CHAPTER

The Human Factor

2

The deepest challenges in human-computer interaction (HCI) lie in the human factor. Humans are complicated. Computers, by comparison, are simple. Computers are designed and built and they function in rather strict terms according to their programmed capabilities. There is no parallel with humans. Human scientists (including those in HCI) confront something computer scientists rarely think about: variability. Humans differ. We're young, old, female, male, experts, novices, left-handed, right-handed, English-speaking, Chinese-speaking, from the north, from the south, tall, short, strong, weak, fast, slow, able-bodied, disabled, sighted, blind, motivated, lazy, creative, bland, tired, alert, and on and on. The variability humans bring to the table means our work is never precise. It is always approximate. Designing systems that work well, period, is a lofty goal, but unfortunately, it is not possible to the degree we would like to achieve. A system might work well for a subset of people, but venture to the edges along any dimension (see list above), and the system might work poorly, or not at all. It is for this reason that HCI designers have precepts like "know thy user" (Shneiderman and Plaisant, 2005, p. 66).

Researchers in HCI have questions—lots of them. We are good at the small ones, but the big ones are difficult to answer: Why do humans make mistakes? Why do humans forget how to do things? Why do humans get confused while installing apps on their computers? Why do humans have trouble driving while talking on a mobile phone? Why do humans enjoy *Facebook* so much? Obviously, the human part is hugely important and intriguing. The more we understand humans, the better are our chances of designing interactive systems—interactions—that work as intended. So in this chapter I examine the human, but the computer and the interaction are never far away.

The questions in the preceding paragraph begin with *why*. They are big questions. Unfortunately, they do not lend themselves to empirical enquiry, which is the focus of this book. Take the first question: *Why do humans make mistakes?* From an empirical research perspective, the question is too broad. It cannot be answered with any precision. Our best bet is to narrow in on a defined group of humans (a *population*) and ask them to do a particular task on a particular system in a particular environment. We observe the interaction and measure the behavior. Along the way, we log the mistakes, classify them, count them, and take note of where and how the mistakes occurred. If our methodology is sound, we might assimilate enough information to put forth an answer to the *why* question—in a narrow sense.

If we do enough research like this, we might develop an answer in a broad sense. But a grounded and rigorous approach to empirical research requires small and narrowly focused questions.

Descriptive models, which will be discussed in Chapter 7, seek to delineate and categorize a problem space. They are tools for thinking, rather than tools for predicting. A descriptive model for "the human" would be useful indeed. It would help us get started in understanding the human, to delineate and categorize aspects of the human that are relevant to HCI. In fact there are many such models, and I will introduce several in this chapter.

2.1 Time scale of human action

Newell's *Time Scale of Human Action* is a descriptive model of the human (Newell, 1990, p. 122). It delineates the problem space by positioning different types of human actions in timeframes within which the actions occur. (See Figure 2.1.) The model has four bands, a *biological band*, a *cognitive band*, a *rational band*, and a *social band*. Each band is divided into three levels. Time is ordered by seconds and appears on a logarithmic scale, with each level a factor of ten longer than the level below it. The units are microseconds at the bottom and months at the top. For nine levels, Newell ascribes a label for the human system at work (e.g., *operations* or *task*). Within these labels we see a connection with HCI. The labels for the bands suggest a worldview or theory of human action.

The most common dependent variable in experimental research in HCI is time—the time for a user to do a task. In this sense, Newell's time-scale model is relevant to HCI. The model is also appropriate because it reflects the multidisciplinary nature of the field. HCI research is both *high level* and *low level*, and we see this in the model. If desired, we could select a paper at random from an HCI conference proceedings or journal, study it, then position the work somewhere in Figure 2.1. For example, research on selection techniques, menu design, force or auditory feedback, text entry, gestural input, and so on, is within the cognitive band. The tasks for these interactions typically last on the order of a few hundred milliseconds (ms) to a few dozen seconds. Newell characterizes these as deliberate acts, operations, and unit tasks.

Up in the rational band, users are engaged in tasks that span minutes, tens of minutes, or hours. Research topics here include web navigation, user search strategies, user-centered design, collaborative computing, ubiquitous computing, social navigation, and situated awareness. Tasks related to these research areas occupy users for minutes or hours.

Tasks lasting days, weeks, or months are in the social band. HCI topics here might include workplace habits, groupware usage patterns, social networking, online dating, privacy, media spaces, user styles and preferences, design theory, and so on.

Another insight Newell's model provides pertains to research methodology. Research at the bottom of the scale is highly quantitative in nature. Work in the biological band, for example, is likely experimental and empirical—at the level of neural

Scale (sec)	Time Units	System	World (theory)
10^7	Months		SOCIAL BAND
10^6	Weeks		
10^5	Days		
10^4	Hours	Task	RATIONAL BAND
10^3	10 min	Task	
10^2	Minutes	Task	
10^1	10 sec	Unit task	COGNITIVE BAND
10^0	1 sec	Operations	
10^{-1}	100 ms	Deliberate act	
10^{-2}	10 ms	Neural circuit	BIOLOGICAL BAND
10^{-3}	1 ms	Neuron	
10^{-4}	100 μs	Organelle	

FIGURE 2.1

Newell's time scale of human action.

(From Newell, 1990, p. 122)

impulses. At the top of the scale, the reverse is true. In the social band, research methods tend to be qualitative and non-experimental. Techniques researchers employ here include interviews, observation, case studies, scenarios, and so on. Furthermore, the transition between qualitative and quantitative methods moving from top to bottom in the figure is gradual. As one methodology becomes more prominent, the other becomes less prominent. Researchers in the social band primarily use qualitative methods, but often include some quantitative methods. For example, research on workplace habits, while primarily qualitative, might include some quantitative methods (e.g., counting the number of personal e-mails sent each day while at work). Thus, qualitative research in the social band also includes some quantitative assessment. Conversely, researchers in the cognitive band primarily use quantitative methods but typically include some qualitative methods. For example, an experiment on human performance with pointing devices, while primarily quantitative, might include an interview at the end to gather comments and suggestions on the interactions. Thus, quantitative, experimental work in the cognitive band includes some qualitative assessment as well.

Newell speculates further on bands above the social band: a *historical band* operating at the level of years to thousands of years, and an *evolutionary band* operating at the level of tens of thousands to millions of years (Newell, 1990, p. 152). We will forgo interpreting these in terms of human-computer interaction.

2.2 Human factors

There are many ways to characterize the human in interactive systems. One is the model human processor of Card et al. (1983), which was introduced in Chapter 1.

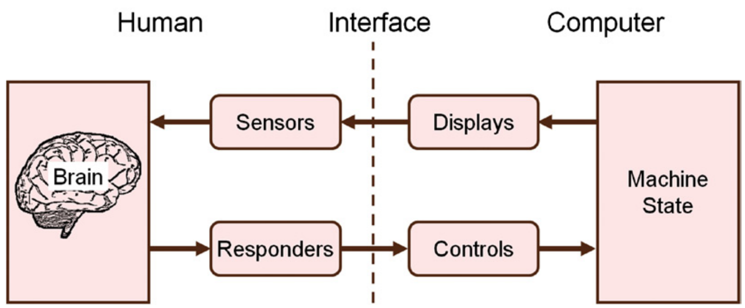


FIGURE 2.2

Human factors view of the human operator in a work environment.

(After Kantowitz and Sorkin, 1983, p. 4)

Other characterizations exist as well. Human factors researchers often use a model showing a human operator confronting a machine, like the image in Figure 2.2. The human monitors the state of the computer through sensors and displays and controls the state of the computer through responders and controls. The dashed vertical line is important since it is at the interface where interaction takes place. This is the location where researchers observe and measure the behavioral events that form the interaction.

Figure 2.2 is a convenient way to organize this section, since it simplifies the human to three components: sensors, responders, and a brain.

2.3 Sensors

Rosa: You deny everything except what you want to believe. That's the sort of man you are.
Bjartur: I have my five senses, and don't see what need there is for more.

(Halldór Laxness, *Independent People*)

The five classical human senses are vision, hearing, taste, smell, and touch. Each brings distinctly different physical properties of the environment to the human. One feature the senses share is the reception and conversion into electrical nerve signals of physical phenomena such as sound waves, light rays, flavors, odors, and physical contact. The signals are transmitted to the brain for processing. Sensory stimuli and sense organs are purely physiological. Perception, discussed later, includes both the sensing of stimuli and use of the brain to develop identification, awareness, and understanding of what is being sensed. We begin with the first of the five senses just noted: vision.

2.3.1 Vision (Sight)

Vision, or sight, is the human ability to receive information from the environment in the form of visible light perceived by the eye. The visual sensory channel

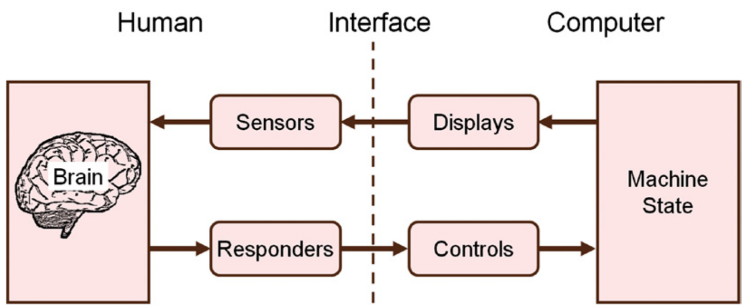


FIGURE 2.2

Human factors view of the human operator in a work environment.

(After Kantowitz and Sorkin, 1983, p. 4)

Other characterizations exist as well. Human factors researchers often use a model showing a human operator confronting a machine, like the image in Figure 2.2. The human monitors the state of the computer through sensors and displays and controls the state of the computer through responders and controls. The dashed vertical line is important since it is at the interface where interaction takes place. This is the location where researchers observe and measure the behavioral events that form the interaction.

Figure 2.2 is a convenient way to organize this section, since it simplifies the human to three components: sensors, responders, and a brain.

2.3 Sensors

Rosa: You deny everything except what you want to believe. That's the sort of man you are.
Bjartur: I have my five senses, and don't see what need there is for more.

(Halldór Laxness, *Independent People*)

The five classical human senses are vision, hearing, taste, smell, and touch. Each brings distinctly different physical properties of the environment to the human. One feature the senses share is the reception and conversion into electrical nerve signals of physical phenomena such as sound waves, light rays, flavors, odors, and physical contact. The signals are transmitted to the brain for processing. Sensory stimuli and sense organs are purely physiological. Perception, discussed later, includes both the sensing of stimuli and use of the brain to develop identification, awareness, and understanding of what is being sensed. We begin with the first of the five senses just noted: vision.

