Laura was running late for an appointment in a large, unfamiliar city and relied on her new navigation device to guide her. She had read the somewhat confusing instructions and realized the importance of the voice display mode so that she could hear the directions to her destination without taking her eyes off the road. She had reminded herself to activate it before she got into heavy traffic, but the traffic suddenly increased, and she realized that she had forgotten to do so. Being late, however, she did not pull over but tried to remember the sequence of mode switches necessary to activate the voice mode. She couldn't get it right, but she managed to activate the electronic map. However, transposing its north-up representation to accommodate her south-bound direction of travel was too confusing. Finally lost, she pulled out her cellular phone to call her destination, glanced at the number she had written down, 303-462-8553, and dialed 303-462-8533. Getting no response, she became frustrated. She looked down to check the number and dial it carefully. Unfortunately, she did not see the car rapidly converging along the entrance ramp to her right, and only at the last moment the sound of the horn alerted her that the car was not yielding. Slamming on the brakes, heart beating fast, she pulled off to the side to carefully check her location, read the instructions, and place the phone call in the relative safety of the roadside.

Each day, we process large amounts of information from our environment to accomplish various goals and make our way successfully through the world. The previous illustration represents a typical problem that one might experience because of a poor match between man-made equipment (or the environment) and the human information-processing system. Sometimes these mismatches cause misperceptions, and sometimes people just experience memory failures. While the scenario described above may seem rather mundane, there are dozens of other cases where difficulties result in injury or death (Casey, 1993; Wickens & Hollands, 2000). Some of these cases are discussed in Chapter 14 on safety. In
this chapter, we consider the basic mechanisms by which people perceive, think, and remember, processes generally grouped under the label of cognition, and we provide a framework for understanding how such information is processed. As we learn about the various limitations of the human cognitive system, we consider the implications of, and some solutions for, design problems.

The human information-processing system is conveniently represented by different stages at which information gets transformed: (1) perception of information about the environment, (2) central processing or transforming that information, and (3) responding to that information. We highlight the first and second stages as the processes involved in cognition and most typically represented in the study of applied cognitive psychology (Durso, 1999), although we present a more elaborate picture than the simple three-stage model. This chapter then picks up where our discussions of the more sensory aspects of auditory and visual processing left off in the previous two chapters. In Chapter 7, we describe more complex cognitive processes that form the basis of decision making, in Chapter 8 we discuss the implications of perception and cognition for display design, and in Chapter 9 we discuss the implications for control. Finally, our discussions of memory have many direct implications for learning, as discussed in Chapter 18.

**INFORMATION PROCESSING MODELS**

Shown in Figure 6.1 is a model of information processing that highlights those aspects that typically influence cognition: perceiving, thinking about, and understanding the world. The senses, shown to the left of the figure, gather information, which is then perceived, providing a meaningful interpretation of what is sensed as aided by prior knowledge, through a mechanism that we described in Figure 4.6 as top-down processing. This prior knowledge is stored in long-term memory.

Sometimes, perception leads directly to the selection and execution of a response, as when the driver swerved to avoid the converging car in the opening story. Quite often, however, an action is delayed, or not executed at all, as we "think about" or manipulate perceived information in working memory. This stage of information processing plays host to a wide variety of mental activities that are in our consciousness, such as rehearsing, planning, understanding, visualizing, decision making, and problem solving. Working memory is a temporary, effort-demanding store. One of the activities for which working memory is used is to create a more permanent representation of the information in long-term memory, where it may be retrieved minutes, hours, days, or years later. These are the processes of learning (putting information into long-term memory) and retrieval. As we see in the figure, information from long-term memory is retrieved every time we perceive familiar information.

At the top of the figure we note that many of the stages of information processing depend upon mental or cognitive resources, a sort of pool of attention or mental effort that is of limited availability and can be allocated to processes as required. In particular, the figure highlights an important distinction that has
been quite visible in the research on attention. On the left, we see the role of attention, the limited resources in selecting sensory channels for further information processing, as when our eyes focus on one part of the world, and not another. In contrast, the other dashed arrows suggest the role of attention in supporting all aspects of performance as well as in dividing attention between tasks. These two aspects of attention, selection and division, are treated separately in this chapter.

Finally, we note the feedback loop. Our actions often generate new information to be sensed and perceived. The sequence of information processing may start anywhere. For example, sometimes we initiate an action from a decision with no perception driving it. We then may evaluate the consequence of that decision later, through sensation and perception. We consider the importance of this closed feedback loop in Chapter 9.

**SELECTIVE ATTENTION**

Laura was not attending to the roadway at the time she was looking at her cell phone, and this failure of selective attention nearly caused an accident. We shall see in Chapter 17 that failures of attention, many of them selective, are the major cause of automobile accidents (Malaterre, 1990). Correspondingly, the cause of the greatest number of fatal accidents in commercial aviation, controlled flight
into terrain, when a pilot flies a perfectly good airplane into the ground, represents a failure of selective attention to all those sources of information regarding the plane’s altitude above the ground (Phillips, 2001; Wiener, 1977).

Selective attention does not guarantee perception, but it is usually considered necessary to achieve it. Stated in other terms, we normally look at the things we perceive and perceive the things we are looking at. We considered the role of visual scanning in selective attention in Chapter 4. While we do not have “ear-balls” that can index selective auditory attention as we have eyeballs in the visual modality, there is nevertheless a corresponding phenomenon in auditory selection. For example, we may tune our attention to concentrate on one conversation in a noisy workplace while filtering out the distraction of other conversations and noises.

The selection of channels to attend (and filtering of channels to ignore) is typically driven by four factors: salience, effort, expectancy, and value (Wickens et al., 2003, in press). They can be represented in the same contrasting framework of stimulus-driven bottom-up processes versus knowledge-driven top-down processes that we applied to perception in chapters 4 and 5. Salience is a bottom-up process, characterizing what is described as attentional capture (Yantis, 1993). The car horn, for example, clearly captured Laura’s attention. Salient stimulus dimensions are chosen by designers to signal important events via alarms and alerts. Abrupt onsets (Yantis, 1993), distinct stimuli and auditory stimuli (Spence & Driver, 2000; Banbury et al., 2001), and tactile stimuli (Sklar & Sarter, 1999) are particularly salient. In contrast, many events that do not have these characteristics may not be noticed, even if they are significant, a phenomenon known as change blindness or attentional blindness (Rensink 2002).

Expectancy and value together define what are characteristicaly called as top-down or knowledge-driven factors in allocating attention. That is, we tend to look at, or “sample,” the world where we expect to find information. Laura looked downward because she expected to see the phone number there. As an example in visual search, discussed in Chapter 4, a radiologist looks most closely at those areas of an x-ray plate most likely to contain an abnormality. Correspondingly, a pilot looks most frequently at the instrument that changes most rapidly because here is where the pilot expects to see change (Senders, 1964; Bellenkes et al., 1997). Conversely, the frequency of looking at or attending to channels is also modified by how valuable it is to look at (or how costly it may be to miss an event on a channel; Moray, 1986). This is why a trained airplane pilot will continue to scan the world outside the cockpit for traffic even if that traffic is rare (Wickens et al., 2002); the costs of not seeing the traffic (and colliding with it) are large.

Finally, selective attention may be inhibited if it is effortful. We prefer to scan short distances rather than long ones, and we often prefer to avoid head movements to select information sources. It is for this reason that drivers, particularly fatigued ones (who have not much “effort to give”), fail to look behind them to check their blind spot when changing lanes.

In addition to understanding the high frequency with which failures to notice may contribute to accidents (Jones & Endsley, 1996) and considering ways
of training selective attention (which we discuss at the end of this chapter), understanding bottom-up processes of attentional capture are important for the design of alarms (Woods, 1995; Chapters 5 & 8) and automated cueing (Chapter 16). Understanding the role of effort in inhibiting attention movement is important in both designing integrated displays (Chapter 8) and configuring the layout of workspaces (Chapter 10).

**Perception**

The most direct consequence of selective attention selection is perception, which involves the extraction of meaning from an array (visual) or sequence (auditory) of information processed by the senses. Our driver, Laura, eventually looked to the roadside (selection) and perceived the hazard of the approaching vehicle. Sometimes, meaning may be extracted (perception) without attention. In this way, our attention at a party can be "captured" in a bottom-up fashion when a nearby speaker utters our name even though we were not initially selecting that speaker. This classic phenomenon is sometimes labeled the "cocktail party effect" (Moray 1969; Wood & Cowan, 1995). Correspondingly, the driver may not be consciously focusing attention on the roadway, even though he is adequately perceiving roadway information enough to steer the car.

**Three Perceptual Processes**

To elaborate on our discussion in chapters 4 and 5, perception proceeds by three often simultaneous and concurrent processes: (1) bottom-up feature analysis, (2) unitization, and (3) top-down processing. The latter two are based on long-term memory, and all of them have different implications for design.

Perception proceeds by analyzing the raw features of a stimulus or event, whether it is a word (the features may be letters), a symbol on a map (the features may be the color, shape, size, and location), or a sound (the features may be the phonemes of the word or the loudness and pitch of an alarm). Every event could potentially consist of a huge combination of features. However, to the extent that past experience has exposed the perceiver to sets of features that occur together and their co-occurrence is familiar (i.e., represented in long-term memory), these sets are said to become unitized. The consequence of unitization is more rapid and automatic than perceptual processing. Thus, the difference between perceiving the printed words of a familiar and an unfamiliar language is that the former can be perceived as whole units, and their meaning is directly accessed (retrieved from long-term memory), whereas the latter may need to be analyzed letter by letter, and the meaning is more slowly and effortfully retrieved from long-term memory. This distinction between the effortless processing of feature analysis and the more automatic processing of familiar unitized feature combinations (whose combined representation is stored in long-term memory), can be applied to almost any perceptual experience, such as perceiving symbols and icons (see Chapter 8), depth cues (Chapter 4), or alarm sounds (Chapter 5).

