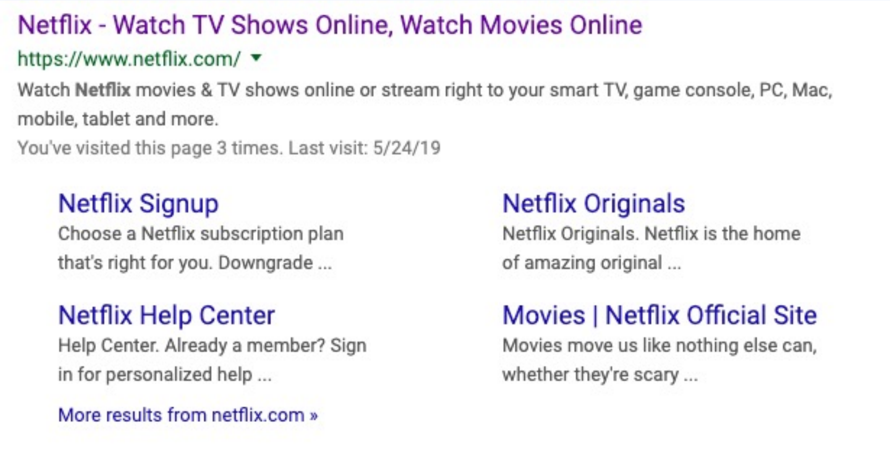
Application of Miller Magic Number

- Applications in Interaction Design
- Chucking Numbers

16047559385	**Hard to remember**
1 604 755 9385	The breaking down of the number into more “logical” chunks makes the number easier to remember.