2.3.1 Vision (Sight)

Vision, or sight, is the human ability to receive information from the environment in the form of visible light perceived by the eye. The visual sensory channel

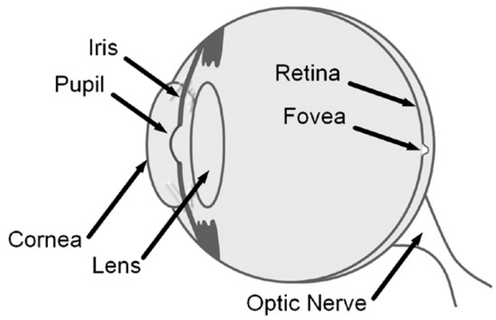


FIGURE 2.3

The eye.

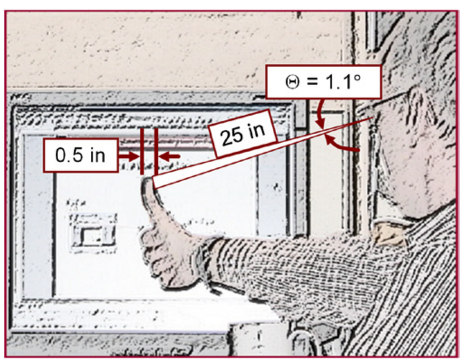


FIGURE 2.4

The fovea image spans a region a little more than one degree of visual angle.

is hugely important, as most people obtain about 80 percent of their information though the sense of light (Asakawa and Takagi, 2007). The act of seeing begins with the reception of light through the eye's lens. The lens focuses the light into an image projected on to the retina at the back of the eye. (See Figure 2.3.) The retina is a transducer, converting visible light into neurological signals sent to the brain via the optic nerve.

Near the center of the retina is the fovea, which is responsible for sharp central vision, such as reading or watching television. The fovea image in the environment encompasses a little more than one degree of visual angle, approximately equivalent to the width of one's thumb at arm's length (see Figure 2.4). Although the fovea is only about 1 percent of the retina in size, the neural processing associated with the fovea image engages about 50 percent of the visual cortex in the brain.

As with other sensory stimuli, light has properties such as intensity and frequency.

Frequency. Frequency is the property of light leading to the perception of color. Visible light is a small band in the electromagnetic spectrum, which ranges from

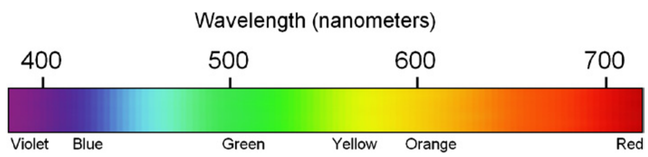


FIGURE 2.5

The visible spectrum of electromagnetic waves.

radio waves to x-rays and gamma rays. Different colors are positioned within the visible spectrum of electromagnetic waves, with violet at one end (390 nanometers) and red at the other (750 nm). (See Figure 2.5; colors not apparent in grayscale print).

Intensity. Although the frequency of light is a relatively simple concept, the same cannot be said for the intensity of light. Quantifying light intensity, from the human perspective, is complicated because the eye's light sensitivity varies by the wavelength of the light and also by the complexity of the source (e.g., a single frequency versus a mixture of frequencies). Related to intensity is *luminance*, which refers to the amount of light passing through a given area. With luminance comes *brightness*, a subjective property of the eye that includes perception by the brain. The unit for luminance is candela per square meter (cd/m^2).

Fixations and saccades. Vision is more than the human reception of electromagnetic waves having frequency and intensity. Through the eyes, humans look at and perceive the environment. In doing so, the eyes engage in two primitive actions: *fixations* and *saccades*. During a fixation, the eyes are stationary, taking in visual detail from the environment. Fixations can be long or short, but typically last at least 200 ms. Changing the point of fixation to a new location requires a saccade—a rapid repositioning of the eyes to a new position. Saccades are inherently quick, taking only 30–120 ms. Early and influential research on fixations and saccades was presented in a 1965 publication in Russian by Yarbus, translated as *Eye Movements and Vision* (reviewed in Tatler, Wade, Kwan, Findlay, and Velichkovsky, 2010). Yarbus demonstrated a variety of inspection patterns for people viewing scenes. One example used *The Unexpected Visitor* by painter Ilya Repin (1844–1930). Participants were given instructions and asked to view the scene, shown in Figure 2.6a. Eye movements (fixations and saccades) were recorded and plotted for a variety of tasks. The results for one participant are shown in Figure 2.6b for the task "remember the position of people and objects in the room" and in Figure 2.6c for the task "estimate the ages of the people." Yarbus provided many diagrams like this, with analyses demonstrating differences within and between participants, as well as changes in viewing patterns over time and for subsequent viewings. He noted, for example, that the similarity of inspection patterns for a single viewer was greater than the patterns between viewers.

HCI research in eye movements has several themes. One is analyzing how people read and view content on web pages. Figure 2.7 shows an example of a

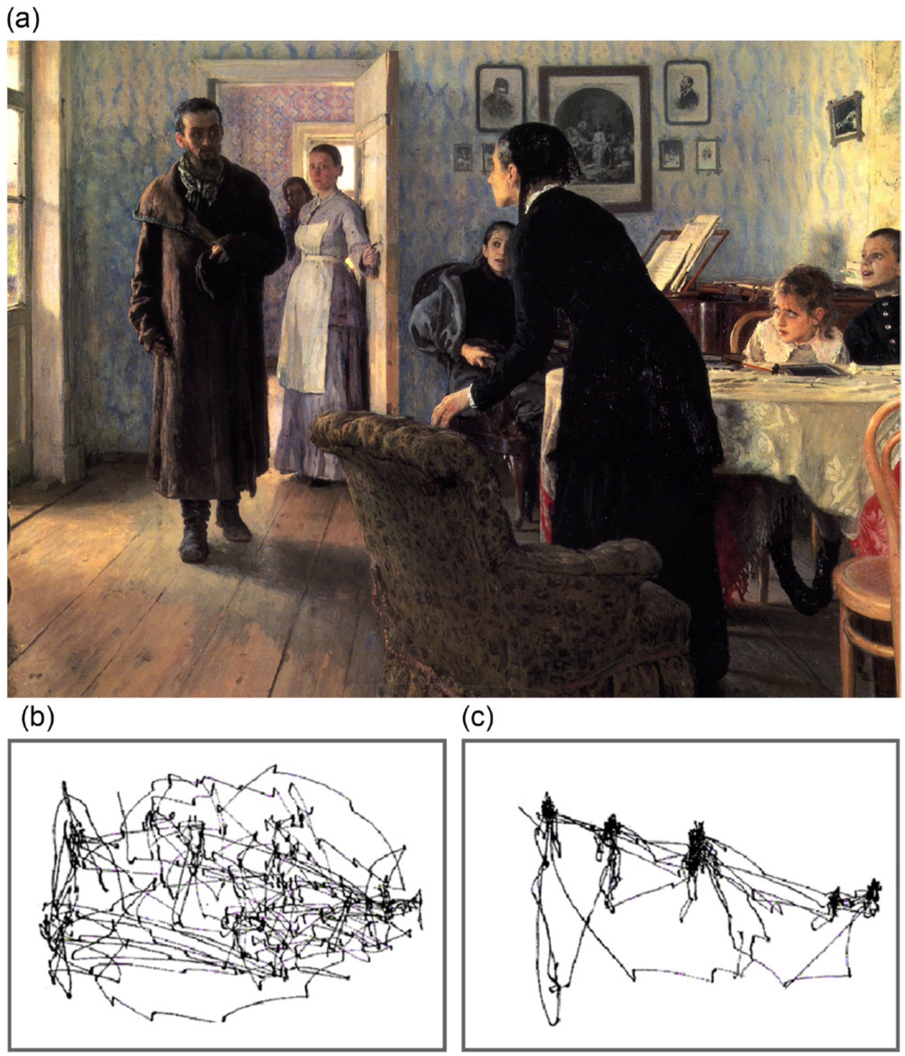


FIGURE 2.6

Yarbus' research on eye movements and vision (Tatler et al., 2010). (a) Scene. (b) Task: *Remember the position of the people and objects in the room.* (c) Task: *Estimate the ages of the people.*

scanpath (a sequence of fixations and saccades) for a user viewing content at different places on a page. (See also J. H. Goldberg and Helfman, 2010, Figure 2.) The results of the analyses offer implications for page design. For example, advertisers might want to know about viewing patterns and, for example, how males and females differ in viewing content. There are gender differences in eye movements

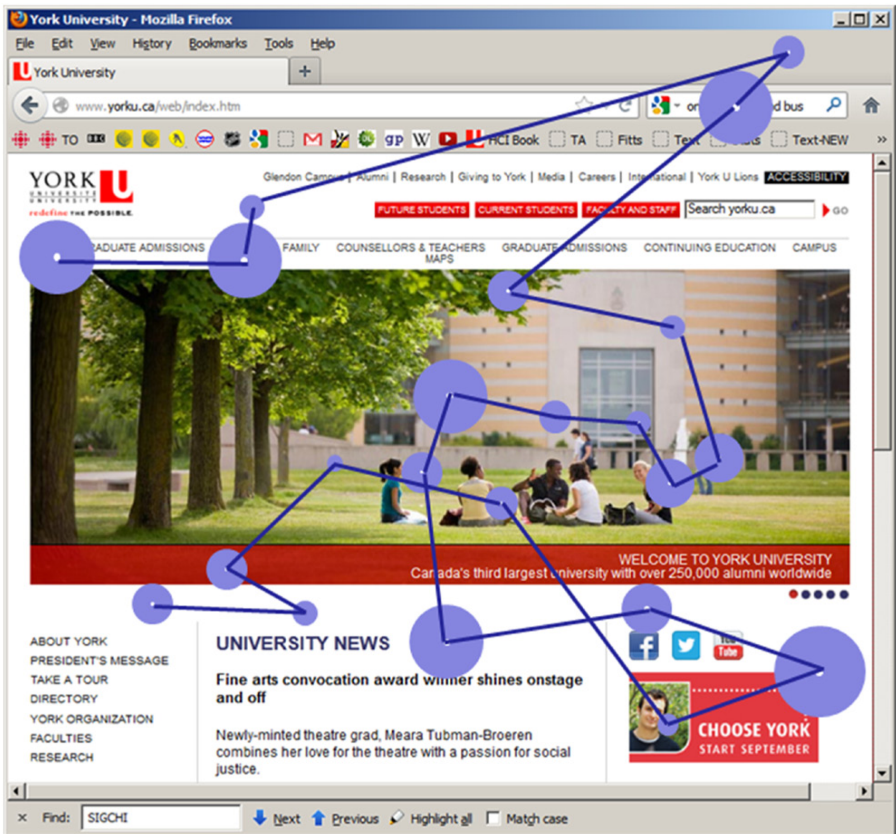


FIGURE 2.7

Scanpath for a user locating content on a web page.

(Pan et al., 2004), but it remains to be demonstrated how low-level experimental results can inform and guide design.