Whether unitized or not, stimulus elements and events may be perceived in clear visual or auditory form (reading large text in a well-lighted room or hear-
ing a clearly articulated speech) or may be perceived in a degraded form. For a visual stimulus, short glances, tiny text, and poor illumination or contrast represent such degradation. For an auditory event, masking noise and low intensity or unfamiliar accents produce degradation. We describe these degrading characteristics as producing poor bottom-up processing. The successful perception of such stimuli or events are beneficial to the extent that they are unitized and familiar. However, the third aspect of perceptual processing, top-down processing, provides another support to offset the degradation of bottom-up processing.

Top-down processing may be conceived as the ability to correctly guess what a stimulus or event is, even in the absence of clear physical (bottom-up) features necessary to precisely identify it. Such guesses are based upon expectations, and these expectations are based upon past experience, which is, by definition, stored in long-term memory. That is, we see or hear what we expect to see or hear (see Fig 4.6). High expectations are based on events that we have encountered frequently in the past. They are also based on associations between the perceived stimulus or event, and other stimuli or events that are present in the same context and have been joined in past experience.

The concepts of frequency and context in supporting top-down processing can be illustrated by the following example. A status indicator for a piece of very reliable equipment can be either green, indicating normal operations, or red, indicating failure.

- Given our past experience of red and green in human-designed systems, the association of these two colors to their meaning is made fairly automatically.
- A brief glance at the light, in the high glare of the sun, makes it hard to see which color it is (poor bottom-up processing).
- The past high reliability of the system allows us to “guess” that it is green (top-down processing based upon frequency) even if the actual color is hard to make out. Hence, the quick glance perceives the light to be green.
- The sound of smooth running and good system output provides a context to amplify the “perception of greenness” (top-down processing based upon context).
- An abnormal sound gradually becomes evident. The context has now changed, and red becomes somewhat more expected. The same ambiguous stimulus (hard to tell the color) is now perceived to be red (changing context).
- Now a very close look at the light, with a hand held up to shield it from the sun, reveals that it in fact is red, (improved bottom-up processing), and it turns out that it was red all along. Perception had previously been deceived by expectancy.

We now consider two other examples of the interplay between, and complementarily of, bottom-up and top-down processing. As one example, in reading, bottom-up processing is degraded by speed (brief glances) as well as by legibility, factors discussed in Chapter 4. When such degradation is imposed, we can
read words more easily than random digit strings (phone numbers, basketball scores, or stock prices), because each word provides an expectancy-based context for the letters within, and when text is presented, the sentence words provide context for reading degraded words within. For example, if we read the sentence “Turn the machine off when the red light indicates failure” and find the fourth word to be nearly illegible (poor bottom-up), the context of the surrounding words allows us to guess that the word is probably “off.”

As a second example, scrolling messages viewed in a small window can present a tradeoff between bottom-up and top-down processing. If words within sentences are presented so that context is available, then it may be better to use small text

However, if random digits are to be displayed within the same space, top-down processing cannot assist perception, so we must maximize bottom-up processing by making the digits larger (and presenting fewer of them)

\[
\begin{array}{c}
72184 \\
64992.
\end{array}
\]

If a line in a phone book can be thought of as a space-limited “message window” in the above example, then the same analysis can be applied, and it makes better sense to display the phone number in a larger font than the name, because the name is more likely to provide contextual cues for its spelling. Furthermore, there are usually less serious consequences for failing to perceive the name correctly than for failing to perceive the phone number correctly. The latter will always lead to a dialing error. Like the digits in the phone number, the letters in an email address should also be larger, since the lack of standardization of email addresses (and the fact that many people don’t know the middle initial of an addressee) removes context that could otherwise help support top-down processing. In short,

Adam Humfac: Adamjhumfa@xxx.yyy 444-455-2995.

is a better design than is

Adam Humfac adamjhumfa@xxx.yyy 444-455-2995.

**Human Factors Guidelines in Perception**

The proceeding examples and others lead us to a few simple guidelines for supporting perception.

1. **Maximize bottom-up** processing (Chapters 4 and 5). This involves not only increasing visible legibility (or audibility of sounds), but also pay-
ing careful attention to confusion caused by similarity of message sets that could be perceived in the same context.

2. **Maximize** automaticity and **unitization** by using familiar perceptual representations (those encountered frequently in long-term memory). Examples include the use of familiar fonts and lowercase text (Chapter 4), meaningful icons (Chapter 8), and words rather than abbreviations.

3. Maximize top-down processing when bottom-up processing may be poor (as revealed by analysis of the environment and the conditions under which perception may take place), and when unitization may be missing (unfamiliar symbology or language). This can be done by providing the best opportunities for guessing. For example,

- Avoid confusions: Maximize discriminating features.
- Use a smaller vocabulary. This has a double benefit of improving guess rate and allowing the creation of a vocabulary with more discriminating features.
- Create context. For example, the meaning of “your fuel is low” is better perceived than that of the shorter phrase “fuel low,” particularly under noisy conditions (Simpson, 1976).
- Exploit redundancy. This is quite similar to creating context, but redundancy often involves direct repetition of content in a different format. For example, simultaneous display of a visual and auditory message is more likely to guarantee correct perception in a perceptually degraded environment. The phonetic alphabet exploits redundancy by having each syllable convey a message concerning the identity of a letter (alpha = a).
- When doing usability testing of symbols or icons, make sure that the context in which these will eventually be used is instated for the testing conditions (Wolff & Wogalter, 1998). This provides a more valid test of the effective perception of the icons.
- Be wary of the “conspiracy” to invite perceptual errors when the unexpected may be encountered and when bottom-up processing is low (as revealed by task and environmental analysis). Examples of such conditions are: flying at night and encountering unusual aircraft attitudes, which can lead to illusions; or driving at night and encountering unexpected roadway construction. In all cases, as top-down processing attempts to compensate for the bottom-up degradation, it encourages the perception of the expected, which will not be appropriate. Under such conditions, perception of the unusual must be supported by providing particularly salient cues. A special case here is the poor perception of negation in sentences. For example, “do not turn off the equipment” could perceived as “turn off the equipment” if the message is badly degraded, because our perceptual system appears to treat the positive meaning of the sentence as the “default” state of a message (Clark & Chase, 1972). We return to this issue in our discussion of comprehension and working memory. If negation is to be used, it should be highlighted.
One downside of the redundancy and context, as these are employed to support top-down processing, is that the **length of perceptual messages is increased**, thereby reducing the **efficiency of information transfer**. For example, "alpha" and "your fuel is low" both take longer to articulate than "A" and "fuel low" (although they do not necessarily take longer to understand). The printed message "failure" occupies more space than the letter "F" or a small red light. Thus, redundancy and context can help to gain perceptual accuracy, but at the expense of efficiency. This is a tradeoff that designers must explicitly consider by carefully analyzing the consequences of perceptual errors and the extent of environmental factors and stress factors that may degrade bottom-up processing. We consider these stress factors, such as a stress-induced speed-accuracy tradeoff, in more detail in Chapter 13.

**Conclusion**

Perception is assumed to be relatively automatic (but becomes less so as bottom-up processing is degraded and top-down and unitization processes become less effective). However, as the duration of the perceptual process increases, we speak less of perception and more of **comprehension**, which is less automatic. The border between perception and comprehension is a fuzzy one, although we usually think of **perceiving** and word, but **comprehending** a series of words that make up a sentence. As we shall see, comprehension, like perception, is very much driven by top-down processing, from past experience and long-term memory. However, comprehension tends to also rely heavily upon the capabilities of working **memory** in a way that perception does not.

**WORKING MEMORY**

Failures of memory occur for everyone—and relatively frequently (Schacter, 2001). Sometimes, the failures are trivial, such as forgetting a new password that you just created. Other times, memory failures are more critical. For example, in 1915 a railroad switchman at a station in Scotland forgot that he had moved a train to an active track. As a result, two oncoming trains used the same track and the ensuing crash killed over 200 people (Rolt, 1978).

The next few sections focus on the part of cognition that involves human memory systems. Substantial evidence shows that there are two very different types of memory storage. The first, **working memory** (sometimes termed **short-term memory**), is relatively transient and limited to holding a small amount of information that may be rehearsed or "worked on" by other cognitive transformations (Cowan, 2001; Baddeley, 1986, 1990). It is the temporary store that keeps information active while we are using it or until we use it. Some examples are looking up a phone number and then holding it in working memory until we have completed dialing, remembering the information in the first part of a sentence as we hear the later words and integrate them to understand the sentence meaning, "holding" subsums while we multiply two-digit numbers, and constructing an image of the way an intersection will look from a view on a
Working Memory holds two different types of information: verbal and spatial.

The other memory store, long-term memory, involves the storage of information after it is no longer active in working memory and the retrieval of the information at a later point in time. When retrieval fails from either working or long-term memory, it is termed forgetting. Conceptually, working memory is the temporary holding of information that is active, either perceived from the environment or retrieved from long-term memory, while long-term memory involves the relatively passive store of information, which is activated only when it needs to be retrieved. The limitations of working memory hold major implications for system design.

A Model of Working Memory

Working memory can be understood in the context of a model proposed by Baddeley (1986, 1990), consisting of three components. In this model, a central executive component acts as an attentional control system that coordinates information from the two "storage" systems.

The visuospatial sketchpad holds information in an analog spatial form (e.g., visual imagery) while it is being used (Logie, 1995). These images consist of encoded information that has been brought from the senses or retrieved from long-term memory. Thus, the air traffic controller uses the visual-spatial sketchpad to retain information regarding where planes are located in the airspace. This representation is essential for the controller if the display is momentarily lost from view. This spatial working-memory component is also used when a driver tries to construct a mental map of necessary turns from a set of spoken navigational instructions. Part of the problem that Laura had in using her north-up map to drive south into the city was related to the mental rotation in spatial working memory that was necessary to bring the map into alignment with the world out her windshield.

The phonological loop represents verbal information in an acoustical form (Baddeley, 1990). It is kept active, or “rehearsed,” by articulating words or sounds, either vocally or subvocally. Thus, when we are trying to remember a phone number, we subvocally sound out the numbers until we no longer need them.