Application of Miller Magic Number

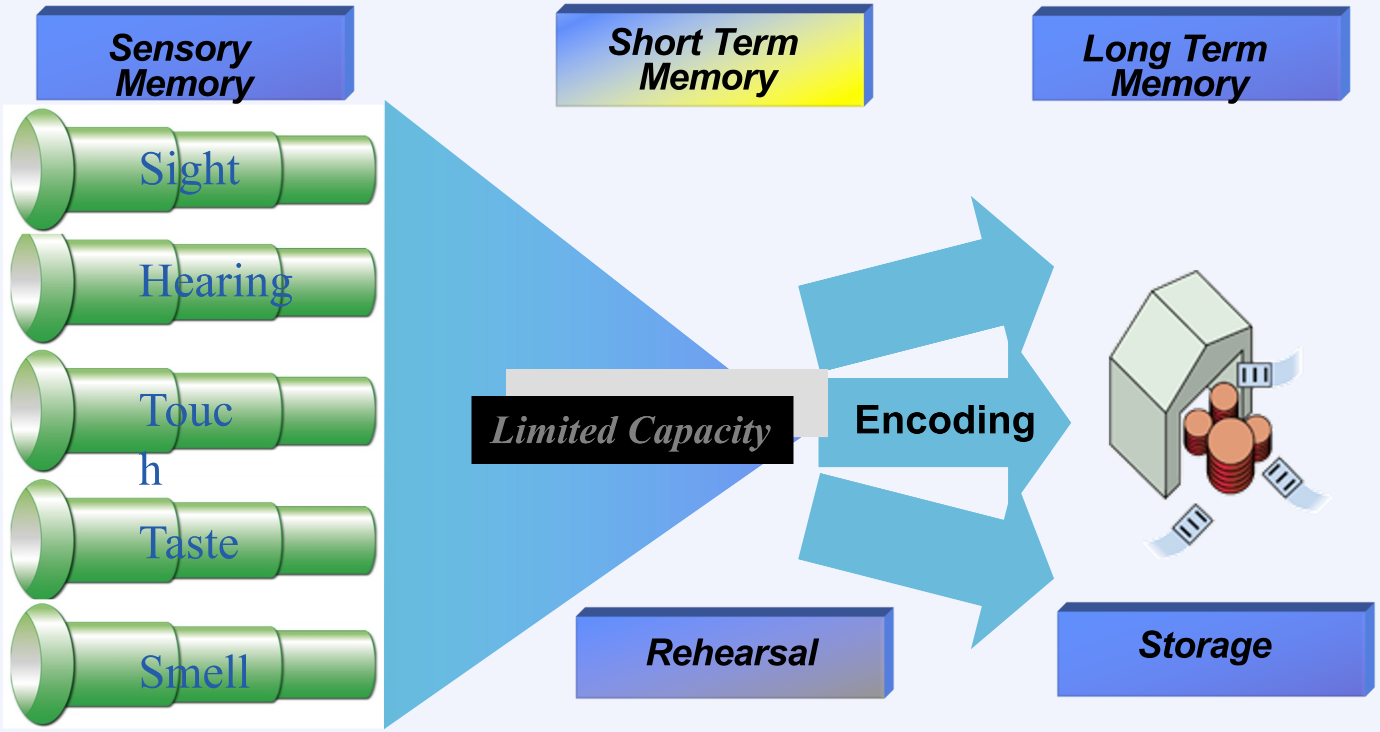
- **Chunking Text Content**



Example


Z0QX0007L
(1P) Part No. Z0QX0007L
LL
(S) Serial No. C02NQCC6FY17
C02 NQC C6F Y17
Model No: A1419
Assembled In China
Other Items as Marked Thereon
© TechTerms.com

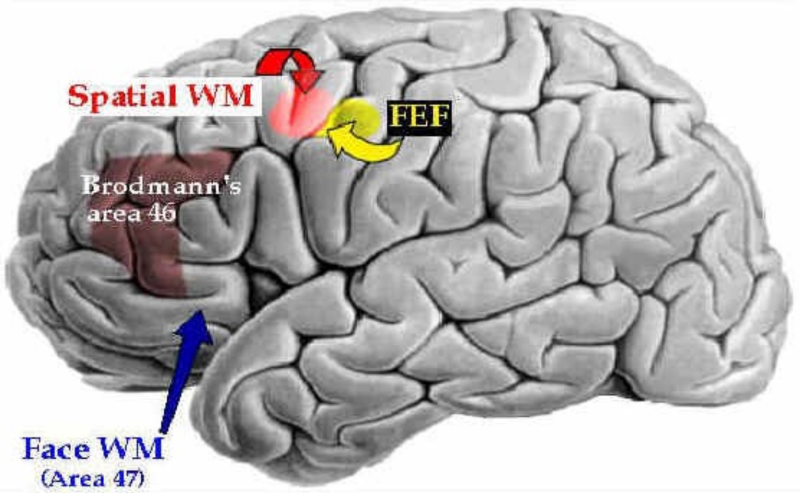
Memory and Attention



Human Information Processing Stage (Leahey, 1989)

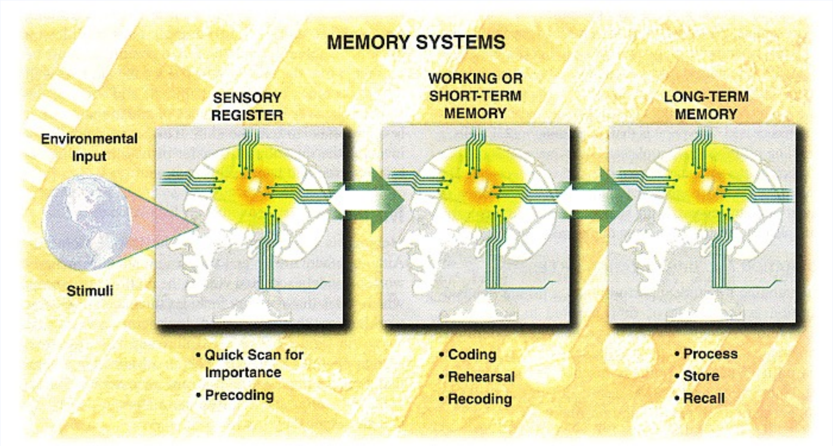
Short-term memory (STM)

- **Scratch-pad for temporary recall**
  - **Rapid access ~ 70ms**
  - **Rapid decay ~ 200ms**
  - **Limited capacity - 7± 2 chunks**