2.3.2 Hearing (Audition)

Hearing, or audition, is the detection of sound by humans. Sound is transmitted through the environment as sound waves—cyclic fluctuations of pressure in a medium such as air. Sound waves are created when physical objects are moved or vibrated, thus creating fluctuations in air pressure. Examples include plucking a string on a guitar, slamming a door, shuffling cards, or a human speaking. In the latter case, the physical object creating the sound is the larynx, or vocal cords, in the throat.

Hearing occurs when sound waves reach a human's ear and stimulate the ear drum to create nerve impulses that are sent to the brain. A single sound from a single source has at least four physical properties: intensity (loudness), frequency (pitch), timbre, and envelope. As a simple example, consider a musical note played from an instrument such as a trumpet. The note may be loud or soft (intensity); high or low (frequency). We hear and recognize the note as coming from a trumpet, as opposed to a flute, because of the note's timbre and envelope. Let's examine each of these properties.

Loudness. Loudness is the subjective analog to the physical property of intensity. It is quantified by *sound pressure level*, which expresses the pressure in a sound wave relative to the average pressure in the medium. The unit of sound pressure level is the decibel (dB). Human hearing begins with sounds of 0–10 dB. Conversational speech is about 50–70 dB in volume. Pain sets in when humans are exposed to sounds of approximately 120–140 dB.

Pitch. Pitch is the subjective analog of frequency, which is the reciprocal of the time between peaks in a sound wave's pressure pattern. The units of pitch are cycles per second, or Hertz (Hz). Humans can perceive sounds in the frequency range of about 20 Hz–20,000 Hz (20 kHz), although the upper limit tends to decrease with age.

Timbre. Timbre (aka richness or brightness) results from the harmonic structure of sounds. Returning to the example of a musical note, harmonics are integer multiples of a note's base frequency. For example, a musical note with base frequency of 400 Hz includes harmonics at 800 Hz, 1200 Hz, 1600 Hz, and so on. The relative amplitudes of the harmonics create the subjective sense of timbre, or richness, in the sound. While the human hears the note as 400 Hz, it is the timbre that distinguishes the tone as being from a particular musical instrument. For example, if notes of the same frequency and loudness are played from a trumpet and an oboe, the two notes sound different, in part, because of the unique pattern of harmonics created by each instrument. The purest form of a note is a sine wave, which includes the base frequency but no harmonics above the base frequency. The musical notes created by a flute are close to sine waves.

Envelope. Envelope is the way a note and its harmonics build up and transition in time—from silent to audible to silent. There is considerable information in the onset envelope, or attack, of musical notes. In the example above of the trumpet and oboe playing notes of the same frequency and same loudness, the attack also assists in distinguishing the source. If the trumpet note and oboe note were recorded and played back with the attack removed, it would be surprisingly difficult to distinguish the instruments. The attack results partly from inherent properties of instruments (e.g., brass versus woodwind), but also from the way notes are articulated (e.g., staccato versus legato).

Besides physical properties, sound has other properties. These have to do with human hearing and perception. Sounds, complex sounds, can be described as being harmonious (pleasant) or discordant (unpleasant). This property has to do with how different frequencies mix together in a complex sound, such as a musical chord. Sounds may also convey a sense of urgency or speed.

Humans have two ears, but each sound has a single source. The slight difference in the physical properties of the sound as it arrives at each ear helps humans in identifying a sound's location (direction and distance). When multiple sounds from multiple sources are heard through two ears, perceptual effects such as stereo emerge.

Sounds provide a surprisingly rich array of cues to humans, whether walking about while shopping or sitting in front of a computer typing an e-mail message. Not surprisingly, sound is crucial for blind users, for example, in conveying information about the location and distance of environmental phenomena (Talbot and Cowan, 2009).

2.3.3 Touch (Tactition)

Although touch, or tactition, is considered one of the five traditional human senses, touch is just one component of the somatosensory system. This system includes sensory receptors in the skin, muscles, bones, joints, and organs that provide information on a variety of physical or environmental phenomena, including touch, temperature, pain, and body and limb position. Tactile feedback, in HCI, refers to information provided through the somatosensory system from a body part, such as a finger, when it is in contact with (touching) a physical object. Additional information, such as the temperature, shape, texture, or position of the object, or the amount of resistance, is also conveyed.

All user interfaces that involve physical contact with the user's hands (or other body parts) include tactile feedback. Simply grasping a mouse and moving it brings considerable information to the human operator: the smooth or rubbery feel of the mouse chassis, slippery or sticky movement on the desktop. Interaction with a desktop keyboard is also guided by tactile feedback. The user senses the edges and shapes of keys and experiences resistance as a key is pressed. Tactile identifiers on key tops facilitate eyes-free touch typing. Identifiers are found on the 5 key for numeric keypads and on the F and J keys for alphanumeric keyboards. Sensing the identifier informs the user that the home position is acquired. (See Figure 2.8a.)

Augmenting the user experience through active tactile feedback is a common research topic. Figure 2.8b shows a mouse instrumented with a solenoid-driven pin below the index finger (Akamatsu et al., 1995). The pin is actuated (pulsed) when the mouse cursor crosses a boundary, such as the edge of a soft button or window. The added tactile feedback helps inform and guide the interaction and potentially reduces the demand on the visual channel. A common use of tactile feedback in mobile phones is vibration, signaling an incoming call or message. (See Figure 2.8c.)

2.3.4 Smell and taste

Smell (olfaction) is the ability to perceive odors. For humans, this occurs through sensory cells in the nasal cavity. Taste (gustation) is a direct chemical reception of sweet, salty, bitter, and sour sensations through taste buds in the tongue and oral cavity. Flavor is a perceptual process in the brain that occurs through a partnering of the in identifying a sound's location (direction and distance). When multiple sounds from multiple sources are heard through two ears, perceptual effects such as stereo emerge.

Sounds provide a surprisingly rich array of cues to humans, whether walking about while shopping or sitting in front of a computer typing an e-mail message. Not surprisingly, sound is crucial for blind users, for example, in conveying information about the location and distance of environmental phenomena (Talbot and Cowan, 2009).

2.3.3 Touch (Tactition)

Although touch, or tactition, is considered one of the five traditional human senses, touch is just one component of the somatosensory system. This system includes sensory receptors in the skin, muscles, bones, joints, and organs that provide information on a variety of physical or environmental phenomena, including touch, temperature, pain, and body and limb position. Tactile feedback, in HCI, refers to information provided through the somatosensory system from a body part, such as a finger, when it is in contact with (touching) a physical object. Additional information, such as the temperature, shape, texture, or position of the object, or the amount of resistance, is also conveyed.

All user interfaces that involve physical contact with the user's hands (or other body parts) include tactile feedback. Simply grasping a mouse and moving it brings considerable information to the human operator: the smooth or rubbery feel of the mouse chassis, slippery or sticky movement on the desktop. Interaction with a desktop keyboard is also guided by tactile feedback. The user senses the edges and shapes of keys and experiences resistance as a key is pressed. Tactile identifiers on key tops facilitate eyes-free touch typing. Identifiers are found on the 5 key for numeric keypads and on the F and J keys for alphanumeric keyboards. Sensing the identifier informs the user that the home position is acquired. (See Figure 2.8a.)

Augmenting the user experience through active tactile feedback is a common research topic. Figure 2.8b shows a mouse instrumented with a solenoid-driven pin below the index finger (Akamatsu et al., 1995). The pin is actuated (pulsed) when the mouse cursor crosses a boundary, such as the edge of a soft button or window. The added tactile feedback helps inform and guide the interaction and potentially reduces the demand on the visual channel. A common use of tactile feedback in mobile phones is vibration, signaling an incoming call or message. (See Figure 2.8c.)

2.3.4 Smell and taste

Smell (olfaction) is the ability to perceive odors. For humans, this occurs through sensory cells in the nasal cavity. Taste (gustation) is a direct chemical reception of sweet, salty, bitter, and sour sensations through taste buds in the tongue and oral cavity. Flavor is a perceptual process in the brain that occurs through a partnering of the

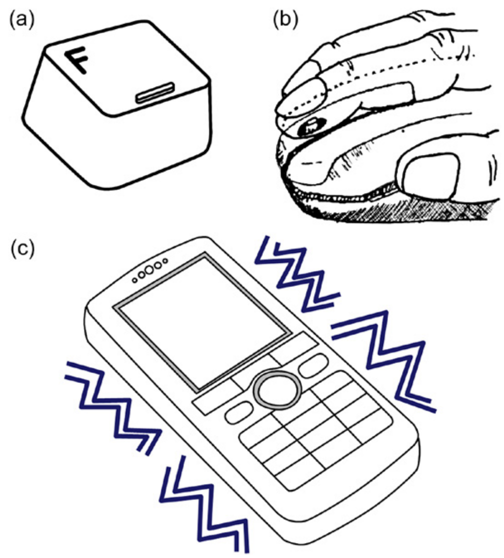


FIGURE 2.8

Tactile feedback: (a) Identifier on key top. (b) Solenoid-driven pin under the index finger. (c) Vibration signals an in-coming call.

(Adapted from Akamatsu, MacKenzie, and Hasbrouq, 1995)

smell and taste senses. Although smell and taste are known intuitively by virtually all humans—and with expert-like finesse—they are less understood than the visual and auditory senses. Complex smells and tastes can be built up from simpler elements, but the perceptual processes for this remain a topic of research. For example, classification schemes have been developed for specific industries (e.g., perfume, wine) but these do not generalize to human experiences with other smells and tastes.

While humans use smell and taste all the time without effort, these senses are not generally "designed in" to systems. There are a few examples in HCI. Brewster et al. (2006) studied smell as an aid in searching digital photo albums. Users employed two tagging methods, text and smell, and then later used the tags to answer questions about the photos. Since smell has links to memory, it was conjectured that smell cues might aid in recall. In the end, recall with smell tags was poorer than with word tags. Related work is reported by Bodnar et al. (2004) who compared smell, auditory, and visual modalities for notifying users of an interruption by an incoming message. They also found poorer performance with smell. Notable in both examples, though, is the use of an empirical research methodology to explore the potential of smell in a user interface. Both studies included all the hallmarks of experimental research, including an independent variable, dependent variables, statistical significance testing, and counterbalancing of the independent variable.

from the fixations shown in Figure 2.7, where the user was simply viewing content on a web page. In Figure 2.13b, the interaction includes numerous fixations meeting the required dwell time criterion to select soft keys. There is also a fixation (with two corresponding saccades) to view the typed text.

2.5 The brain

The brain is the most complex biological structure known. With billions of neurons, the brain provides humans with a multitude of capacities and resources, including pondering, remembering, recalling, reasoning, deciding, and communicating. While sensors (human inputs) and responders (human outputs) are nicely mirrored, it is the brain that connects them. Without sensing or experiencing the environment, the brain would have little to do. However, upon experiencing the environment through sensors, the brain's task begins.