Whether material is verbal (in the phonetic loop) or spatial (in the visuospatial sketchpad), our ability to maintain information in working memory is limited in four interrelated respects: how much information can be kept active (its capacity), how long it can be kept active, how similar material is to other elements of working memory and ongoing information processing, and how much attention is required to keep the material active. We describe each of these influences in turn.

Limits of Working Memory

Capacity. Researchers have defined the upper limit or the capacity of working memory to be around 7 ± 2 chunks of information (Miller, 1956), although even this limit may be somewhat optimistic (Cowan, 2001). A chunk is the unit of
working memory space, defined jointly by the physical and cognitive properties that bind items within the chunk together. Thus, the sequence of four unrelated letters, XFDU, consists of four chunks, as does the sequence of four digits, 8479. However, the four letters DOOR or the four digits 2004 consist of only one chunk, because these can be coded into a single meaningful unit. As a result, each occupies only one “slot” in working memory, and so our working memory could hold 7 (±2) words or familiar dates as well as 7 ±2 unrelated letters or digits.

What then binds the units of an item together to make a single chunk? As the examples suggest, it is familiarity with the links or associations between the units, a familiarity based upon past experience and therefore related to long-term memory. The operation is analogous to the role of unitization in perception, discussed earlier. As a child learns to read, the separate letters in a word gradually become unified to form a single chunk. Correspondingly, as the skilled expert gains familiarity with a domain, an acronym or abbreviation that was once several chunks (individual letters) now becomes a single chunk.

Chunking benefits the operations in working memory in several ways. First, and most directly, it reduces the number of items in working memory and therefore increases the capacity of working memory storage. Second, chunking makes use of meaningful associations in long-term memory, and this aids in retention of the information. Third, because of the reduced number of items in working memory, material can be more easily rehearsed and is more likely to be transferred to long-term memory (which then reduces load on working memory).

Chunks in working memory can be thought of as “memory units,” but they also have physical counterparts in that perceptual chunks may be formed by providing spatial separation between them. For example, the social security number 123456789 contains three physical chunks. Such physical chunking is helpful to memory, but physical chunking works best when it is combined with cognitive chunking. In order to demonstrate this, ask yourself which of the following would be the easiest to perceive and remember: FBI CIA USA, or FBICIAU.

**Time.** The capacity limits of working memory are closely related to the second limitation of working memory, the limit of how long information may remain. The strength of information in working memory decays over time unless it is periodically reactivated, or “pulsed” (Cowan, 2001), a process called maintenance rehearsal (Craik & Lockhart, 1972). Maintenance rehearsal for acoustic items in verbal working memory is essentially a serial process of subvocally articulating each item. Thus, for a string of items like a phone number or a personal identity number (PIN), the interval for reactivating any particular item depends on the length of time to proceed through the whole string. For a seven-digit phone number, we can serially reactivate all items in a relatively short time, short enough to keep all items active (i.e., so that the first digit in the phone number will still be active by the time we have cycled through the last item). The more chunks contained in working memory (like a seven-digit phone number plus a three-digit area code), the longer it will take to cycle through the items in main-
tenance rehearsal, and the more likely it will be that items have decayed beyond the point where they can be reactivated.

Two specific features should be noted in the proceeding example, relevant to both time and capacity. First, seven digits is right about at the working memory limit, but 10 digits clearly exceeds it. Hence, requiring area codes to be retained in working memory, particularly unfamiliar ones, is a bad human factors design (and a costly one when wrong numbers are dialed in long-distance calls). Second, familiar area codes create one chunk, not three, and a familiar prefix also reduces three chunks to one. Thus, a familiar combination, such as one’s own phone number, will occupy six, not 10, slots of working memory capacity.

To help predict working memory decay for differing numbers of chunks, Card, Moran, and Newell (1986) combined data from several studies to determine the “half-life” of items in working memory (the delay after which recall is reduced by half). The half-life was estimated to be approximately 7 seconds for a memory store of three chunks and 70 seconds for one chunk.

Confusability and Similarity. Just as perceptual confusability was seen as a source of error in Chapter 4, so also in working memory high confusability—similarity—between the features of different items means that as their representation decays before reactivation, it is more likely that the discriminating details will be gone. For example, the ordered list of letters E G B D V C is less likely to be correctly retrieved from working memory than is the list E N W R U J because of the greater confusability of the acoustic features of the first list. (This fact, by the way, demonstrates the dominant auditory aspect of the phonetic loop, since such a difference in working memory confusion is observed no matter whether the lists are heard or seen). Thus, decay and time are more disruptive on material that is more similar, particularly when such material needs to be recalled in a particular order (Cowan, 2001). The repetition of items also leads to confusability. A particularly lethal source of errors concerns the confusability of which items are repeated. For example, as Laura discovered in the driving example, the digit string 8553 is particularly likely to be erroneously recalled as 8533. Smith (1981) provides good data on the most likely sources of confusion in digit and letter sequences.

Attention and Similarity. Working memory, whether verbal or spatial, is resource-limited. In the context of Figure 6.1 working memory depends very much upon the limited supply of attentional resources. If such resources are fully diverted to a concurrent task, rehearsal will stop, and decay will be more rapid. In addition, if the activity toward which resources are diverted uses similar material, like diverting attention to listening to basketball scores while trying to retain a phone number, the added confusion may be particularly lethal to the contents of working memory. The diversion of attention need not be conscious and intentional in order to disrupt working memory. For example, Banbury and colleagues (2001) describe the particular way that sounds nearly automatically intrude on the working memory for serial order. We return to this issue of auditory disruption at the end of the chapter, just as we highlighted its attention-capturing properties in our discussion of selective attention. In terms of
Baddeley’s model of working memory, the visual spatial scratchpad is more disrupted by other spatial tasks, like pointing or tracking, and the phonetic loop is more disrupted by other verbal or language-based tasks, like listening or speaking (Wickens et al., 1983; Wickens, 2002).

**Human Factors Implications of Working Memory Limits**

1. **Minimize working memory load.** An overall rule of thumb is that both the time and the number of alphanumeric items that human operators have to retain in working memory during task performance should be kept to a minimum (Loftus et al., 1979). In general, designers should try to avoid human use of long codes of arbitrary digit or numerical strings (Peacock & Peacock-Goebel, 2002). Hence, any technique that can offload more information in working memory sooner is of value. Windows in computer systems allow comparisons between side-by-side information sources without requiring the larger demands on working memory imposed by sequencing between screens. Electronic “notepads” can accomplish the same general purpose (Wright et al., 2000).

2. **Provide visual echoes.** Wherever synthetic voice is used to convey verbal messages, these messages can, and ideally should, be coupled with a redundant visual (print) readout of the information so that the human’s use of the material is not vulnerable to working memory failures. For example, since automated telephone assistance can now “speak” phone numbers with a synthetic voice, a small visual panel attached to the phone could display the same number in the form of a “visual echo.” The visual material can be easily rescanned. In contrast, auditory material whose memory may be uncertain cannot be reviewed without an explicit request to “repeat.”

3. **Provide placeholders for sequential tasks.** Tasks that require multiple steps, whose actions may be similar in appearance or feedback, benefit from some visual reminder of what steps have been completed, so that the momentarily distracted operator will not return to the task, forgetting what was done, and needing to start from scratch (Gray, 2000).

4. **Exploit chunking.** We have seen how chunking can increase the amount of material held in working memory and increase its transfer to long-term memory. Thus, any way in which we can take advantage of chunking is beneficial. There are several ways in which this can be done:

   - **Physical chunk size.** For presenting arbitrary strings of letters, numbers, or both, the optimal chunk size is three to four numbers or letters per chunk (Bailey, 1989; Peacock & Peacock-Goebel, 2002; Wickelgren, 1964).

   - **Meaningful sequences.** The best procedure for creating cognitive chunks out of random strings is to find or create meaningful sequences within the total string of characters. A meaningful sequence should already have an integral representation in long-term memory. This means that the sequence is retained as a single item rather than a set of the individual characters. Meaningful sequences include things such as 555, 4321, or a friend’s initials.


- **Superiority of letters over numbers.** In general, letters induce better chunking than numbers because of their greater potential for meaningfulness. Advertisers have capitalized on this principle by moving from numbers such as 1-800-663-5900, which has eight chunks, to letter-based chunking such as 1-800-GET HELP, which has three chunks (“1-800” is a sufficiently familiar string that it is just one chunk). Grouping letters into one word, and thus one chunk, can greatly increase working memory capabilities.

- **Keeping numbers separate from letters.** If displays must contain a mixture of numbers and letters, it is better to keep them separated (Preczewski & Fisher, 1990). For example, a license plate containing one numeric and one alphabetic chunk, such as 458 GST, will be more easily kept in working memory than a combination such as 4G58ST or 4G58 ST.

5. **Minimize confusability.** Confusability in working memory can be reduced by building physical distinctions into material to be retained. We have already noted that making words and letters sound more different reduces the likelihood that they will be confused during rehearsal. This can sometimes be accommodated by deleting common elements between items that might otherwise be confused. For example, confusion between 3 and 2 is less likely than between A5433 and A5423. Spatial separation also reduces confusability (Hess, Detweiler, and Ellis, 1999). A display that has four different windows for each of four different quantities will be easier to keep track of than a single window display in which the four quantities are cycled. Spatial location represents a salient, discriminating cue to reduce item confusability in such cases.

6. **Avoid unnecessary zeros in codes to be remembered.** The zeros in codes like 002385, which may be created because of an anticipated hundredfold increase in code number, will occupy excessive slots of working memory.

7. **Consider working memory limits in instructions.** The sentences presented in instructions must be accurately comprehended. There may be no tolerance for error in such instructions when they are designed to support emergency procedures. To understand how we comprehend sentences, it is useful to assume that most words in a sentence may need to be retained in working memory until the sentence meaning is interpreted (Wickens & Carswell, 1997; Kintsch & Van Dijk, 1978; Carlson et al., 1989). Thus, long sentences obviously create vulnerabilities. So too do those with unfamiliar words or codes. Particularly vulnerable are those instructions in which information presented early must be retained (rather than "dumped") until the meaning of the whole string is understood. Such an example might be procedural instructions that read:

**Before doing X and Y, do A.**

**Here, X and Y must be remembered until A is encountered. Better would be the order**

**Do A. Then do X and Y.**
Wickens and Hollands (2000) refer to this improved design as one that maintains congruence between the order of text and the order of action. Congruence is a good design principle that reduces working memory load.

