Engineering Psychology and Human Performance

THIRD EDITION

CHRISTOPHER D. WICKENS
University of Illinois at Champaign-Urbana

JUSTIN G. HOLLANDS
University of Idaho

Prentice Hall
Upper Saddle River, New Jersey 07458
Attention in Perception and Display Space

OVERVIEW

The limitations of human attention represent one of the most formidable bottlenecks in human information processing. We can easily recall times when we failed to notice the words of a speaker because we were distracted or when we had so many tasks to perform that some were neglected. These intuitive examples of failures of attention may be described more formally in terms of three categories:

1. Selective attention. In some instances we select inappropriate aspects of the environment to process. For example, as we discuss in Chapter 8, decision makers sometimes select the cues that stand out rather than useful, diagnostic cues. The van driver described in Chapter 1 was so engrossed in reading the map display that he could not attend to the roadway ahead. We could say that his attention was too selective, in that important roadway information (the stalled car) was ignored until it was too late. Another dramatic example is provided by the behavior of the flight crew of an Eastern Airlines L–1011 flight that crashed in the Florida Everglades. Because they were preoccupied with a malfunction elsewhere in the cockpit, no one on the flight deck attended to the critical altimeter reading and to subsequent warnings that the plane was gradually descending to the ground (Wiener, 1977; see also Chapter 13). Here again, attention was too selective, a situation sometimes referred to as cognitive tunneling.

2. Focused attention. Occasionally we are unable to concentrate on one source of information in the environment; in spite of our desires to do so that is, we have a tendency to be distracted. The clerical worker transcribing a tape in a room filled with extraneous conversation encounters such a problem. So also does the translator who must ignore the feedback provided by his or her own voice.
to concentrate solely on the incoming message. Another example is the process
control room operator attempting to locate a critical item of information in
the midst of a “busy” display consisting of many changing variables. The dif-
ference between failures of selective and focused attention is that in the former
case there is an intentional but unwise choice to process nonoptimal envi-
ronmental sources. In the latter case this processing of nonoptimal sources is “driv-
en” by external environmental information despite the operator’s efforts to
shut it out (Yantis, 1993). Attention could not be focused on the appropriate
stimulus source.

3. Divided attention. When problems of focused attention are encountered, some
of our attention is inadvertently directed to stimuli or events we do not wish to
process. When problems of divided attention are encountered, we are unable to
divide our attention among stimuli or tasks, all of which we wish to process.
Here we may again consider our van driver, who must scan the highway for road
signs while maintaining control of the vehicle, or a fault diagnostician who must
maintain several hypotheses in working memory while scanning the environ-
ment for diagnostic information and also entering this information into a
recording device. Thus the limits of divided attention sometimes describe our
limited ability to time-share performance of two or more concurrent tasks, and
sometimes describe the limits in integrating multiple information sources.

Attention may be described by the metaphor of a searchlight (Wachtel, 1967). Two
properties of the searchlight are relevant: its breadth and direction. The beam’s breadth
can be subdivided into two components: that which we want to process (focused at-
tention), and that which we must process but do not want to (divided attention). The
direction of the searchlight—how it knows when, what, and where in the environment
to illuminate—describe the properties of selective attention. Each of these will be con-
sidered in detail as we consider examples of how operators search the complex stimu-
lus world for critical information and how the information is processed once found.

The searchlight metaphor describes the various characteristics of attention with re-
spect to perception, the topic of this chapter. Yet the concept of attention is relevant to
a range of activities beyond perception. We can speak of dividing attention between two
tasks no matter what stage of processing they require. The broader issue of divided at-
tention as it relates to the time-sharing of activities will be the concern of Chapter 11,
after we have discussed other stages of information processing. In this chapter we will
present an overview of the experimental findings of selective, focused, and divided at-
tention in perception and their relevance to display layout, addressing first those aspects
of attention that are serial (e.g., visual scanning) before considering its parallel charac-
teristics in vision and audition.