2.5.1 Perception

Perception, the first stage of processing in the brain, occurs when sensory signals are received as input from the environment. It is at the perceptual stage that associations and meanings take shape. An auditory stimulus is perceived as harmonious or discordant. A smell is pleasurable or abhorrent. A visual scene is familiar or strange. Touch something and the surface is smooth or rough, hot or cold. With associations and meaning attached to sensory input, humans are vastly superior to the machines they interact with:

> *People excel at perception, at creativity, at the ability to go beyond the information given, making sense of otherwise chaotic events. We often have to interpret events far beyond the information available, and our ability to do this efficiently and effortlessly, usually without even being aware that we are doing so, greatly adds to our ability to function.*
>
> **(Norman, 1988, p. 136)**

Since the late 19th century, perception has been studied in a specialized area of experimental psychology known as *psychophysics*. Psychophysics examines the relationship between human perception and physical phenomena. In a psychophysics experiment, a human is presented with a physical stimulus and is then asked about the sensation that was felt or perceived. The link is between a measurable property of a real-world phenomenon that stimulates a human sense and the human's subjective interpretation of the phenomenon. A common experimental goal is to measure the *just noticeable difference* (JND) in a stimulus. A human subject is presented with two stimuli, one after the other. The stimuli differ in a physical property, such as frequency or intensity, and the subject is asked if the stimuli are the same or different. The task is repeated over a series of trials with random variations in the magnitude of the difference in the physical property manipulated. Below a from the fixations shown in Figure 2.7, where the user was simply viewing content on a web page. In Figure 2.13b, the interaction includes numerous fixations meeting the required dwell time criterion to select soft keys. There is also a fixation (with two corresponding saccades) to view the typed text.

2.5 The brain

The brain is the most complex biological structure known. With billions of neurons, the brain provides humans with a multitude of capacities and resources, including pondering, remembering, recalling, reasoning, deciding, and communicating. While sensors (human inputs) and responders (human outputs) are nicely mirrored, it is the brain that connects them. Without sensing or experiencing the environment, the brain would have little to do. However, upon experiencing the environment through sensors, the brain's task begins.

2.5.1 Perception

Perception, the first stage of processing in the brain, occurs when sensory signals are received as input from the environment. It is at the perceptual stage that associations and meanings take shape. An auditory stimulus is perceived as harmonious or discordant. A smell is pleasurable or abhorrent. A visual scene is familiar or strange. Touch something and the surface is smooth or rough, hot or cold. With associations and meaning attached to sensory input, humans are vastly superior to the machines they interact with:

> *People excel at perception, at creativity, at the ability to go beyond the information given, making sense of otherwise chaotic events. We often have to interpret events far beyond the information available, and our ability to do this efficiently and effortlessly, usually without even being aware that we are doing so, greatly adds to our ability to function.*
>
> **(Norman, 1988, p. 136)**

Since the late 19th century, perception has been studied in a specialized area of experimental psychology known as *psychophysics*. Psychophysics examines the relationship between human perception and physical phenomena. In a psychophysics experiment, a human is presented with a physical stimulus and is then asked about the sensation that was felt or perceived. The link is between a measurable property of a real-world phenomenon that stimulates a human sense and the human's subjective interpretation of the phenomenon. A common experimental goal is to measure the *just noticeable difference* (JND) in a stimulus. A human subject is presented with two stimuli, one after the other. The stimuli differ in a physical property, such as frequency or intensity, and the subject is asked if the stimuli are the same or different. The task is repeated over a series of trials with random variations in the magnitude of the difference in the physical property manipulated. Below a

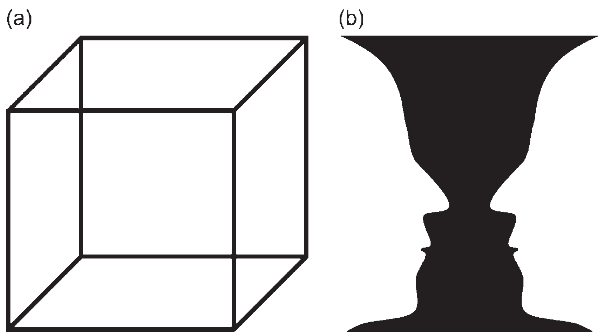


FIGURE 2.14

Ambiguous images: (a) Necker cube. (b) Rubin vase.

certain threshold, the difference between the two stimuli is so small that it is not perceived by the subject. This threshold is the JND. JND has been highly researched for all the human senses and in a variety of contexts. Does the JND depend on the absolute magnitude of the stimuli (e.g., high intensity stimuli versus low intensity stimuli)? Does the JND on one property (e.g., intensity) depend on the absolute value of a second property (e.g., frequency)? Does the JND depend on age, gender, or other property of the human? These are basic research questions that, on the surface, seem far afield from the sort of research likely to bear on human-computer interfaces. But over time and with new research extending results from previous research, there is indeed an application to HCI. For example, basic research in psychophysics is used in algorithms for audio compression in MP3 audio encoding.

Another property of perception is ambiguity—the human ability to develop multiple interpretations of a sensory input. Ambiguous images provide a demonstration of this ability for the visual sense. Figure 2.14a shows the Necker wireframe cube. Is the top-right corner on the front surface or the back surface? Figure 2.14b shows the Rubin vase. Is the image a vase or two faces? The very fact that we sense ambiguity in these images reveals our perceptual ability to go beyond the information given.

Related to ambiguity is illusion, the deception of common sense. Figure 2.15a shows Ponzo lines. The two black lines are the same length; however, the black line near the bottom of the illustration appears shorter because of the three-dimensional perspective. Müller-Lyer arrows are shown in Figure 2.15b. In comparing the straight-line segments in the two arrows, the one in the top arrow appears longer when in fact both are the same length. Our intuition has betrayed us.

If illusions are possible in visual stimuli, it is reasonable to expect illusions in the other senses. An example of an auditory illusion is the Shepard musical scale. It is perceived by humans to rise or fall continuously, yet it somehow stays the same. A variation is a continuous musical tone known as the Shepard-Risset glissando—a tone that continually rises in pitch while also continuing to stay at the same pitch. Figure 2.16 illustrates this illusion. Each vertical line represents a sine

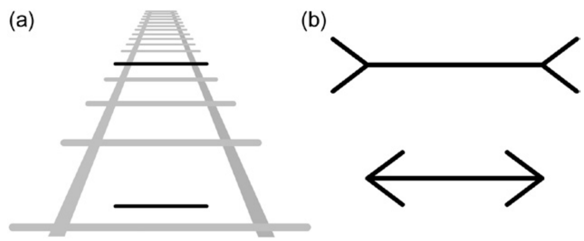


FIGURE 2.15

Visual illusion: (a) Ponzo lines. (b) Müller-Lyer arrows.

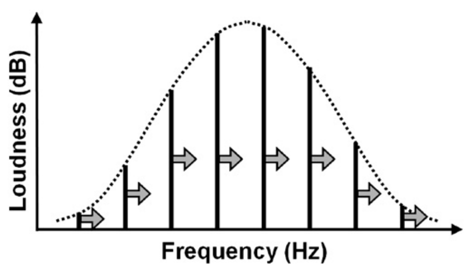


FIGURE 2.16

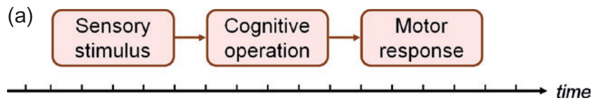
Auditory illusion. A collection of equally spaced sine waves rise in frequency. The human hears a tone that rises but stays the same.

wave. The height of each line is the perceived loudness of the sine wave. Each wave is displaced from its neighbor by the same frequency; thus, the waves are harmonics of a musical note with a base frequency equal to the displacement. This is the frequency of the single tone that a human perceives. If the sine waves collectively rise in frequency (block arrows in the figure), there is a sense that the tone is rising. Yet because the sine waves are equally spaced, there is a competing sense that the tone remains the same (because the frequency perceived is the distance between harmonics). Sine waves at the high end of the frequency distribution fade out, while new sine waves enter at the low end. Examples of the Shepard scale and the Shepard-Risset glissando can be heard on *YouTube*.

Tactile or *haptic* illusions also exist. A well-documented example is the "phantom limb." Humans who have lost a limb through amputation often continue to sense that the limb is present and that it moves along with other body parts as it did before amputation (Halligan, Zemen, and Berger, 1999).

Beyond perception, sensory stimuli are integrated into a myriad of other experiences to yield ideas, decisions, strategies, actions, and so on. The ability to excel at these higher-level capabilities is what propels humans to the top tier in classification schemes for living organisms. By and large it is the human ability to think and reason that affords this special position.

(a)



(b)

Operation	Typical time (ms)
Sensory reception	1 – 38
Neural transmission to brain	2 – 100
Cognitive processing	70 – 300
Neural transmission to muscle	10 – 20
Muscle latency and activation	30 – 70
Total:	113 – 528

FIGURE 2.17

Cognitive operation in a reaction time task: (a) Problem schematic. (b) Sequence of operations (Bailey, 1996, p. 41).

2.5.2 Cognition

Among the brain's vital faculties is cognition—the human process of conscious intellectual activity, such as thinking, reasoning, and deciding. Cognition spans many fields—from neurology to linguistics to anthropology—and, not surprisingly, there are competing views on the scope of cognition. Does cognition include social processes, or is it more narrowly concerned with deliberate goal-driven acts such as problem solving? It is beyond the reach of this book to unravel the many views of cognition. The task is altogether too great and in any case is aptly done in other references, many of them in human factors (e.g., B. H. Kantowitz and Sorkin, 1983; Salvendy, 1987; Wickens, 1987).

Sensory phenomena such as sound and light are easy to study because they exist in the physical world. Instruments abound for recording and measuring the presence and magnitude of sensory signals. Cognition occurs within the human brain, so studying cognition presents special challenges. For example, it is not possible to directly measure the time it takes for a human to make a decision. When does the measurement begin and end? Where is it measured? On what input is the human deciding? Through what output is the decision conveyed? The latter two questions speak to a sensory stimulus and a motor response that bracket the cognitive operation. Figure 2.17a illustrates this. Since sensory stimuli and motor responses are observable and measurable, the figure conveys, in a rough sense, how to measure a cognitive operation. Still, there are challenges. If the sensory stimulus is visual, the retina converts the light to neural impulses that are transmitted to the brain for perceptual processing. This takes time. So the beginning of the cognitive operation is not precisely known. Similarly, if the motor response involves a finger pressing a button, neural associations for the response are developed in the brain with nerve signals transmitted to the hand before movement begins. So the precise ending of the cognitive operation is also unknown. This sequence of events is shown in Figure 2.17b, noting the operations and the typical time for each step. The most remarkable observation here is the wide range of values—an indication of the difficulty in pinpointing where and how the measurements are made. Despite these challenges, techniques exist for measuring the duration of cognitive operations. These are discussed shortly.