Finally, reiterating a point made in the context of perception, designers of comprehension material should remember that negation imposes an added chunk in working memory. Even if the negation may be perceived in reading or hearing an instruction, it may be forgotten from working memory as that instruction is retained before being carried out. In such circumstances, the default memory of the positive is likely to be retained, and the user may do the opposite of what was instructed. This is another reason to advocate using positive assertions in instructions where possible (Wickens & Hollands, 2000). More details on text and instructional design are given in Chapter 18.

**LONG-TERM MEMORY**

We constantly maintain information in working memory for its immediate use, but we also need a mechanism for storing information and retrieving it at later times. This mechanism is termed long-term memory. Learning is the processing of storing information in long-term memory, and when specific procedures are designed to facilitate learning, we refer to this as instruction or training, an issue treated in depth in Chapter 18. Our emphasis in the current chapter is on retrieval and forgetting and the factors that influence them.

Long-term memory can be distinguished by whether it involves memory for general knowledge, called semantic memory (memory for facts or procedures), or memory for specific events, called event memory.

The ability to retrieve key information from long-term memory is important for many tasks in daily life. We saw at the beginning of this chapter that Laura’s failure to recall instructions was a major source of her subsequent problems. In many jobs, forgetting to perform even one part of a job sequence can have catastrophic consequences. In this section, we review the basic mechanisms that underlie storage and retrieval of information from long-term memory and how to design around the limitations of the long-term memory system.

**Basic Mechanisms**

Material in long-term memory has two important features that determine the ease of later retrieval: its strength and its associations.

**Strength.** The strength of an item in long-term memory is determined by the frequency and recency of its use. Regarding frequency, if a password is used every day (i.e., frequently) to log onto a computer, it will probably be well represented in long-term memory and rarely forgotten. Regarding recency, if a pilot spends a day practicing a particular emergency procedure, that procedure will be better recalled (and executed) if the emergency is encountered in flight the very next day than if it is encountered a month later. In this regard, the fact that emergency procedures are generally not used frequently in everyday practice suggests that their use should be supported by external visual checklists rather than reliance upon memory.
Associations. Each item retrieved in long-term memory may be linked or associated with other items. For example, the sound of a foreign word is associated with its meaning or with its sound in the native language of the speaker. As a different example, a particular symptom observed in an abnormal system failure will, in the mind of the skilled troubleshooter, be associated with other symptoms caused by the same failure as well as with memory of the appropriate procedures to follow given the failure. Associations between items have a strength of their own, just as individual items do. As time passes, if associations are not repeated, they become weaker. For example, at some later point a worker might recognize a piece of equipment but be unable to remember its name.

Working Memory and Long-term Memory. Information in long-term memory becomes more available as a function of the richness or number of associations that can be made with other items. Like strings tied to an underwater object, the more strings there are, the greater likelihood that any one (or several) can be found and pulled to retrieve the object. Thus, thinking about the material you learn in class in many different contexts, with different illustrative examples, improves your ability to later remember that material. Doing the mental work to form meaningful associations between items describes the active role of working memory in learning (Carlson et al., 1989; see Chapter 18). As we noted in the discussion of working memory, storing such relations in long-term memory results in the formation of chunks, which are valuable in reducing the load on working memory. Sometimes, however, when rehearsing items through simple repetition (i.e., the pure phonetic loop) rather than actively seeking meaning through associations, our memories may be based solely on frequency and recency, which is essentially rote memory. Rote memory is more rapidly forgotten. This is a second reason that advertisers have moved from solely digit-based phone numbers to items such as 1-800-GET-RICH. Such phone numbers have both fewer items (chunks) and more associative meaning.

Forgetting. The decay of item strength and association strength occurs in the form of an exponential curve, where people experience a very rapid decline in memory within the first few days. This is why evaluating the effects of training immediately after an instructional unit is finished does not accurately indicate the degree of one’s eventual memory. Even when material is rehearsed to avoid forgetting, if there are many associations that must be acquired within a short period of time, they can interfere with each other or become confused, particularly if the associations pertain to similar material. New trainees may well recall the equipment they have seen and the names they have learned, but they confuse which piece of equipment is called which name as the newer associations interfere with the older ones.

Thus, memory retrieval often fails because of (1) weak strength due to low frequency or recency, (2) weak or few associations with other information, and (3) interfering associations. To increase the likelihood that information will be remembered at a later time, it should be processed in working memory frequently and in conjunction with other information in a meaningful way.
Different forms of long-term memory retrieval degrade at different rates. In particular, recall, in which one must retrieve the required item (fact, name, or appropriate action), is lost faster than recognition, in which a perceptual cue is provided in the environment, which triggers an association with the required item to be retrieved. For example, a multiple-choice test visually presents the correct item, which must be recognized and discriminated from a set of "foils." In contrast, short-answer questions require recall. In human–computer interaction, discussed in Chapter 15, command languages require recall of the appropriate commands to make something happen. In contrast, menus allow visual recognition of the appropriate command to be clicked.

Organization of Information in Long-Term Memory

It is apparent from the description of working memory that we do not put isolated pieces of information in long-term memory the way we would put papers in a filing cabinet. Instead, we store items in connection with related information. The information in long-term memory is stored in associative networks where each piece of information (or image or sound) is associated with other related information. Much of our knowledge that we use for daily activities is semantic knowledge, that is, the basic meaning of things. Cognitive psychologists have performed research showing that our knowledge seems to be organized into semantic networks where sections of the network contain related pieces of information. Thus, you probably have a section of your semantic network that relates all of your knowledge about college professors, both general information and specific instances, based on previous experience. These semantic networks are then linked to other associated information, such as images, sounds, and so on.

A semantic network has many features in common with the network structure that may underlie a database or file structure, such as that used in an index, maintenance manual, or computer menu structure. It is important that the designer create the structure of the database to be compatible or congruent with the organization of the user's semantic network (Roske-Hofstrand & Paap, 1986; Seidler & Wickens, 1992). In this way, items that are close together, sharing the same node in the semantic network, will be close together in the database representation of the information. For example, if the user of a human factors database represents perception and displays as closely associated, the database should also contain links between these two concepts. We see in Chapter 8 how this process can be aided by good displays.

In addition to networks, there are three other ways that psychologists have described the organization of information: schemas, mental models, and cognitive maps.

Schemas and Scripts. The information we have in long-term memory is sometimes organized around central concepts or topics. The entire knowledge structure about a particular topic is often termed a schema. People have schemas about all aspects of their world, including equipment and systems that they use. Examples of common schemas are semantic networks associated with college courses,
cups, or vacations. Schemas that describe a **typical sequence of activities**, like getting online in a computer system, shutting down a piece of industrial equipment, or dealing with a crisis at work, are called **scripts** (Schank & Abelson, 1977).

**Mental Models.** People also have schemas about equipment or systems. The schemas of dynamic systems are often called **mental models** (Gentner & Stevens, 1983; Norman, 1988; Rouse & Morris, 1986; Wilson & Rutherford, 1989). Mental models typically include our understanding of system components, how the system works, and how to use it. In particular, mental models generate a set of **expectancies** about how the equipment or system will behave. Mental models may vary on their degree of **completeness** and **correctness**. For example, a correct mental model of aerodynamics posits that an aircraft stays aloft because of the vacuum created over the wings. An incorrect model assumes that it stays aloft because of the speed through which it travels through airspace. Mental models may also differ in terms of whether they are personal (possessed by a single individual) or are similar across large groups of people. In the latter case the mental model defines a **population stereotype** (Smith, 1981). Designs that are consistent with the population stereotype are said to be **compatible** with the stereotype (such as turning a knob clockwise should move a radio dial to the right). Later chapters on displays (Chapter 8), controls (Chapter 9), and computer design (Chapter 15) illustrate the importance of knowing the user’s mental model.

**Cognitive Maps.** Mental representations of spatial information, like the layout of a city, a room, or a workplace, are referred to as cognitive maps. They represent the long-term memory analogy to the visual-spatial scratchpad in working memory. Such maps may not necessarily be accurate renderings of the space they represent (Wickens & Hollands, 2000). For example, cognitive maps of a geographical area often simplify by “mentally straightening” corners that are not at right angles (Chase & Chi, 1979). People also have a preferred or “canonical” orientation by which they typically represent an environment (Sholl, 1987). This may often represent the direction in which you most frequently view the environment. For example, your cognitive map of a classroom may have the orientation of the direction you face when you sit in it. Reorienting one’s perspective of a cognitive map through “mental rotation” requires mental effort (Tversky & Franklin, 1981). As we discuss in our treatment of map displays in Chapter 8, this has some implications for how maps are configured.

**Long-Term Memory Implications for Design**

Designers frequently fail to realize or predict the difficulty people will experience in using their system. One reason is that they are extremely familiar with the system and have a very detailed and complete mental model (Norman, 1988). They **know how the system works**, when it will do various things, and how to control the system to do what the user wishes. They **fail to realize that the average user does not have this mental model** and may never interact with the system **enough to develop one**. When people have to do even simple tasks on an infrequent basis, they forget things. Manufacturers write owners’ manuals as if
they will be read thoroughly and all of the information will be remembered for
the life of the equipment. Neither is necessarily the case. Even if we have very
clear and explicit instructions for operating our programmable VCR (which is
unlikely), what average owner wants to get the instructions out every time he or
she must perform a task?

The following are some ways that we can design the environment and sys-
tems within it so that people do not have problems, errors, accidents, and incon-
veniences due to poor retrieval from long-term memory.

1. **Encourage regular use of information to increase frequency and recency.**

2. **Encourage active verbalization or reproduction of information that is to be recalled.** For example, taking notes in class or requiring active recitation or read-
back of heard instructions increases the likelihood that the information will be
remembered.