**SELECTIVE ATTENTION**

**Visual Sampling**

Our discussion of selective attention begins with the eye and with *visual sampling*, that
is, when the operator seeks information and searches for targets. Although selective at-
tention can occur without a change in direction of gaze (Egeth & Yantis, 1997), it is still
the case that for much of the time, our gaze is driven by our need to attend. Thus we can learn a lot about selective attention by studying visual scanning behavior, a close analog to the attentional searchlight (Fisher, Monty, & Senders, 1981; Moray, 1986).

Before we describe models of visual sampling, it is important to understand a few basic characteristics of the eye fixation system. First, only a small region of the visual field perceives detail. This region, the fovea, is about 2 degrees of visual angle. To keep objects in foveal vision, to “look at” them, the eyeball exhibits two different kinds of movement. Pursuit movements occur when the eye follows a target moving across the visual field. As you follow the trajectory of a ball or a flying bird, your eyes will show pursuit movements of roughly constant velocity. Saccadic movements are discrete, jerky movements that jump from one stationary point in the visual field to the next. They can sometimes be superimposed on pursuit movements. If the velocity of the moving ball or flying bird is too fast for pursuit movement, a saccade will be used to “catch up” and bring the target back into foveal vision (Young & Stark, 1963).

The saccadic behavior used in visual sampling has two components: the saccade and the fixation. During the saccade, the visual system suppresses visual input (Chase & Kalil, 1972), and so display information can be properly processed only during fixation. The fixation is characterized by a location (the center of the fixation), a useful field of view (diameter around the central location from which information is extracted), and a dwell time (how long the eye remains at that location).

Visual sampling behavior has been studied in two somewhat different applied contexts. In what we shall refer to as the supervisory control context, the operator scans the display of a complex system under supervision—an aircraft cockpit, for example—and allocates attention through visual fixations to various instruments, as these represent sources of information. In the target search context, the operator scans a region of the visual world, looking for something at an unknown location: it may be a failure in a circuit board examined by a quality control inspector (see Chapter 2), a search and rescue mission for a downed aircraft, or a receiver suddenly breaking into the open on the football field. In the supervisory control context, the location of the target (or targets) is known, but in target search, the observer must find a target whose location and existence is unknown (Liu & Wickens, 1992). We will discuss each of these situations in turn.

**Supervisory Control Sampling**

**Optimality of Selective Attention** In the aircraft cockpit or the process control console, many information sources must be sampled periodically. In these situations, engineering psychologists have studied how optimal performance is when the observer must select relevant stimuli at the appropriate times. As in our discussions of signal detection theory (Chapter 2), optimal was defined in terms of a behavior that will maximize expected value or minimize expected cost. For example, the van driver in Chapter 1 who continuously sampled the map display while ignoring the road ahead is not behaving optimally. If he sampled both the road and the map but never checked the fuel gauge, he is doing better but performance is still not optimal, for he is incurring the expected costs of missing an important event (running out of gas).

Engineering psychologists often divide the stimulus environment into channels, along which critical events may periodically occur. They assume that environmental
sampling is guided by the expected cost that results when an event is missed. The probability of missing an event in turn is directly related to event frequency and uncertainty (discussed in the vigilance section of Chapter 2). Those events that occur often are more likely to be missed if the channels along which they occur are not sampled, and when the timing of events is uncertain regularly scheduled samples will become less effective. In addition, the probability of missing an event on a channel typically increases with the amount of time since the channel was last sampled. For example, the probability of speeding increases with the time that has passed since the driver last looked at the speedometer.

When optimum sampling is examined in the laboratory, the subject is typically presented with two or more channels of stimulus information, along which events may arrive at semipredictable rates. For example, a channel might be an instrument dial, with an "event" defined as the needle moving into a danger zone, as in Figure 3.1 (e.g., Senders, 1964). Six general conclusions of these studies are described below. Some of these conclusions are based on summaries by Moray (1981, 1986).