The range of cognitive operations applicable to Figure 2.17 is substantial. While driving a car, the decision to depress a brake pedal in response to a changing signal light is simple enough. Similar scenarios abound in HCI. While using a mobile phone, one might decide to press the REJECT CALL key in response to an incoming call. While reading the morning news online, one might decide to click the CLOSE button on a popup ad. While editing a document, one might switch to e-mail in response to an audio alert of a new message. These examples involve a sensory stimulus, a cognitive operation, and a motor response, respectively.

Other decisions are more complicated. While playing the card game 21 (aka Blackjack), perhaps online[3], if a card is drawn and the hand then totals 16, the decision to draw another card is likely to produce a cognitive pause. What is the chance the next card will bring the hand above 21? Which cards 6 to KING are already dealt? Clearly, the decision in this scenario goes beyond the information in the sensory stimulus. There are strategies to consider, as well as the human ability to remember and recall past events—cards previously dealt. This ability leads us to another major function of the brain—memory.

2.5.3 Memory

Memory is the human ability to store, retain, and recall information. The capacity of our memory is remarkable. Experiences, whether from a few days ago or from decades past, are collected together in the brain's vast repository known as *long-term memory*. Interestingly enough, there are similarities between memory in the brain and memory in a computer. Computer memory often includes separate areas for data and code. In the brain, memory is similarly organized. A declarative/explicit area stores information about events in time and objects in the external world. This is similar to a data space. An implicit/procedural area in the brain's memory stores information about how to use objects or how to do things. This is similar to a code space.[4]

Within long-term memory is an active area for *short-term memory* or *working memory*. The contents of working memory are active and readily available for access. The amount of such memory is small, about seven units, depending on the task and the methodology for measurement. A study of short-term memory was

[3]The parenthetic "perhaps online" is included as a reminder that many activities humans do in the physical world have a counterpart in computing, often on the Internet.

[4]The reader is asked to take a cautious and loose view of the analogy between human memory and computer memory. Attempts to formulate analogies from computers to humans are fraught with problems. Cognitive scientists, for example, frequently speak of human cognition in terms of operators, operands, cycles, registers, and the like, and build and test models that fit their analogies. Such *reverse anthropomorphism*, while tempting and convenient, is unlikely to reflect the true inner workings of human biology.

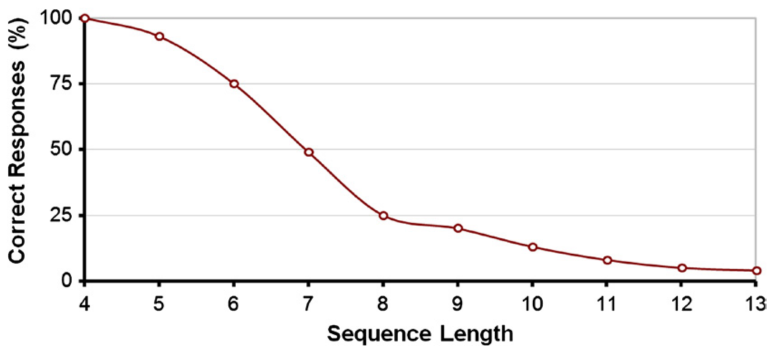


FIGURE 2.18

Results of a test of short-term memory.

published in 1956 in a classic essay by Miller, aptly titled "The Magic Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information" (G. A. Miller, 1956).[5] Miller reviewed a large number of studies on the absolute judgment of stimuli, such as pitch in an auditory stimulus or salt concentration in water in a taste stimulus. Humans are typically able to distinguish about seven levels of a uni-dimensional stimulus.[6]

Miller extended this work to human memory, describing an experiment where participants were presented with a sequence of items and then asked to recall the items. He found that the human ability with such tasks is, similarly, about seven items ($\pm$2). A simple demonstration of Miller's thesis is shown in Figure 2.18. For this "mini-experiment," log sheets were distributed to students in a class on human-computer interaction ($n \approx 60$). The instructor dictated sequences of random digits, with sequences varying in length from four digits to 13 digits. After each dictation, students copied the sequence from short-term memory onto the log sheet. The percentage of correct responses by sequence length is shown in the figure. At length seven the number of correct responses was about 50 percent. At lengths five and nine the values were about 90 percent and 20 percent, respectively.[7] See also student exercise 2-2 at the end of this chapter.

[5]Miller's classic work is referred to as an *essay* rather than a *research paper*. The essay is casual in style and, consequently, written in the first person; for example, "I am simply pointing to the obvious fact that…" (G. A. Miller, 1956, p. 93). Research papers, on the other hand, are generally plain in style and avoid first-person narratives (cf. "This points to the fact that…").

[6]The human ability to distinguish levels is greater if the stimulus is multidimensional; that is, the stimulus contains two or more independent attributes, such as a sound that varies in pitch and intensity.

[7]A response was deemed correct only if all the items were correctly recalled. For the longer sequences, many responses were "mostly correct." For example, at sequence length = 7, many of the responses had five or six items correct.

Miller extended his work by revealing and analyzing a simple but powerful process within the brain: our ability to associate multiple items as one. So-called *chunking* is a process whereby humans group a series of low-level items into a single high-level item. He described an example using binary digits. For example, a series of 16 bits, such as 1000101101110010, would be extremely difficult to commit to memory. If, however, the bits are collected into groups of four and chunked into decimal digits, the pattern is much easier to remember: 1000101101110010→1000, 1011, 0111, 0010→8, 11, 7, 2. Card et al. (1983, 36) give the example of BSCBMICRA. At nine units, the letter sequence is beyond the ability of most people to repeat back. But the sequence is similar to the following three groups of three-letter sequences: CBS IBM RCA. Shown like this, the sequence contains three chunks and is relatively easy to remember provided the person can perform the recoding rapidly enough. The process of chunking is mostly informal and unstructured. Humans intuitively build up chunked structures recursively and hierarchically, leading to complex organizations of memory in the brain.

2.6 Language

Language—the mental faculty that allows humans to communicate—is universally available to virtually all humans. Remarkably, language as speech is available without effort. Children learn to speak and understand speech without conscious effort as they grow and develop. Writing, as a codification of language, is a much more recent phenomenon. Learning to write demands effort, considerable effort, spanning years of study and practice. Daniels and Bright distinguish language and writing as follows: "Humankind is defined by language; but civilization is defined by writing." (Daniels and Bright, 1996, p. 1). These words are a reminder that the cultural and technological status associated with civilization is enabled by systems of writing. Indeed, the term *prehistory*, as applied to humans, dates from the arrival of human-like beings, millions of years ago, to the emergence of writing. It is writing that presaged *recorded history*, beginning a mere six thousand years ago.

In HCI, our interest in language is primarily in systems of writing and in the technology that enables communication in a written form. *Text* is the written material on a page or display. How it gets there is a topic that intrigues and challenges HCI researchers, as well as the engineers and designers who create products that support text creation, or text entry. Although text entry is hugely important in HCI, our interest here is language itself in a written form.

One way to characterize and study a language in its written form is through a corpus—a large collection of text samples gathered from diverse and representative sources such as newspapers, books, e-mails, and magazines. Of course, it is not possible for a corpus to broadly yet precisely represent a language. The sampling process brings limitations: During what timeframe were the samples written? In what country? In what region of the country? On what topics are the samples focused and who wrote them? A well-known corpus is the British National Corpus

Word Rank	English	French	German	Finnish	SMS English	SMS Pinyin
1	the	de	der	ja	u	wo (我)
2	of	la	die	on	i	ni (你)
3	and	et	und	ei	to	le (了)
4	a	le	in	että	me	de (的)
5	in	à	den	oli	at	bu (不)
…	…	…	…	…	…	…
1000	top	ceci	konkurrenz	muista	ps	jiu (舅)
1001	truth	mari	stieg	paikalla	quit	tie (贴)
1002	balance	solution	notwendig	varaa	rice	ji (即)
1003	heard	expliquer	sogenannte	vie	sailing	jiao (角)
1004	speech	pluie	fahren	seuran	sale	ku (裤)
…	…	…	…	…	…	…

FIGURE 2.19

Sample words from word-frequency lists in various languages.

(BNC), which includes samples totaling 100 million words.[8] The sources are written in British English and are from the late 20th century. So analyses gleaned from the BNC, while generally applicable to English, may not precisely apply, for example, to American English, to present day English, or to the language of teenagers sending text messages.

To facilitate study and analysis, a corpus is sometimes reduced to a word-frequency list, which tabulates unique words and their frequencies in the corpus. One such reduction of the BNC includes about 64,000 unique words with frequencies totaling 90 million (Silfverberg, MacKenzie, and Korhonen, 2000). Only words occurring three or more times in the original corpus are included. The most frequent word is *the*, representing about 6.8 percent of all words.

Figure 2.19 includes excerpts from several corpora, showing the five most frequently used words and the words ranked from 1000 to 1004. The English entries are from the British National Corpus. There are additional columns for French (New, Pallier, Brysbaert, and Ferrand, 2004), German (Sporka et al., 2011), Finnish, SMS English, and SMS Pinyin (Y. Liu and Räihä, 2010). The Finnish entries are from a database of text from a popular newspaper in Finland, *Turun Sanomat*. The SMS English entries are from a collection of about 10,000 text messages, mostly from students at the University of Singapore.[9] SMS text messaging is a good example of the dynamic and context-sensitive nature of language. Efforts to characterize SMS English are prone to the limitations noted above. Note that there is no overlap in the entries 1–5 under English and SMS English.

The right-hand column in Figure 2.19 is for SMS Pinyin. Pinyin has been the standard coding system since 1958, using the Latin alphabet for Mandarin Chinese characters. The entries are pinyin marks, not words. Each mark maps to the Chinese

[8] See www.natcorp.ox.ac.uk.

[9] Available at www.comp.nus.edu.sg/~rpnlpir/smsCorpus.

```
THE ROOM WAS NOT VERY LIGHT A SMALL OBLONG
----ROO------NOT-V-----I------SM----OB----

READING LAMP ON THE DESK SHED GLOW ON
REA----------O------D----SHED-GLO--0-

POLISHED WOOD BUT LESS ON THE SHABBY RED CARPET
P-L-S-----O---BU--L-S--O-------SH----RE--C-----
```

FIGURE 2.22

Shannon's letter-guessing experiment.

(Adapted from Shannon, 1951)

beginning. As guessing proceeds, the phrase is revealed to the participant, letter by letter. The results are recorded as shown in the line below each phrase in the figure. A dash ("-") is a correct guess; a letter is an incorrect guess. Shannon called the second line the "reduced text." In terms of redundancy and entropy, a dash represents redundancy (what is known), while a letter represents entropy (what is not known). Among the interesting observations in Figure 2.22 is that errors are more common at the beginning of words, less common as words progress. The statistical nature of the language and the participant's inherent understanding of the language facilitate guessing within words.