3. **Standardize.** One way that we can decrease the load on long-term
memory is to standardize environments and equipment, including controls,
displays, symbols, and operating procedures. An example from the automotive
industry where a control is being standardized is the shift pattern, and where a
control has still not been standardized is the location and operation of elec-
tronic windows and lighting. **Standardization results in development of strong yet simple schemas and mental models that are applicable to a wide variety of
circumstances.** Of course, the conflict between standardizing across industries
and still preserving uniqueness of product style remains a difficult design chal-
lenge.

4. **Use memory aids.** When a task will be performed infrequently or when
correct task performance is critical, designers should provide computer-based
or hardcopy memory aids or job aids as discussed in Chapter 18. These consist
of information critical for task performance and can be as simple as a list of
procedures.

Norman (1988) characterizes memory aids as putting “knowledge in the
world” (i.e., perception) so that the operator does not have to rely on “knowl-
edge in the head” (i.e., long-term memory). In the context of command lan-
guages and menus, such aids often replace recall requirements with recognition
opportunities. This important human factors topic is reconsidered in Chapters
15 and 18.

5. **Carefully design information to be remembered.** Information that must be
remembered and later retrieved unaided should have characteristics such as the
following:

- Meaningful to the individual and semantically associated with other
  information.
- Concrete rather than abstract words when possible.
- Distinctive concepts and information (to reduce interference).
- Well-organized sets of information (grouped or otherwise associ-
at ed).
- Able to be guessed based on other information (top-down process-
ing).
- Little technical jargon.
6. Design to support development of correct mental models. One way to develop correct mental models is to apply the concept of visibility, as suggested by Norman (1988). This guideline suggests that a device has visibility if the user can immediately and easily determine the state of the device and the alternatives for action. For example, switches that have different positions when activated have visibility, whereas push/toggle switches do not. The concept of visibility also relates to the ability of a system to show variables intervening between an operator’s action and the ultimate system response. An example is an oven display showing that an input has been read, the heat system is warming up, and the temperature has not reached the target temperature. Mental model development can also be encouraged by the appropriate wording of instructional manuals that describe why a particular action is required as well as what the action is.

Episodic Memory for Events

In contrast to both procedural and declarative knowledge, which is often embodied in schemas, scripts, and skills and acquired from multiple experiences, the personal knowledge or memory of a specific event or episode is, almost by definition, acquired from a single experience. This may be the first encounter with an employer or coworker, a particular incident or accident at home or the workplace, or the eyewitness view of a crime or accident. Such memories are very much based on visual imagery, but the memories themselves are not always faithful “video replays” of the events, having a number of biases.

Episodic memory is of tremendous importance to the psychology of eyewitness testimony. While this is certainly of great importance to legal criminal proceedings (Wright & Davies, 1999; Devenport et al., 1999), it also has considerable relevance to the field of accident investigation, which we discuss in Chapter 14. That is, what does the witness to an accident recall about its circumstances when later interviewed by the investigator?

Through a simple cognitive task analysis, we can represent the processes involved in the formation, storage, and retrieval of episodic memories as shown in Figure 6.2. Here an “event” occurs, which defines some ground truth of what actually happened. The witness observes the event and encodes information about it, which reflects the allocation of selective attention and may reflect some of the top-down biases of expectancy on perception that we described earlier. As time passes, the memory of the episode is maintained in long-term memory, where it will show some degradation (forgetting), and the memory may be distorted by influences related to both schema memory and specific intervening events (Bartlett, 1932). Finally, the memory may be retrieved in a variety of circumstances: For example, a witness picks out a suspect from a police lineup, the witness is interviewed by police as the prosecution develops its case, or the witness responds to queries during actual courtroom testimony. These retrieval tests may have characteristics of both recall and recognition.

Extensive research on eyewitness testimony has revealed that the episodic memory process is far from perfect (e.g., Wright & Davies, 1999; Schacter, 2001; Wells & Seelau, 1995). In one study of police lineup recognition, for example, Wright and McDaid (1996) estimated that an innocent person was chosen (as a guilty perpetrator) approximately 20 percent of the time. The sources of such
Episodic memory. The processes involved in episodic memory characteristics and influences on these processes are shown in the box at the bottom. It will be noted that retrieval in a courtroom (testimony) often starts another memory cycle: that of the jury who encodes the witness testimony for later retrieval during the jury deliberations and judgment.

Biases can occur at all three stages. For example, at encoding, a well-established bias is the strong focus of witness attention on a weapon when one is used at the scene of the crime. In light of what we know about the limits of attention, it should come as no surprise that this focus degrades the encoding of other information in the scene, particularly the physical appearance of the suspect's face relative to crimes where no weapon is employed (Loftus et al., 1987). In a different application of attention research, Lassiter and his colleagues (e.g., Lassiter, 2002) show how the focus of a video camera during interrogation, on the suspect alone rather than on the suspect and the interviewer, can bias the judgment of a jury who views such a video. Focus on the suspect alone leads jurors to substantially increase their later judgment that the suspect is guilty, independent of the contents of the interrogation.

Episodic memory also has an auditory component, but this too may be flawed. John Dean, the former council to President Richard M. Nixon, had a reputation for having a particularly precise memory. Some even called him a "human tape recorder" (Neisser, 1982). His confident recall of dozens of conversations helped bring down the Nixon administration. After his testimony, tapes of the actual conversations were released and Neisser compared the recorded and recalled conversations. He found Dean's memory to be seriously flawed regarding the details of specific conversations. Dean was not a "human tape recorder"; however, he was quite accurate in capturing the general theme or gist of the conversations. Instead of a verbatim recording of conversations, memory
relies on extracting the gist and reconstructing the details. The reconstruction of the details may be distorted by the cultural background and self-interests of the individual (Bartlett, 1932).

As Figure 6.2 suggests, two qualitatively different forms of bias may influence the memory during storage (Wright & Davies, 1999). First, a degraded visual recollection may be partially replaced by a long-term memory schema of what the crime might "typically" look like. For example, it may be replaced by the witness’s memory of the appearance of the "typical criminal" or by the assumption that the typical automobile accident will occur at a high rate of speed, thereby leading to an overestimation of vehicle speed in a crash. Second, certain events during the storage interview can also bias memory. For example, a chance encounter with a suspect in handcuffs in the hallway prior to a police lineup might increase the likelihood that the suspect will be selected in the lineup. Sometimes, the way questions are phrased in a witness interview can also "suggest" that a particular suspect is guilty or that events occurred in a different way than they actually did, and as a consequence, distort the accuracy of episodic recall, which may be used in trial.

Finally, biases at retrieval can sometimes be represented in the form of a signal detection task discussed in Chapter 4 when recognition tests are used, as they are in a police lineup (Wells, 1993). As shown in the lower right corner of Figure 6.2, a "signal" can be represented as the witness’s accurate episodic memory of the suspect’s appearance. The witness’s response is represented as either selecting the suspect from the lineup (yes) or failing to do so (no). This defines the four classes of events in which the "hit" is the most important for accurately developing the police’s case. In contrast, a false alarm, in which an innocent person is positively identified, is clearly an undesirable event and one that has dangerous implications for society.

Within this context, it is in the interest of all parties to maximize the sensitivity (keeping misses and false alarms to a minimum). However, it is also important to avoid a "guilty bias" where witnesses are likely to see the suspect as being guilty. Even one who had not encoded the crime at all would still have a 20 percent chance of picking the suspect from a police lineup if the witness felt certain that the actual perpetrator was in the lineup. Wells and Seelau (1995) describe ways of conducting eyewitness line-up procedures to avoid the unfortunate consequences of a guilty bias (see also Wright & Davies, 1999). For example, they suggest that witnesses be clearly informed that the perpetrator might not be in the lineup; furthermore, witnesses can initially be shown a blank lineup in which the suspect is not included. Witnesses who "recognize" the suspect in such a lineup can be assumed to have a strong guilty bias and their testimony can therefore be discounted. If they do not respond yes to the blank lineup, then they can be shown the actual lineup with the suspect included.

Since those who judge a witness’s testimony often have no independent means of assessing the accuracy of that testimony (e.g., they do not know the "sensitivity" of a recognition memory test), we might think that asking witnesses to express the confidence in their memory should provide a means of assessing this accuracy. Unfortunately, however, extensive research has shown that the self-rated confidence of a witness’s judgment is only weakly correlated with the
accuracy of that judgment (Wells & Seelau, 1995; Wright & Davies, 1999; Wells et al., 1979). People aren’t very well calibrated in estimating the strength of their own episodic memory.

In one important application of memory research to episodic retrieval, Fisher and Geiselman (1992; Fisher, 1999) developed what is called the cognitive interview (CI) technique for assisting police in interviewing witnesses in order to maximize the retrieval of information. Their approach is to avoid recognition tasks because, they argue persuasively, classic recognition tests, approximated by asking witnesses a series of yes-or-no questions (“Did the suspect have red hair?”) can be quite biasing and leave vast quantities of encoded information untapped. Instead, they apply a series of principles from cognitive psychology to develop effective recall procedures. For example, the CI technique

- Encourages the witness to reinstate the context of the original episode, thereby possibly exploiting a rich network of associations that might be connected with the episodic memory.
- Avoids time-sharing requirements where the witness must divide cognitive resources between searching episodic memory for details of the crime and listening to the interrogator ask additional questions. We learn about the consequences of such time-sharing later in the chapter.
- Avoids time stress, allowing the witness plenty of time to retrieve information about the crime and ideally allowing the witness multiple opportunities to recall. These multiple opportunities will take advantage of the rich network of associations.

The CI technique has been shown to allow witnesses to generate between 35 and 100 percent more information than standard police interview procedures and to do so without any substantial loss of accuracy; it has been adopted by a number of police forces (Fisher, 1999; Wright & Davies, 1999).

A final important issue regarding cognitive psychological principles in legal proceedings pertains to the admissibility of testimony from expert psychologists regarding the sorts of eyewitness biases described above (Devenport et al., 1999; Levett & Kovera, 2002). Judges may disagree on whether such scientific recommendations are themselves “true” and hence admissible evidence. Research shows that to the extent that such expert testimony from psychologists is admissible, it has two effects on jury belief. First it leads to some general down-weighting of the impact of that testimony, as jurors themselves become more skeptical of the “video tape” analogy to episodic memory. Second, it allows the jurors to become more sensitive in discriminating accurate from inaccurate testimony (Devenport et al., 1999).