Sensory memory

- **Buffers for stimuli received through senses**
  - iconic memory: visual stimuli
  - echoic memory: aural stimuli
  - haptic memory: tactile stimuli
- **Examples**
  - “sparkler” trail
  - stereo sound
- **Continuously overwritten**



[redacted]

Memory

There are three types of memory function:

Sensory memories

Short-term memory or working memory

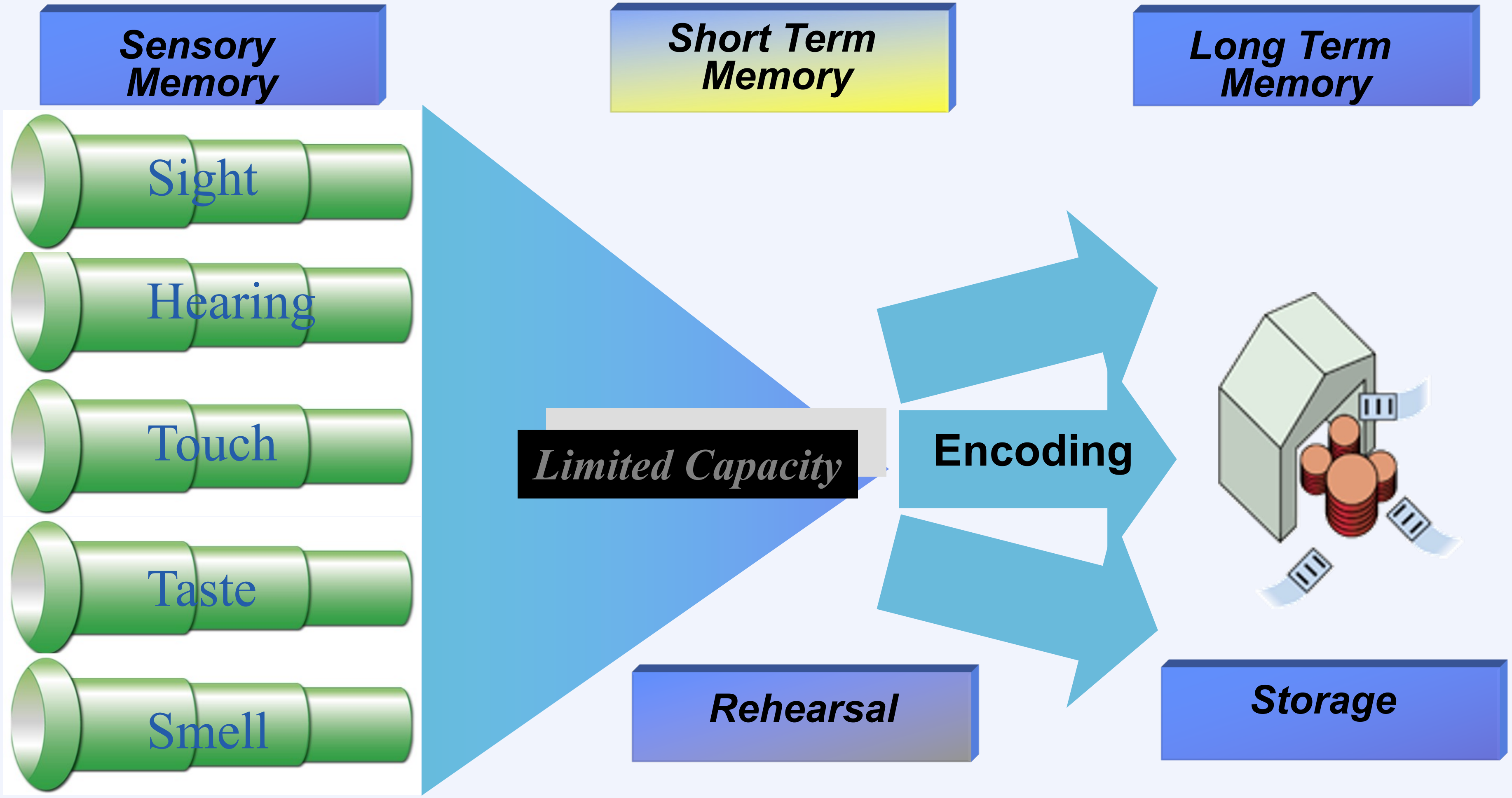
Attention

Long-term memory

Rehearsal

Selection of stimuli governed by level of arou

Memory and Attention



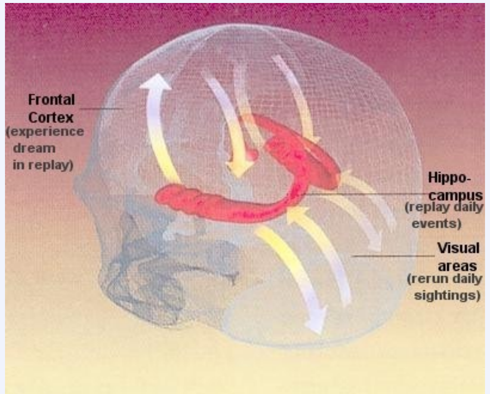
Human Information Processing Stage (Leahey, 1989)

Long-term memory (LTM)

- Repository for all our knowledge
  - slow access ~ 1/10 second
  - slow decay, if any
  - huge or unlimited capacity

- Two types
  - episodic – serial memory of events
  - semantic – structured memory of facts,concepts, skills