The letter-guessing experiment in Figure 2.22 is more than a curiosity. Shannon was motivated to quantify the entropy of English in information-theoretic terms. He pointed out, for example, that both lines in each phrase-pair contain the same information in that it is possible, with a good statistical model, to recover the first line from the second. Because of the redundancy in printed English (viz. the dashes), a communications system need only transmit the reduced text. The original text can be recovered using the statistical model. Shannon also demonstrated how to compute the entropy of printed English. Considering letter frequencies alone, the entropy is about 4.25 bits per letter.[11] Considering previous letters, the entropy is reduced because there is less uncertainty about forthcoming letters. Considering long range statistical effects (up to 100 letters), Shannon estimated the entropy of printed English at about one bit per letter with a corresponding redundancy of about 75 percent.

See also student exercise 2-2 at the end of this chapter.

2.7 Human performance

Humans use their sensors, brain, and responders to do things. When the three elements work together to achieve a goal, human performance arises. Whether the

[11]The data set and calculation are given in Chapter 7 (see Figure 7.19).

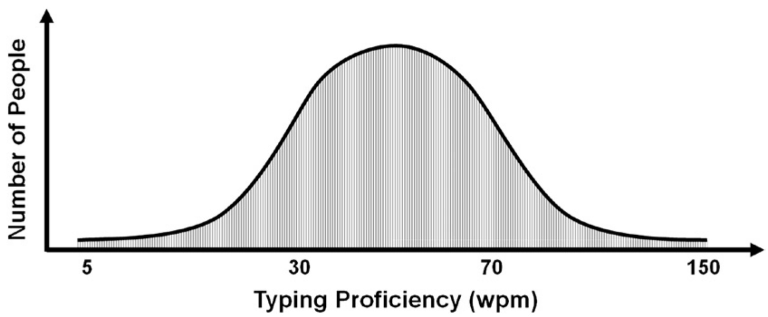


FIGURE 2.23

Variability of people in performing a task such as typing.

action is tying shoelaces, folding clothes, searching the Internet, or entering a text message on a mobile phone, human performance is present. Better performance is typically associated with faster or more accurate behavior, and this leads to a fundamental property of human performance—the *speed-accuracy trade-off*: go faster and errors increase; slow down and accuracy improves. Reported in academic papers dating back more than a century (see Swensson, 1972, for a review), mundane and proverbial ("Haste makes waste"), and steeped in common sense (we instinctively slow down to avoid errors), it is hard to imagine a more banal feature of human performance. Clearly, research on a new interface or interaction technique that seeks to determine the speed in doing a task must consider accuracy as well.

Humans position themselves on the speed-accuracy trade-off in a manner that is both comfortable and consistent with their goals. Sometimes we act with haste, even recklessly; at other times we act with great attention to detail. Furthermore, we may act in the presence of a secondary task, such as listening to the radio, conversing with a friend, or driving a car. Clearly, context plays an important role, as do the limits and capabilities of the sensors, the brain, and the responders.

With human performance, we begin to see complexities and challenges in HCI that are absent in traditional sciences such as physics and chemistry. Humans bring diversity and variability, and these characteristics bring imprecision and uncertainty. Some humans perform tasks better than others. As well, a particular human may perform a task better in one context and environment than when performing the same task in a different context and environment. Furthermore, if that same human performs the same task repeatedly in the same context and environment, the outcome will likely vary.

Human diversity in performing tasks is sometimes illustrated in a distribution, as in Figure 2.23. Here the distribution reveals the number of people performing a task (y-axis) versus their proficiency in doing it (x-axis). The example assumes computer users as the population and illustrates typing on a conventional computer keyboard as the task. Most people fall somewhere in the middle of the distribution.

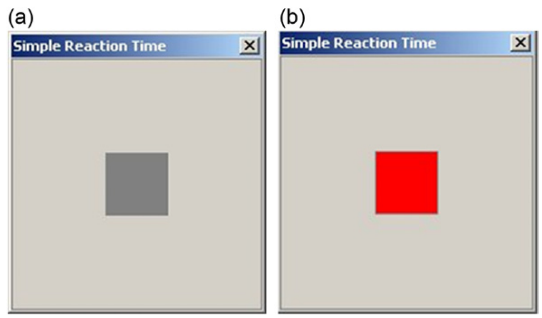


FIGURE 2.24

Simple reaction time: (a) The user fixates on the grey box. (b) After a delay, the box turns red whereupon the user presses a key as quickly as possible.

Typing speeds here are in the range of, say, 30–70 words per minute. Some people are slower, some faster. However, a small number of people will be exceedingly fast, say, 150 words per minute or faster. Yet others, also a small number, exhibit difficulty in achieving even a modest speed, such as 5 words per minute, equivalent to one word every 12 seconds.

2.7.1 Reaction time

One of the most primitive manifestations of human performance is *simple reaction time*, defined as the delay between the occurrence of a single fixed stimulus and the initiation of a response assigned to it (Fitts and Posner, 1968, p. 95). An example is pressing a button in response to the onset of a stimulus light. The task involves the three elements of the human shown in Figure 2.17. The cognitive operation is trivial, so the task is relatively easy to study. While the apparatus in experimental settings is usually simple, humans react to more complex apparatus all the time, in everyday pursuits and in a variety of contexts, such as reacting to the ring of a phone, to a traffic light, or to water in a bath (hot!). These three examples all involve a motor response. But the sensory stimuli differ. The ring of a phone is an auditory stimulus; a changing traffic light is a visual stimulus; hot water touching the skin is a tactile stimulus. It is known that simple reaction times differ according to the stimulus source, with approximate values of 150 ms (auditory), 200 ms (visual), 300 ms (smell), and 700 ms (pain) (Bailey, 1996, p. 41).

To explore reaction times further, a Java-based application was developed to experimentally test and demonstrate several reaction time tasks.[12] (See also Appendix A.) After describing each task, the results of an experiment are presented. For *simple reaction*, the interface is shown in Figure 2.24. A trial begins with the

[12]The software, a detailed API, and related files are in `ReactionTimeExperiment.zip`, available on this book's website.

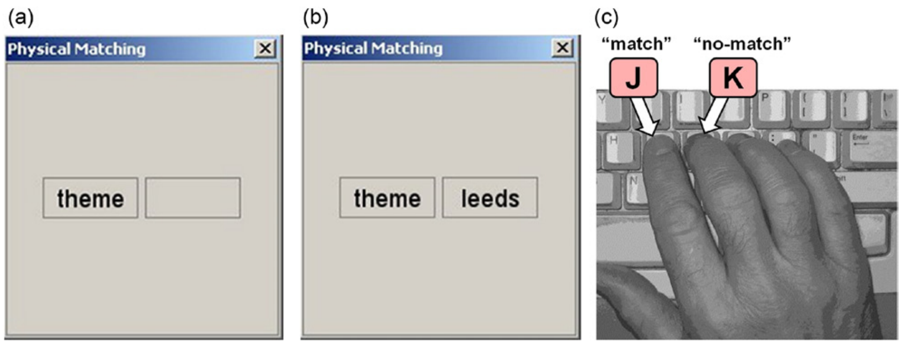


FIGURE 2.25

Physical matching: (a) Initial stimulus. (b) After a delay, a second stimulus appears. (c) Setup.

appearance of a grey box in a GUI window. Following a delay, the box turns red (color is not apparent in grayscale print). This is the sensory stimulus. The user's goal is to press a key on the system keyboard as quickly as possible after the stimulus appears. The delay between the grey box appearing and the box turning red is randomized to prevent the user from anticipating the onset of the stimulus.

The software implements three extensions of simple reaction tasks: physical matching, name matching, and class matching. Each adds a layer of complexity to the cognitive operation. The tasks were modeled after descriptions by Card et al. (1983, 65–71). For *physical matching*, the user is presented with a five-letter word as an initial stimulus. After a delay a second stimulus appears, also a five-letter word. The user responds as quickly as possible by pressing one of two keys: a "match" key if the second stimulus matches the first stimulus, or a "no-match" key if the second stimulus differs from the first stimulus. Matches occur with 50 percent probability. An example experimental setup is shown in Figure 2.25.

Obviously, physical matching is more complicated than simple reaction, since the user must compare the stimulus to a code stored in working memory. There are many examples of similar tasks in HCI, such as entering text on a mobile phone using predictive input (*T9*). While entering a word, the user has in her or his mind an intended word. This is the initial stimulus. With the last keystroke, the system presents a word. This is the second stimulus. If the presented word matches the intended word, the user presses 0 to accept the word. If the presented word does not match the intended word, the user presses * to retrieve the next alternative word matching the key sequence. (Details vary depending on the phone.)

Name matching is the same as physical matching except the words vary in appearance: uppercase or lowercase, mono-spaced or sans serif, plain or bold, 18 point or 20 point. A match is deemed to occur if the words are the same, regardless of the look of the fonts. See Figure 2.26. Name matching should take longer than physical matching because "the user must now wait until the visual code has been

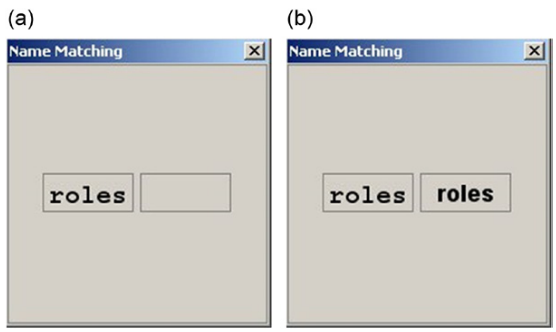


FIGURE 2.26

Name matching: (a) Initial stimulus. (b) Second stimulus.

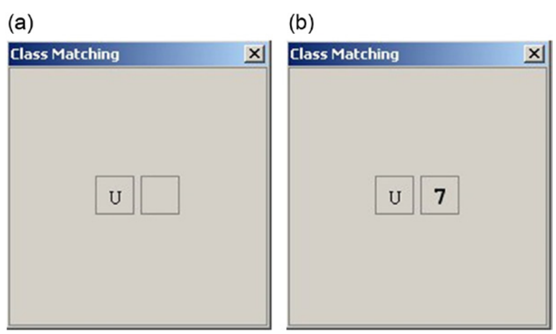


FIGURE 2.27

Class matching: (a) Initial stimulus. (b) Second stimulus.

recognized and an abstract code representing the name of the letter is available" (Card et al., 1983, p. 69).

For *class matching*, the initial stimulus contains a letter or digit. After a delay a second stimulus appears, also containing a letter or digit. The font is mono-spaced or sans serif, plain or italic, 18 point or 20 point. A match is deemed to occur if both symbols are of the same class; that is, both are letters or both are digits. Class matching takes longer still, because "the user has to make multiple references to long-term memory" (Card et al., 1983, p. 70). To avoid confusion, 0 (digit) and O (letter) are not included, nor are 1 (digit) and I (letter). (See Figure 2.27.)