In conclusion, it is evident that the cognitive psychology of memory and attention has tremendous importance for the quality of criminal and other legal proceedings. We will also see the relevance of the study of decision making in the next chapter. One final implication for every reader is that when you witness a serious episode about which you might be later queried, it is good advice to write down everything about it as soon as the episode has occurred and at that time think clearly about and indicate your degree of certainty or uncertainty.
about the events within the incident. Your written record will now be "knowledge in the world," not susceptible to forgetting.

**Prospective Memory for Future Events**

Whereas failures of episodic memory are inaccurate recollection of things that happened in the past, failures of prospective memory are forgetting to do something in the future (Harris & Wilkins, 1982). Laura, in the story at the beginning of the chapter, forgot to activate the voice mode while the traffic was still light. In 1991, an air traffic controller positioned a commuter aircraft at the end of a runway and later forgot to move the aircraft to a different location. The unfortunate aircraft was still positioned there as a large transport aircraft was cleared to land on the same runway. Several lives were lost in the resulting collision (NTSB, 1992).

Failures of prospective memory are sometimes called absentmindedness. Several system and task design procedures are incorporated in systems to support prospective memory. Strategies can be adopted to implement reminders (Herrmann et al., 1999). These may be things like tying a string around your finger, setting a clock or programming a personal data assistant (PDA) to sound an alarm at a future time, taping a note to the steering wheel of your car, or putting a package you need to mail in front of the door so that you will be sure to notice it (if not trip on it!) on your way out. In systems with multiple operators, sharing the knowledge of what one or the other is to do decreases the likelihood that both will forget that it is to be done. Also, loss of prospective memory is reduced by verbally stating or physically taking some action (e.g., writing down or typing in) regarding the required future activity the moment it is scheduled. Checklists are aids for prospective memory (Degani & Wiener, 1991). Herrmann and colleagues (1999) describe characteristics of ideal reminding devices like the PDA.

**SITUATION AWARENESS**

In the dynamic sequence of events leading up to Laura’s near accident, she was unaware of the converging vehicle until her attention was captured by its horn. Designers, researchers, and users of complex dynamic systems often employ the cognitive concept of situation awareness, or SA, to characterize users' awareness of the meaning of dynamic changes in their environment (Durso & Gronlund, 1999; Adams et al., 1995). A pilot loses SA whenever he or she suffers a catastrophic controlled-flight into terrain (Strauch, 1997; Wiener, 1977), and as we shall see in Chapter 16, control room operators at the Three Mile Island nuclear power plant lost SA when they believed the water level in the plant to be too high rather than too low, a misdiagnosis that led to a catastrophic release of radioactive material into the atmosphere (Rubinstein & Mason, 1979).

Endsley (1995) defines SA as “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” (p 36). These three stages, perception (and selective attention), understanding, and prediction, must be applied to a specific situation. Thus, a user cannot be simply said to have SA without specifying what that awareness is (or should be) about. A vehicle driver may
have good awareness of time and navigational information (where I am and how much time it will take me to drive to where I need to be), but little awareness of the local traffic tailgating behind.

Many elements that are necessary to support SA have been covered elsewhere in this chapter. Selective attention is necessary for the first stage, while the second stage of understanding depends very much upon both working memory and long-term memory. The third stage, projection and prediction, however, is an important construct in cognitive psychology that we have not yet discussed but consider in more detail later when we discuss planning and scheduling.

It is important to note that SA is distinct from performance. SA can be maintained even when there is no performance to be observed. For example, a passerenger in a vehicle may have very good awareness of the traffic and the navigational situation, even as he or she carries out no actions (other than visual scanning). Great differences in the ability of pilots to deal with unexpected occurrences within their automation system are observed as a function of how well they are aware of changes in an automated state during periods when the pilots are totally passive observers (Sarter & Woods, 2000; Sarter et al., 1997). We can also identify instances in which very good performance is observed with low SA, as when you are so absorbed in doing a task that you lose awareness of the time. A key issue here is that the importance of SA is not so much for understanding and describing the quality of routine performance (e.g., the accuracy in staying in a lane or maintaining speed while driving) as it is for understanding the appropriate and timely response to unexpected events (Wickens, 2000).

### Measuring SA

The importance of SA can often be realized after an accident by inferring that the loss of SA was partially responsible. In controlled-flight-into-terrain accidents it is almost always assumed that the pilot lost awareness of the aircraft’s altitude over or trajectory toward the terrain (Strauch, 1997). However, “measuring” SA after the fact by assuming its absence (SA = 0) is not the same as measuring how well a particular system or operator preserves SA in the absence of an unexpected event (Endsley & Garland, 2000). A popular technique for SA measurement is the SA global assessment technique (SAGAT; Endsley, 1995) in which the operator is briefly interrupted in the performance of a dynamic task and asked questions about it; for example, identify the location of other road traffic (Gugerty, 1997) or identify the direction of the nearest hazardous terrain (Wickens & Prevett, 1995). While SA can sometimes be measured by a subjective evaluation (“rate your SA on a scale of 1 to 10; Selcon et al., 1991), a concern about the validity of such self-rating techniques is that people are not always aware of what they are not aware. This issue of metacognition is addressed later in this chapter.

### Importance of SA to Human Factors

Probably the first and most direct application of the SA concept to human factors is its implications for designing easy-to-interpret displays of dynamic systems that can help people notice what is going on (stage 1), interpret and
understand the meaning—a challenge when there are several coupled display elements in a complex system (stage 2), and predict the future implications—a challenge when the system is slow or lagged, like a supertanker, industrial oven, or air traffic system. Human factors practitioners have noted how easy it is to lose SA when automation carries out much of the processing for complex systems and hence how critical it is to have SA-supporting displays in the unexpected event that automation does not perform as intended (Parasuraman et al., 2000; Sarter et al., 1997; Sarter & Woods, 2000; see Chapter 16). In this regard, the design of products to maximize the effectiveness of routine performance may not be the same as the design of those to support SA (Wickens, 2000). The support for SA typically imposes the need for added information display to support appropriate behavior when unexpected things go wrong. This information must be carefully integrated in order to avoid issues of information overload.

Second, SA can be an important tool for accident analysis, understanding when its loss was a contributing factor (Strauch, 1997). To the extent that accidents may be caused by SA loss, an added implication is that systems should be designed and, when appropriate, certified to support SA (Wickens, 2000). This becomes important when federal regulators are responsible for certification, such as the case with new aircraft or nuclear power plants.

Third, the SA concept has important implications for training. Training for routine performance may conflict with training to maintain SA. One particularly relevant aspect concerns the training of attentional skills (stage 1 SA) to scan the environment with enough breadth to assure that important and relevant dynamic events are noticed when they occur (Gopher, 1993).

**PROBLEM SOLVING AND TROUBLESHOOTING**

The cognitive phenomena of problem solving and troubleshooting are often closely linked because they have so many overlapping elements. Both start with a difference between an initial "state" and a final "goal state" and typically require a number of cognitive operations to reach the latter. The identity of those operations is often not immediately apparent to the human engaged in problem-solving behavior. Troubleshooting is often embedded within problem solving in that it is sometimes necessary to understand the identity of a problem before solving it. Thus, we may need to understand why our car engine does not start (troubleshoot) before trying to implement a solution (problem solving). Although troubleshooting may often be a step within a problem-solving sequence, problem solving may occur without troubleshooting if the problem is solved through "trial and error" or if a solution is accidentally encountered through serendipity.

While both problem solving and troubleshooting involve attaining a state of knowledge, both also typically involve performance of specific actions. Thus, troubleshooting usually requires a series of tests whose outcomes are used to diagnose the problem, whereas problem solving usually involves actions to implement the solution. Both are considered to be iterative processes of perceptual, cognitive, and response-related activities involving the full cycle of processing shown in Figure 6.1.
Challenges

Both problem solving and troubleshooting impose heavy cognitive activity, and human performance is therefore often limited (Wickens & Hollands, 2000; Casner, 1994; Teague & Allen, 1997). In troubleshooting, for example, people usually maintain no more than two or three active hypotheses in working memory as to the possible source of a problem (Rasmussen, 1981; Wickens, 1992). More than this number overloads the limited capacity of working memory, since each hypothesis is complex enough to form more than a single chunk. Furthermore, when testing hypotheses, there is a tendency to focus on only one hypothesis at a time in order to confirm it or reject it. Thus, the engine troubleshooter will probably assume one form of the problem and perform tests specifically defined to confirm that it is the problem.

Naturally, troubleshooting success depends closely upon attending to the appropriate cues and test outcomes. This dependency makes troubleshooting susceptible to attention and perceptual biases. The operator may attend selectively to very salient outcomes (bottom-up processing) or to outcomes that are anticipated (top-down processing). As we consider the first of these potential biases, it is important to realize that the least salient stimulus or event is the “nonevent.” People do not easily notice the absence of something (Wickens & Hollands, 2000; Hunt & Rouse, 1981). Yet the absence of a symptom can often be a very valuable and diagnostic tool in troubleshooting to eliminate faulty hypotheses of what might be wrong. For example, the fact that a particular warning light might not be on could eliminate from consideration a number of competing hypotheses.

An important bias in troubleshooting, resulting from top-down or expectancy-driven processing, is often referred to as cognitive tunneling, or confirmation bias (Woods & Cook, 1999; Woods et al., 1994; see Chapter 7). In troubleshooting, this is the tendency to stay fixated on a particular hypothesis (that chosen for testing), look for cues to confirm it (top-down expectancy guiding attention allocation), and interpret ambiguous evidence as supportive (top-down expectancy guiding perception). In problem solving, the corresponding phenomenon is to become fixated on a particular solution and stay with it even when it appears not to be working.