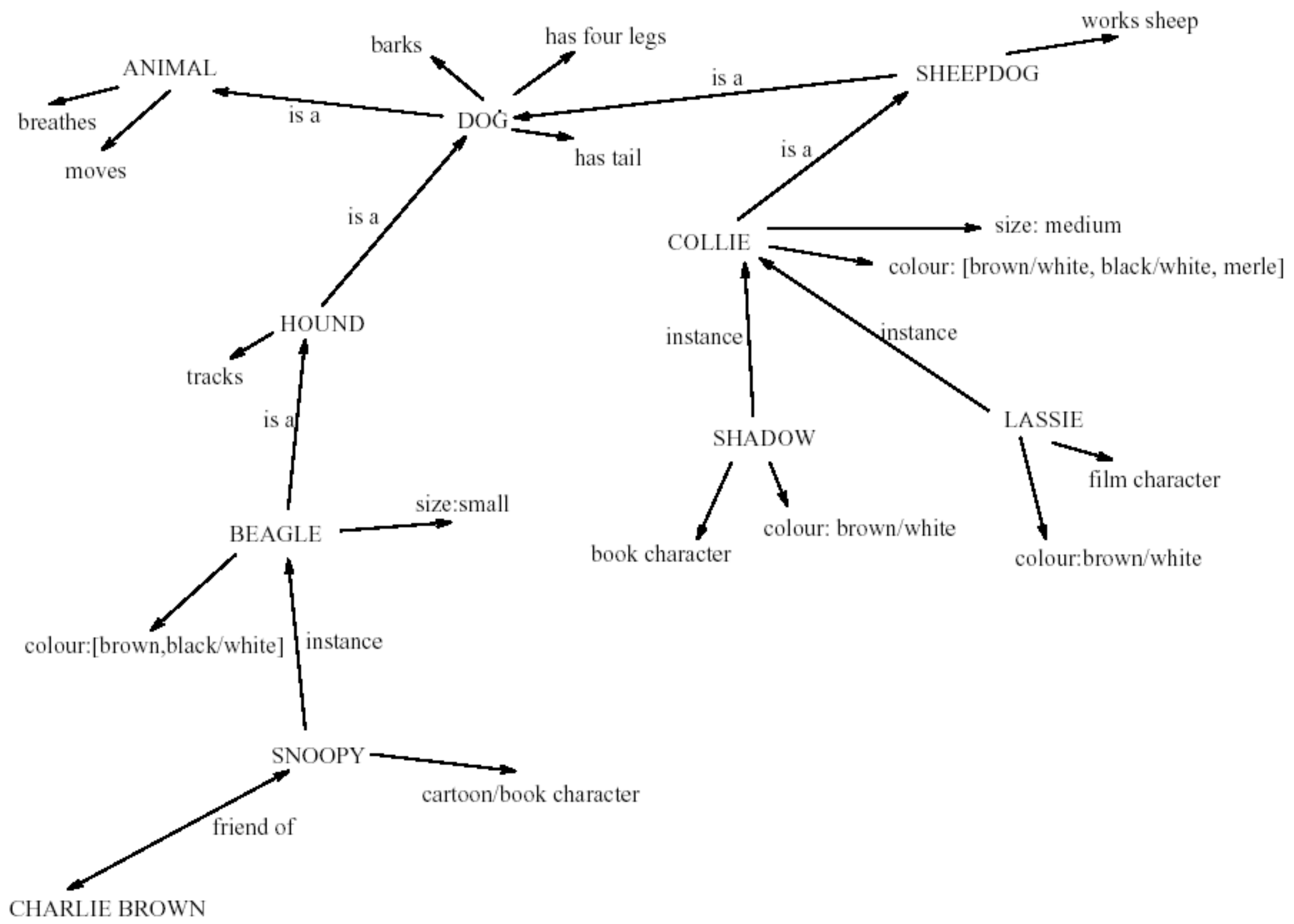
semantic LTM derived from episodic LTM



Long-term memory (cont.)

- **Semantic memory structure**
  - provides access to information
  - represents relationships between bits of information
  - supports inference
- **Model: semantic network**
  - inheritance – child nodes inherit properties of parent nodes
  - relationships between bits of information explicit
  - supports inference through inheritance

LTM - Semantic Network



Models of LTM - Frames

- Information organized in data structures
- Slots in structure instantiated with values for instance of data
- Type–subtype relationships

DOG

Fixed
- legs: 4

Default
- diet: carniverous
- sound: bark

Variable
- size:
- colour

COLLIE

Fixed
- breed of: DOG
- type: sheepdog

Default
- size: 65 cm

Variable
- colour

Models of LTM - Scripts

Model of stereotypical information required to interpret situation

Script has elements that can be instantiated with values for context

Script for a visit to the vet

Entry conditions:	*dog ill* *vet open* *owner has money*	Roles:	*vet examines* *diagnoses* *treats* *owner brings dog in* *pays* *takes dog out*
Result:	*dog better* *owner poorer* *vet richer*	Scenes:	*arriving at reception* *waiting in room* *examination* *paying*