The interfaces described above were tested in the lab component of a course on HCI. Fourteen students served as participants and performed three blocks of ten trials for each condition. The first block was considered practice and was discarded. To offset learning effects, participants were divided into two groups of equal size. One group

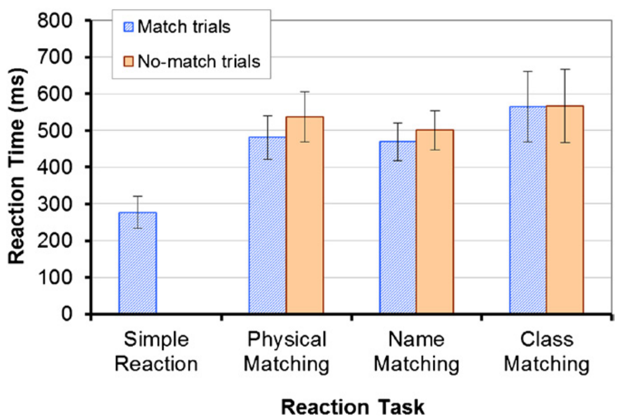


FIGURE 2.28

Results of an experiment comparing several reaction tasks. Error bars show ±1 *SD*.

preformed the simple reaction task first, followed in order by the physical, name, and class matching tasks. The other group performed the tasks in the reverse order.

The results are shown in Figure 2.28. The mean time for simple reaction was 276 ms. This value is nicely positioned in the 113 to 528 ms range noted earlier for reaction time tasks (see Figure 2.17). Note that the time measurement began with the arrival of the second stimulus and ended with the key event registered in the software when a key was pressed; thus, the measurement includes the time for the motor response.

Physical matching took about twice as long as simple reaction, depending on whether the second stimulus was a match (482 ms) or a no-match (538 ms). Interestingly enough, name matching did not take longer than physical matching. One explanation is that the words in the name-matching task had insufficient variability in appearance to require additional cognitive processing. Class matching was the hardest of the tasks, with means of about 565 ms for both the match and no-match conditions.

Choice reaction is yet another type of reaction time task. In this case, the user has *n* stimuli, such as lights, and *n* responders, such as switches. There is a one for one correspondence between stimulus and response. Choice reaction time is discussed in Chapter 7 on modeling.

2.7.2 Visual search

A variation on reaction time is *visual search*. Here, the user scans a collection of items, searching for a desired item. Obviously, the time increases with the number of items to scan. The software described above includes a mode for visual search, with the search space configurable for 1, 2, 4, 8, 16, or 32 items. An example for $N=16$ is shown in Figure 2.29. The initial stimulus is a single letter. After a random

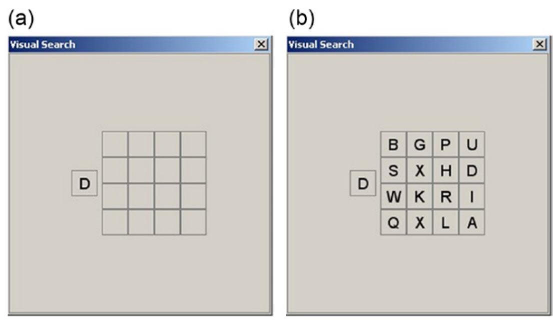


FIGURE 2.29

Visual search: (a) Initial stimulus. (b) After a delay a collection of letters appears.

delay of two to five seconds, the squares on the right are populated with letters selected at random. The initial stimulus appears on the right with 50 percent probability. The user presses a “match” or “no-match” key, as appropriate.

A small experiment was conducted with the same 14 students from the experiment described above, using a similar procedure. The results are shown in Figure 2.30 in two forms. In (a), reaction time (*RT*) versus number of items (*N*) is plotted. Each marker reveals the mean of $14 \times (10 + 10) = 280$ trials. The markers are connected and a linear regression line is superimposed. At $R^2 = .9929$, the regression model is an excellent fit. Clearly, there is a linear relationship between reaction time in a visual search task and the number of items to scan. This is well known in the HCI literature, particularly from research on menu selection (e.g., Cockburn, Gutwin, and Greenberg, 2007; Hornof and Kieras, 1997; Landauer and Nachbar, 1985). For this experiment,

$$RT = 498 + 41 \times N \text{ ms} \tag{1}$$

$N=1$ is a special case since there is only one item to scan. This reduces the task to physical matching. The task is slightly different than in the physical matching experiment, since the user is matching a letter rather than a word. Nevertheless, the result is consistent with the physical matching result in Figure 2.28 ($RT \approx 500$ ms).

In Figure 2.30b, the results are given separately for the match trials and the no-match trials. The no-match trials take longer. The reason is simple. If the initial stimulus is not present, an exhaustive search is required to determine such before pressing the no-match key. If the initial stimulus is present, the user presses the match key immediately when the initial stimulus is located in the right-side stimuli. The effect only surfaces at $N=16$ and $N=32$, however.

Before moving on, here is an interesting reaction time situation, and it bears directly on the title of this section, Human Performance. Consider an athlete competing in the 100 meter dash in the Olympics. Sometimes at the beginning of a race there is a “false start.” The definition of a false start is rather interesting: a false start occurs if an athlete reacts to the starter’s pistol before it is sounded *or within*

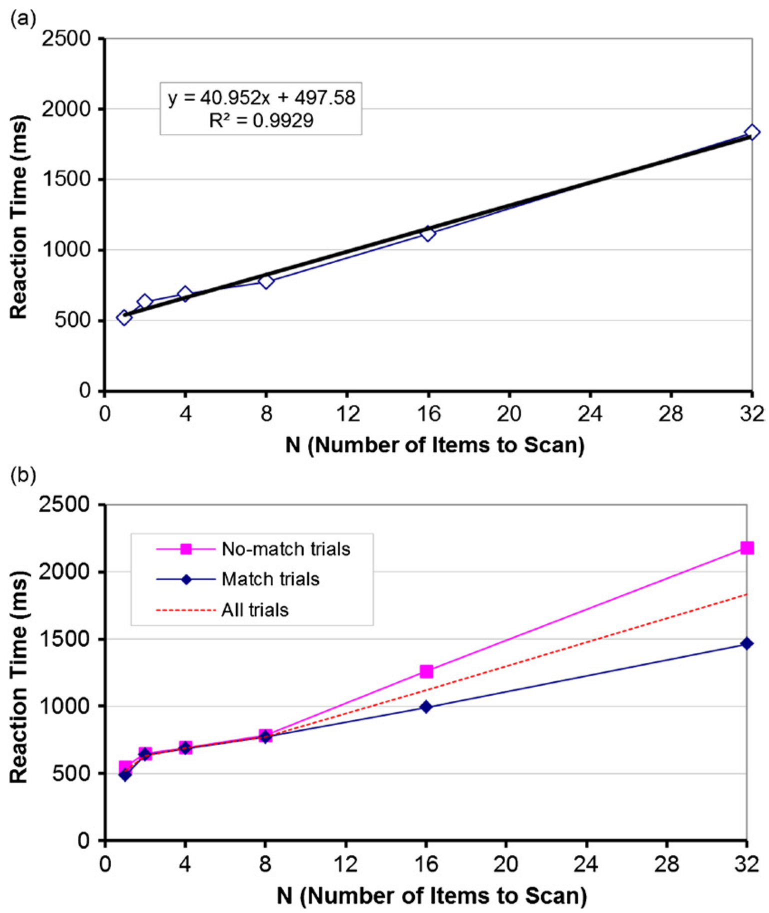


FIGURE 2.30

Results of visual search experiment: (a) Overall result with linear regression model. (b) Results by match and no-match trials.

100 ms after.[13] Clearly, an athlete who reacts before the starter's pistol sounds is anticipating, not reacting. Interesting in the definition, however, is the criterion that a false start has occurred if the athlete reacts within 100 ms *after* the starter's pistol is sounded. One hundred milliseconds is precariously close to the lower bound on reaction time, which is cited in Figure 2.17 as 113 ms. Card et al. peg the lower bound at 105 ms (Card et al., 1983, p. 66). World records are set, and gold medals won, by humans at the extreme tails of the normal distribution. Is it possible that

[13]Rule 161.2 of the International Association of Athletics Federations (IAAF) deems a false start to occur "when the reaction time is less than 100/1000ths of a second." See www.iaaf.org/mm/Document/imported/42192.pdf (107).

a false start is declared occasionally, very occasionally, when none occurred (e.g., honestly *reacting* 95 ms after the starter's pistol is fired)? There are slight differences between the lower-bound reaction times cited above and the false-start scenario, however. The values cited are for pressing a key with a finger in response to a visual stimulus. The motor response signals in the 100 meter dash must travel farther to reach the feet. This tends to lengthen the reaction time. Also, the stimulus in the 100 meter dash is auditory, not visual. Auditory reaction time is less than visual reaction time, so this tends to shorten the reaction time. Nevertheless, the example illustrates the application of low-level research in experimental psychology to human performance and to the design of human-machine systems.

2.7.3 Skilled behavior

The response time tasks in the previous section are simple: a sensory stimulus initiates a simple cognitive operation, which is followed by a simple motor response. It takes just a few trials to get comfortable with the task and with additional practice there is little if any improvement in performance. However, in many tasks, human performance improves considerably and continuously with practice. For such tasks, the phenomenon of learning and improving is so pronounced that the most endearing property of the task is the progression in performance and the level of performance achieved, according to a criterion such as speed, accuracy, degree of success, and so on. *Skilled behavior*, then, is a property of human behavior whereby human performance necessarily improves through practice. Examples include playing darts, playing chess and, in computing scenarios, gaming or programming. One's ability to do these tasks is likely to bear significantly on the amount of practice done.

The examples just cited were chosen for a reason. They delineate two categories of skilled behavior: sensory-motor skill and mental skill (Welford, 1968, p. 21). Proficiency in darts or gaming is likely to emphasize sensory-motor skill, while proficiency in chess or computer programming is likely to emphasize mental skill. Of course, there is no dichotomy. All skilled behavior requires mental faculties, such as perception, decision, and judgment. Similarly, even the most contemplative of skilled tasks requires coordinated, overt action by the hands or other organs.