These cognitive biases are more likely to manifest when two features characterize the system under investigation. First, high system complexity (the number of system components and their degree of coupling or links) makes troubleshooting more difficult (Meister, 2002; Wohl, 1983). Complex systems are more likely to produce incorrect or “buggy” mental models (Sanderson & Murtaugh, 1990), which can hinder the selection of appropriate tests or correct interpretation of test outcomes. Second, intermittent failures of a given system component turn out to be particularly difficult to troubleshoot (Teague & Allen, 1997).

PLANNING AND SCHEDULING

The cognitive processes of planning and scheduling are closely related to those discussed in the previous section, because informed problem solving and troubleshooting often involve careful planning of future tests and activities. How-
ever, troubleshooting and diagnosis generally suggest that something is “wrong” and needs to be fixed. Planning and scheduling do not have this implication. That is, planning may be invoked in the absence of problem solving, as when a routine schedule of activities is generated. We saw earlier in the chapter that prospective memory could be considered a form of planning.

In many dynamic systems, the future may be broken down into two separate components: the predicted state of the system that is being controlled and the ideal or command state that should be obtained. Thus, a factory manager may have predicted output that can be obtained over the next few hours (given workers and equipment available) and a target output that is requested by external demands (i.e., the factory’s client). When systems cannot change their state or productive output easily, we say they are sluggish, or have “high inertia.” In these circumstances of sluggish systems, longer range planning becomes extremely important to guarantee that future production matches future demands. This is because sudden changes in demand cannot be met by rapid changes in system output. Examples of such sluggish systems—in need of planning—are the factory whose equipment takes time to be brought online, the airspace in which aircraft cannot be instantly moved to new locations, or any physical system with high inertia, like a supertanker or a train.

You will recognize the importance to planning of two cognitive constructs discussed earlier in the chapter. First, stage 3, SA is another way of expressing an accurate estimate of future state and future demands. Second, skilled operators often employ a mental model of the dynamic system to be run through a mental simulation in order to infer the future state from the current state (Klein & Crandall, 1995). The role of mental simulation is discussed in Chapter 7, and the great importance of mental models in controlling complex and sluggish industrial processes is visited in Chapter 16. Here, however, we note the heavy cognitive demands on working memory to run an accurate mental model. Such a task requires a heavy investment of cognitive resources (the “tank” at the top of Figure 6.1). Where these resources are lacking, diverted to other tasks, then prediction and planning may be poor, or not done at all, leaving the operator unprepared for the future.

In general, people tend to avoid complex, optimizing, planning schedules over long time horizons (Tulga & Sheridan, 1980), a decision driven both by a desire to conserve the resources imposed by high working memory load and by the fact that in an uncertain world accurate planning is impossible, and plans may need to be revised or abandoned altogether as the world evolves in a way that is different from what was predicted. Here, unfortunately, people sometimes fail to do so, creating what is known as a plan continuation error (Orasanu et al., 2001; Muthard & Wickens, 2003; Goh & Wiegmann, 2001), a form of behavior that has much in common with cognitive tunneling.

As with problem solving and troubleshooting, a variety of automation tools are proposed to reduce these cognitive demands in planning (Gronland et al., 2002). Most effective are predictive displays that offer visual representations of the likely future, reducing the need for working memory (Wickens et al., 2000). We discuss these in the next chapter. Also potentially useful are computer-based planning aids that can either recommend plans (Layton et al., 1994; Muthard &
Wickens, 2003) or allow fast-time simulation of the consequence of such plans to allow the operator to try them out and choose the successful one (Sheridan, 2002). Air traffic controllers can benefit from such a planning aid known as the User Request Evaluation Tool (URET) to try out different routes to avoid aircraft conflicts (Wickens et al., 1998).

**METACOGNITION AND EFFORT**

Performance of nearly all tasks is supported by some combination of perceptual information and long-term-memory knowledge about the task. Norman (1988) refers to these as “knowledge in the world” and “knowledge in the head” respectively. Psychologists have also identified a qualitatively different source of knowledge that is important in many aspects of performance, metaknowledge or metacognition (Reder, 1996; Bjork, 1999), which refers to people’s knowledge about their own knowledge and abilities. Consider, for example, a troubleshooter who is trying to diagnose and fix an engine problem before restarting. Conditions are such that if the diagnosis is incorrect and a restart is tried (and fails), it could lead to serious damage. She asks herself whether she knows enough about the nature of the problem and the projected effectiveness of her “fix” to be confident that the start will proceed without damage. In short, she assesses her knowledge about her own knowledge. In a corresponding situation, a student may assess whether he knows enough to stop studying for the test and turn to another activity.

Another example of metacognition might be the eyewitness who is about to testify and applies her awareness of the general tendency toward overconfidence in recognition memory in such a way as to consciously “downgrade” her estimates of self-confidence on the witness stand. Thus, metacognition sometimes modulates people’s choices of what they do, assertions of what they know (knowledge in the head), and choices of whether additional information should be sought (knowledge in the world).

Seeking additional information related to selective attention is also related to another construct of metacognition, the anticipated effort required to gain that information (Wright et al., 2000; Fennema & Kleinmuntz, 1995). This construct of anticipated effort is closely linked to the strategies people use with information systems, not just seeking information but also performing a wider range of tasks (Gray, 2000). People often ask themselves, implicitly or explicitly, whether the anticipated effort necessary to access information is worth the potential gains in knowledge from acquiring that information. For example, is it worthwhile traveling across campus to the library to check out a particular book that contains the information I need, or to continue an apparently unproductive search for new information (MacGregor et al., 1987)? In a more general sense, people ask themselves similar tradeoff questions regarding whether the effort required to use a particular system feature balances the gain in productivity from using that feature.

One important metacognitive tradeoff often made is between knowledge in the head and knowledge in the world. Sometimes gaining knowledge in the
world (accessing perceptual information) is more accurate but requires more effort than using knowledge in the head (relying upon potentially faulty memory; Gray & Fu, 2001). It is important, however, for designers to realize that people are often overconfident in the accuracy of their own knowledge (Bjork, 1999), as was the case with Laura’s knowledge of how to activate the voice mode in her vehicle and with the overconfidence of eyewitnesses. Thus, the decision of users to avoid effort-imposing access of perceptual information may not always be a wise one.

Balancing the costs and benefits of attributes like anticipated effort and accuracy is an issue discussed more formally in the context of decision making in Chapter 7. There too, we discuss the manner in which people are effort-conserving in the kinds of decision strategies they use (Bettman et al., 1990), choosing low effort heuristics over high-effort, optimizing decision techniques. With regard to scheduling and planning, people tend to choose simpler schedules rather than complex but optimal ones (Raby & Wickens, 1994). Designers must understand the effort costs generated by potentially powerful features in interfaces. Such costs may be expressed in terms of the cognitive effort required to learn the feature or the mental and physical effort and time cost required to load or program the feature. Many people are disinclined to invest such effort even if the anticipated gains in productivity are high. The feature will go unused as a result. Correspondingly, requiring people to engage in manual activity to retrieve information is more effort-consuming than simply requiring them to scan to a different part of the visual field (Yeh & Wickens, 2001; Gray & Fu, 2001), a characteristic that penalizes the concepts of hidden databases, multilevel menus, and decluttering tools. Solutions to this problem are offered by pop-up messages and other automation features that can infer a user’s information needs and provide them without imposing the effort cost of access (Hammer, 1999).

We now turn to a direct examination of this important concept of effort as it is joined with other information processing features to determine people’s success or failure in carrying out two tasks at the same time: divided attention.

**ATTENTION AND TIME-SHARING**

Earlier in this chapter we spoke of attention as acquiring information about the environment. This was selective attention, a process that sometimes requires effort. In this section we discuss attention as supporting the ability to do two (or more) things at one time—to divide attention between two tasks or mental activities (Wickens, 2002). The two aspects of attention are related, but not identical (Wickens et al., 2003, in press). For example, selecting two sources of information to process—the roadway view and the electronic map for our driver Laura—may or may not allow the successful division of attention between the tasks supported by those sources. In Laura’s case, it did not. Researchers of human time-sharing have identified four major factors that contribute to the success or failure of divided attention (Damos, 1991): resource demand, structure, similarity, and resource allocation or task management.
Mental Effort and Resource Demand

In the prior section, we described the effort required to carry out a task or cognitive activity. People, being effort-conserving, tend to avoid high-effort activities or to do them poorly, such as rehearsing an eight-chunk phone number, engaging in mental rotation, or doing prediction. Furthermore, the high mental effort, difficulty, or resource demand of one activity degrades the ability to carry out a second activity at the same time, as if the resources necessary to support one, shown in the “tank” at the top of Figure 6.1, are limited and are therefore less available to the other. For example, one can converse and drive at the same time if the conversation is simple and the driving task is easy. But when the conversation becomes difficult, perhaps solving a tough problem, resources may be diverted from driving at the cost of safety. Alternatively, if the driving suddenly becomes demanding, conversation may cease. This relationship between single-task difficulty and dual-task divided attention decrements is the fundamental feature of resource theory (Kahneman et al., 1973). Scarc e mental resources are shared by tasks, and more difficult tasks leave fewer resources for concurrent tasks, whose performance declines as a result (line 1 of Figure 6.3).

The concept of mental effort is closely (and inversely) linked to that of automaticity (Schneider, 1985; Logan, 1985). A task that is said to be automated, like signing your name or following a familiar computer log-on procedure, has several properties. It is typically highly practiced, carried out rapidly with little conscious thought, and, most importantly, demands few mental resources for its execution, thereby improving the ability to perform other tasks at the same time. Automaticity is a matter of degree, not an all-or-none “thing.” So, the degree of

![Graph showing the relationship between performance of task B and difficulty of task A.](image-url)

**FIGURE 6.3**
Relation between performance of one task (B) and the difficulty of a second task (A) carried out concurrently, as resources are shared between them. Lines 1, 2, and 3 represent versions of task B that are progressively more automatized. At high levels of automaticity (line 3), perfect performance of task B can still be attained even when task A is quite difficult.
automaticity dictates the level of performance that can be obtained for a given investment of mental resources (Norman & Bobrow, 1975; Navon & Gopher, 1979; Wickens & Hollands, 2000). Lines 2 and 3 of Figure 6.3 represent dual-task performance of versions of task (B) that are progressively more automatized. As noted earlier, automaticity is one feature that results when perceptual elements are unitized.