Props:	*examination table* *medicine* *instruments*	Tracks:	*dog needs medicine* *dog needs operation*

Models of LTM - Production Rules

Representation of procedural knowledge.

Condition/action rules

if condition is matched

then use rule to determine action.

IF dog is wagging tail
THEN pat dog

IF dog is growling
THEN run away

LTM - Storage of information

- Rehearsal
  - information moves from STM to LTM
- Total time hypothesis
  - amount retained proportional to rehearsal time
- Distribution of practice effect
  - optimized by spreading learning over time
- Structure, meaning and familiarity
  - information easier to remember

LTM - Forgetting

Decay

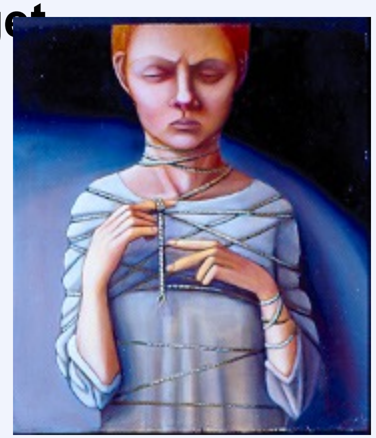
- information is lost gradually but very slowly

Interference

- new information replaces old: retroactive interference
- old may interfere with new: proactive inhibition

so may not forget at all memory is selective …

… affected by emotion – can subconsciously `choose' to forget



LTM - Retrieval

Recall

- information reproduced from memory can be assisted by cues, e.g. categories, imagery

Recognition

- information gives knowledge that it has been seen before
- less complex than recall - information is cue


See You Next Week
Thank You For Your Participation