While tasks such as gaming and computer programming may focus on sensory-motor skill or mental skill, respectively, other tasks involve considerable elements of both. Consider a physician performing minimally invasive surgery, as is common for abdominal procedures. To access the abdominal area, a camera and a light mounted at the end of a laparoscope are inserted through a small incision, with the image displayed on an overhead monitor. Tools are inserted through other incisions for convenient access to an internal organ. The surgeon views the monitor and manipulates the tools to grasp and cut tissue. In Figure 2.31a, the tips of the surgeon's tools for grasping (left) and cutting (top) are shown as they appear on a monitor during a cholecystectomy, or gallbladder removal. The tools are manually operated, external to the patient. Figure 2.31b shows examples of such tools in a training simulator. The tools are complex instruments. Note, for example, that the tips of the tools

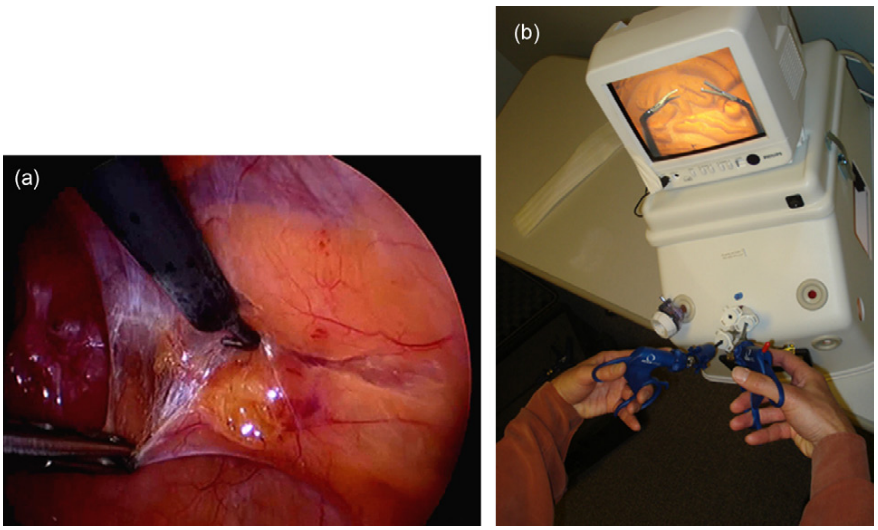


FIGURE 2.31

Sensory-motor skill combined with mental skill during laparoscopic surgery: (a) Tips of tools for grasping and cutting. (b) Exterior view of tools and monitor in a training simulator.

(Photos courtesy of the Centre of Excellence for Simulation Education and Innovation at Vancouver General Hospital)

articulate, or bend, thus providing an additional degree of freedom for the surgeon (Martinec, Gatta, Zheng, Denk, and Swanstrom, 2009). Clearly, the human-machine interaction involves both sensory-motor skill (operating the tools while viewing a monitor) and mental skill (knowing what to do and the strategy for doing it).

One way to study skilled behavior is to record and chart the progression of skill over a period of time. The level of skill is measured in a dependent variable, such as speed, accuracy, or some variation of these. The time element is typically a convenient procedural unit such as trial iteration, block or session number, or a temporal unit such as minutes, hours, days, months, or years. Measuring and modeling the progression of skill is common in HCI research, particularly where users confront a new interface or interaction technique. The methodology for evaluating skilled behavior is presented in Chapter 5 (see Longitudinal Studies) with the mathematical steps for modeling presented in Chapter 7 (see Skill Acquisition). See also student exercise 2-4 at the end of this chapter.

2.7.4 Attention

Texting while driving. It's hard to imagine a more provocative theme to open this discussion on attention. Although driving a car is relatively easy, even the most experienced driver is a potential killer if he or she chooses to read and send text messages while driving. The problem lies in one's inability to attend to both tasks simultaneously. Much like the bottleneck posed by working memory (7 ± 2 items), the human ability to attend is also limited. But what is the limit? More fundamentally, what is attention? Which tasks require attention? Which do not? How is human performance impacted? According to one view, attention is a property of human behavior that occurs when a person who is attending to one thing cannot attend to another (Keele, 1973, p. 4). Typing, for example, requires attention because while typing we cannot engage in conversation. On the other hand, walking requires very little attention since we can think, converse, and do other things while walking. One way to study attention is to observe and measure humans performing two tasks separately and then to repeat the procedure with the two tasks performed simultaneously. A task with performance that degrades in the simultaneous case is said to require attention.

Attention is often studied along two themes: *divided attention* and *selected attention* (B. H. Kantowitz and Sorkin, 1983, p. 179). Divided attention is the process of concentrating on and doing more than one task at time. Texting while driving is an example, and the effect is obvious enough. In other cases, divided attention poses no problem, as in walking and talking. Selected attention (aka *focused attention*) is attending to one task to the exclusion of others. For example, we converse with a friend in a crowded noise-filled room while blocking out extraneous chatter. But there are limits. In that same conversation we are occasionally unable to recall words just spoken because our attention drifted away or was pulled away by a distraction. Selective attention, then, is the human ability to ignore extraneous events and to maintain focus on a primary task. One theory of selective attention holds that our ability to selectively attend bears on the importance of the events to the individual. A person listening to a speech is likely to stop listening if the person's name is spoken from another location (Keele, 1973, p. 140). One's own name is intrinsically important and is likely to intrude on the ability to selectively attend to the speech. Clearly, importance is subjective. Wickens gives an example of an airplane crash where the flight crew were preoccupied with a malfunction in the cockpit that had no bearing on the safety of the flight (Wickens, 1987, p. 249). The crew attended to the malfunction while failing to notice critical altimeter readings showing that the airplane was gradually descending to the ground. The malfunction was of salient importance to the flight crew.

The distinction between divided and selected attention is often explained in terms of channels (Wickens, 1987, p. 254). Events in a single channel (e.g., visual, auditory, motor) are processed in parallel, whereas events in different channels are processed in serial. When processing events in parallel (single channel) one event may intrude on the ability to focus attention on another event. When processing events in serial (different channels), we strive to focus on one event to the exclusion of others or to divide attention in a convenient manner between the channels.

Analyzing accidents is an important theme in human factors, as the aviation example above illustrates, and there is no shortage of incidents. Accidents on the road, in the air, on the seas, or in industry are numerous and in many cases the cause is at least partly attributable to the human element—to distractions or to selectively attending to inappropriate events. One such accident involving a driver and a cyclist occurred because a Tamagotchi digital pet distracted the driver.[14] Evidently, the pet developed a dire need for "food" and was distressed: *bleep, bleep, bleep, bleep, bleep*. The call of the pet was of salient importance to the driver, with a horrific and fatal outcome (Casey, 2006, pp. 255–259). More likely today, it is the call of the mobile phone that brings danger. The statistics are shocking, yet unsurprising—a 23-fold increase in the risk of collision while texting (Richtel, 2009).

Attention has relevance in HCI in for example, office environments where interruptions that demand task switching affect productivity (Czerwinski, Horvitz, and Wilhite, 2004). The mobile age has brought a milieu of issues bearing on attention. Not only are attention resources limited, these resources are engaged while users are on the move. There is a shift toward immediate, brief tasks that demand constant vigilance and user availability, with increasingly demanding expectations in response times. So-called psychosocial tasks compete for and deplete attention resources, with evidence pointing to an eventual breakdown of fluency in the interaction (Oulasvirta, Tamminen, Roto, and Kuorelahti, 2005).

2.7.5 Human error

Human error can be examined from many perspectives. In HCI experiments testing new interfaces or interaction techniques, errors are an important metric for performance. An error is a discrete event in a task, or trial, where the outcome is incorrect, having deviated from the correct and desired outcome. The events are logged and analyzed as a component of human performance, along with task completion time and other measurable properties of the interaction. Typically, errors are reported as the ratio of incorrectly completed trials to all trials, and are often reported as a percent (×100). Sometimes accuracy is reported—the ratio of correctly completed trials to all trials.

Two examples for computing tasks are shown in Figure 2.32. A GUI target selection task is shown on the left in two forms. The top image shows the goal: moving a tracking symbol from a starting position to a target and ending with a select operation. The bottom image shows an error, since the final selection was outside the target. A text entry task is shown on the right. The goal of entering the word *quickly* is shown correctly done at the top. The bottom image shows an error, since the word was entered incorrectly.

Mishaps and miscues in human performance are many. Often, a simple categorization of the outcome of a task as correct or incorrect falls short of fully capturing the behavior. We need look no further than Figure 2.32 for examples. Not only were the results of the tasks on the bottom erroneous in a discrete sense, there were additional behaviors that deviated from perfect execution of the tasks. For the target selection error, the tracking symbol veered off the direct path to the target. For the text entry error, it appears that at least part of the word was correctly entered.

Taking a broader perspective, human error is often studied by examining how and why errors occur. Once again, Figure 2.32 provides insight. In the erroneous

[14] Ample descriptions of the Tamagotchi are found in various online sources (search using "Tamagotchi").

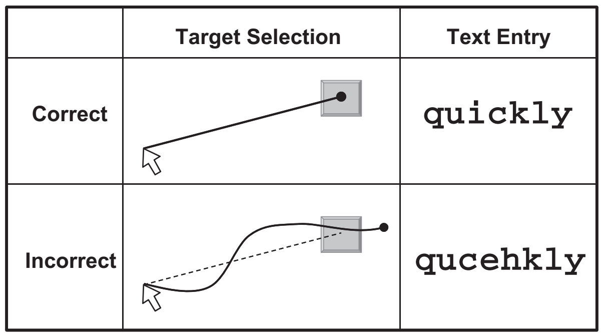


FIGURE 2.32

Common computing tasks completed correctly (*top*) and incorrectly (*bottom*).

target selection task, was there a control problem with the input device? Was the device's gain setting too sensitive? Was the device a mouse, a touchpad, an eye tracker, a game controller, or some other input control? Note as well that the tracking symbol entered then exited the target. Was there a problem with the final target acquisition in the task? In the erroneous text entry task, if input involved a keyboard, were errors due to the user pressing keys adjacent to correct keys? Were the keys too small? If entry involved gestural input using a finger or stylus on a digitizing surface, did the user enter the wrong gesture or an ill-formed gesture? Was the digitizing surface too small, awkwardly positioned, or unstable? Clearly, there are many questions that arise in developing a full understanding of how and why errors occur. Note as well that the questions above are not simply about the human; they also question aspects of the device and the interaction.

An even broader perspective in analyzing errors may question the environmental circumstances coincident with the tasks. Were users disadvantaged due to noise, vibration, lighting, or other environmental conditions? Were users walking or performing a secondary task? Were they distracted by the presence of other people, as might occur in a social setting?

Human factors researchers often examine human error as a factor in industrial accidents where the outcome causes substantial damage or loss of life. Such events rarely occur simply because a human operator presses the wrong button, or commits an interaction error with the system or interface. Usually, the failures are systemic—the result of a confluence of events, many having little to do with the human.

To the extent that a significant accident is determined to have resulted from *human error*, a deeper analysis is often more revealing. Casey's retelling of dozens of such accidents leads to the conclusion that the failures are often *design-induced errors* (Casey, 1998, p. 2006). This point is re-cast as follows: if a human operator mistakenly flicks the wrong switch or enters an incorrect value, and the action results in a serious accident, is the failure due to human error? Partly so, perhaps, but clearly the accident is enabled by the design of whatever he or she is operating. A design that can lead to catastrophic outcomes purely on the basis of an operator's interaction error is a faulty