**Structural Similarity**

Automaticity or resource demand is a property of a single task (or mental activity) that directly predicts its success in time-sharing with another. In contrast, structural similarity is the similarity between key processing structures of both tasks in a concurrently performed pair. Laura failed to monitor the converging traffic, in part because she could not see the road while looking at her cell phone. As she herself realized before starting the trip, she would have been more successful if she could have heard the navigational instructions via the voice mode, dividing attention between the eye and the ear. Researchers have argued that different structures in human information processing behave as if they were supported by separate or multiple resources, so that instead of the single "pool," shown in Figure 6.1, there are multiple pools (Navon & Gopher, 1979; Wickens, 1984, 2002). To the extent that two tasks demand separate resources, time-sharing is improved.

Table 6.1 lists four dichotomous dimensions of multiple resources for which there is generally better time-sharing between than within each end of the dimension. The table provides examples of activities that would "load" each end of a dimension. These four dimensions are partially independent, or "orthogonal," from each other so that, for example, a spatial or a verbal task (code dimension) can involve either perceptual-working memory activity or response activity (stage dimension). However, some of the dichotomies are nested within others. For example, the distinction between the focal and ambient visual channels is one that is only defined within processing that is visual and perceptual-cognitive.

The most important design feature to be derived from the table is that to the extent that two tasks demand common levels on one or more dimension, time-sharing is likely to be worse, and one or the other task will decrease farther from its single task-performance level. For example, a wide variety of research has shown that two tasks involving verbal material on the "code" dimension—like speaking while rehearsing a phone number—interfere more than a verbal and spatial task (Wickens & Liu, 1988). Regarding the modality dimension, research in driving and aviation generally supports the benefits of auditory display of information in the heavy visual demands of vehicle control (Wickens & Seppelt, 2002). Does this mean, for example, that it is always wise to present information auditorily rather than visually to the driver or pilot who has ongoing visual demands of vehicle control? Not necessarily, because sometimes altering the structure of information display may change the resource demand, our first contributor to dual-task interference. As an example, the auditory delivery of
TABLE 6.1 Four Dimensions of Multiple Resources

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Two Levels</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modalities</td>
<td>Auditory vs. Visual</td>
<td>Synthesized voice display, spatially localized tones</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Print, electronic map</td>
</tr>
<tr>
<td>Codes</td>
<td>Spatial vs. Verbal</td>
<td>Tracking, hand pointing, mental rotation, imaging (visuospatial scratchpad)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Listening to speech, rehearsing (phonetic loop), speaking</td>
</tr>
<tr>
<td>Stages</td>
<td>Perceptual–Working Memory vs. Response</td>
<td>Searching, imaging, reading, rehearsing, listening</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pushing, speaking, pointing, manipulating</td>
</tr>
<tr>
<td>Visual Channels</td>
<td>Focal vs. Ambient</td>
<td>Reading, interpreting symbols</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Processing flow fields, visual perception to maintain balance</td>
</tr>
</tbody>
</table>

(From Wickens, 2000).

long messages of five to nine chunks imposes a high resource demand that was not present in the long visual message, since the latter does not need to be rehearsed. Thus, only by considering both resource demand and structural similarity together can the degree of dual-task interference be predicted.

Confusion

We noted that the similarity between items in working memory leads to confusion. We also presented a corresponding argument regarding similarity-based confusion in our discussion of visual sensation in Chapter 4. Here also we find that concurrent performance of two tasks that both have similar material increases task interference (Fracker & Wickens, 1989; Gillie & Broadbent 1989 Wickens & Hollands, 2000). For example, monitoring basketball scores while doing mental arithmetic will probably lead to disruption as digits from one task become confused with digits relevant to the other. Correspondingly, listening to a voice navigational display of turn directions instructing a left turn, while the automobile passenger says, “right . . . that’s what I thought,” could lead to the unfortunate wrong turn. Auditory background information, because of its intrusiveness, may be particularly likely to cause confusion even if it is not part of an ongoing task (Banbury et al., 2001).

Task Management and Interruptions

In the previous section, we described the concept of total interference between two ongoing tasks, determined as a joint function of their combined resource demand, structural overlap, and possibly similarity. If these factors produce interference, then one task or the other will suffer a decrement. But will they both suffer? Or will one or the other be “protected”? In analyzing dual-task performance we typically speak of the primary task as that which should receive the
highest priority and will be buffered from the negative effects of high demand or structural similarity. The task that is degraded is referred to as the secondary task. The dual-task performer's decision to treat one task as primary and another as secondary is an example of task management. Thus, there would be no problem with cell phone use in cars if drivers consistently treated safe driving as the primary task and cell phone use as the secondary task. Unfortunately, not all drivers adhere to such optimum task management strategies, and cell phone-induced accidents are the result (Violanti, 1998).

At a most basic level, task management is simply the allocation of resources to one task or the other. However, this allocation can become considerably more complex than a simple two-state decision. For example, given that most people know (metacognition) that cell phone use (or other in-vehicle tasks) can divert resources from driving and road monitoring, why do drivers still engage in concurrent tasks? One reason is that successful time-sharing strategies can allow an optimal switching of attention between tasks. For example, the driver can sample a competing source of secondary-task visual information at a moment when he or she knows that there is little chance of something happening on the road ahead. When the car is on a straight stretch of freeway, with little traffic on a calm day, the vehicle inertia and absence of hazards can allow the eyes to scan downward for some time. As we described in the context of selective attention, there is little expectancy of important events on the "roadway channel." How long can the eye safely stay "head down"? This depends on a number of factors, such as the speed of the vehicle, the degree of traffic on the highway, and the degree of trust that a driver has that he or she will be warned of an impending event. Thus, the well-skilled driver can develop an accurate mental model of event expectancies and costs to support accurate scheduling of scanning (Moray, 1986; Wickens et al., 2003).

The previous discussion suggests that switching between tasks can be good, and in fact necessary, when parallel processing is impossible, as it is when information to support two tasks is displayed in widely separated locations. Indeed, if attention is switched or alternated fast enough between tasks, the result is indistinguishable from parallel processing. Consistent with this interpretation is the finding that people who more rapidly alternate between tasks may be more effective in their concurrent performance (Raby & Wickens, 1994). At the other end of the spectrum, very slow switching in a multitask environment can lead to cognitive tunneling (Moray & Rotenberg, 1989; Kerstholt et al., 1996); this is the process of keeping attention fixated on one task or channel of information long after a second task or channel should have been attended. In the context of memory failures, one can attribute such errors to forgetting the need to check the neglected task; a breakdown in prospective memory.

Human factors designs to avoid cognitive tunneling are imposed by reminders, as described earlier in the chapter (Herrmann et al., 1999). However, an even more basic human factors solution lies in the design of alarms, as discussed in Chapter 5. Alarms, particularly auditory ones, are specifically designed to interrupt whatever task is ongoing in order to redirect the user's attention to a problem that the system deems worthy of observation (Woods, 1995). It appears
important to train people how to handle interruptions in complex multitask environments like the cockpit (Dismukes 2001); (McFarlane & Latorella, 2002.)

Addressing Time-Sharing Overload

As our discussion suggests, there are a number of ways of addressing the multitask environment of the overloaded office secretary, vehicle driver, airline pilot, or supervisor of an organization in crisis. Briefly, we may subdivide these into four general categories:

1. **Task redesign.** On the one hand, we should avoid asking operators to perform too many tasks that may impose time-sharing requirements. In some environments, the military combat aircraft, for example, there is a temptation to load progressively more “mission tasks” on the pilot (e.g., weapons and surveillance systems). These must inevitably impose challenging time-sharing requirements, inviting overload. We noted earlier in the chapter how the CI interview technique for eyewitnesses explicitly avoids time-sharing of effortful memory retrieval and question comprehension (Fisher, 1999). On the other hand, we can sometimes redesign tasks to make them less resource-demanding. Reducing working memory demands is often successful—for example, users should not be required to remember a 10-digit phone number or even a seven-digit number in multitask situations.

2. **Interface redesign.** Sometimes interfaces can be changed to offload heavily demanded resources. As noted, there are many circumstances in which synthesized voice display can replace visual text when the eyes are needed for continuous vehicle control or monitoring (Dixon & Wickens, 2003; Wickens, Sandry, & Vidulich 1983).

3. **Training.** Explicit or implicit training of the operator, as we discuss in Chapter 18, has two different components in multitask environments. First, repeated and consistent practice at component tasks can develop automaticity (Schneider, 1985), thereby reducing resource demands (see Figure 6.3). Second, training in attention management skills can improve the appropriate allocation of resources (Gopher, 1993; Gopher et al., 1994; Wickens, 1989) and the handling of task switching and interruptions (Dismukes, 2001).

4. **Automation.** Automation also has two aspects relevant to dual-task performance. First, as we discuss in Chapter 16, many aspects of automation can either replace or greatly simplify resource-demanding aspects of performance—cruise control, the computer spell check, and the warning signal are typical examples. Second, designers have recently considered intelligent automation that can serve as a task manager, which can direct users’ selective attention dynamically to neglected tasks or assume performance responsibility for those tasks when required (Hammer, 1999).

CONCLUSION

In this chapter we discussed a number of mental processes that define the contents of cognitive psychology and lie at the core of much information processing in complex environments. These components find their relevance in many other
chapters of this book. Our discussion of perception links to chapters 4 and 5 on visual and auditory sensation, as well as to Chapter 8 on displays, where we consider designer artifacts that can support perception. Our discussions of attention relate to topics in both chapters 4 and 8 as well as to those of workload overload in Chapter 13. Issues of metacognition underlie people's tendency to engage in unsafe behavior, as discussed in Chapter 14. Cognition of all sorts is involved in computer usage (Chapter 15) and in dealing with automation and complex systems (Chapter 16) and transportation systems (Chapter 17). Cognition is knowledge, and knowledge is acquired through learning and training (Chapter 18). Finally, many aspects of cognition of perception and working memory are involved in the all-important task of decision making, the topic to which we turn in the next chapter.