1. Mental model guides sampling. People appear to form a mental model of the statistical properties of events in the environment and use it to guide visual sampling. The mental model consists of a set of expectancies about how frequently and when events will occur on each channel, and about the correlation between events on pairs of channels. As expertise develops, the mental model becomes refined, and sampling changes accordingly (Bellenkes, Wickens, & Kramer, 1997). Because sampling strategies provide estimates of the operator's mental model, the patterns of fixations should help the system designer arrange information

![Figure 3.1 Display typical of those used for studying instrument scanning. Under each display is an example of the time-varying input the operator must sample to ensure that none of the needles moves into the danger zones.](image)
Chapter 3  Attention in Perception and Display Space  73
displays so that optimal performance results. Dating from the pioneering work of Fitts, Jones, and Milton (1950), engineering psychologists have employed scanning data to configure displays according to two principles: Frequently sampled displays should be placed centrally, and pairs of displays that are often sampled sequentially should be located close together (Elkind, Card, Hochberg, & Huey, 1990; Wickens, Vincow, Schopper, & Lincoln, 1997).

2. Adjustment to event rate—sluggish beta. In line with the predictions of signal detection theory, people learn to sample channels with higher event rates more frequently and with lower rates less frequently. However, the sampling rate is not adjusted upward or downward with event frequency as much as it should be. This is similar to the sluggish beta phenomenon discussed in Chapter 2, in which observers were reluctant to adjust the response criterion in signal detection. To elaborate on this second point, some models (e.g., Carbonnell, Ward, & Senders, 1968) propose that the time between samples should be determined by two factors: the growth of uncertainty of the state of the unsampled channel (related to the event rate on that channel) and the cost of taking a sample. These factors trade off. Since sampling, or switching visual attention, has some subjective cost, people will not scan too rapidly across all channels of a dynamic instrument panel. Nor is there need for frequent sampling if channels change their state slowly (Channel 4 in Figure 3.1); hence, the operator’s uncertainty about the state of the unsampled channel grows slowly. But eventually the operator’s uncertainty will reach a high enough level so that it becomes worth the cost of a fixation to find out what is happening there (i.e., to “reset” uncertainty to zero).

Carbonnell, Ward, and Senders found that their model accurately described the fixation patterns of pilots making an instrument landing.

3. Sampling affected by arrangement. Donk (1994) examined eye movements in observers monitoring several instruments and found that they were more likely to make horizontal scans than diagonal scans. Donk also found that operators were reluctant to make diagonal scans to view high event rate channels. Donk proposed that operators use simplifying rules and heuristics based on channel arrangement to decrease attentional demands, which lead to systematic biases in performance. Hence, understanding instrument scanning just in terms of channels and event rates cannot completely account for performance—the arrangement of the instruments matters.

4. Memory imperfect; sampling imperfect. Human memory is imperfect, and sampling reflects this fact. People tend to sample information sources more often than they would need to if they had perfect memory about the status of an information source when it was last sampled. This fact explains the “oversampling” of channels with low event rates described above. Also, people may forget to sample a particular display source entirely if there are many possible sources, as might well be the case for the monitor of a nuclear process control console. Such forgetfulness will be more likely if the channels are not physically represented by a display location but are stored in a computer and must be accessed for inspection on a display screen by a manual or vocal request. These limitations in memory suggest the usefulness of “sampling reminders” (Moray, 1981).
5. **Preview helps.** When people are given a **preview** of scheduled events that are likely to occur in the future, sampling and switching become somewhat more optimal. Now subjects’ sampling can be guided by an “external model,” the display of the previewed events. Thus the dispatcher or industrial scheduler can be helped by having a preview of anticipated demands on different resources (Sanderson, 1989; see also Chapter 13), just as the student is helped by having a preview of upcoming assignments in different courses. However, as the number of channels increases, people fail to take advantage of preview information, apparently because of the heavy load on working memory required to do so (Tulga & Sheridan, 1980). This may be why predictive displays for industrial scheduling have not always been useful (Sanderson, 1989).

6. **Processing strategies—cognitive tunneling.** Scanning behavior may reflect the operator’s mental model of the environment, and therefore also reflect biases in the operator’s strategy. In a study of a simulated process control plant, Moray and Rotenberg (1989), for example, used a scanning analysis to determine that operators engaged in cognitive tunneling on a failed system. When one system under supervision “failed,” operators stopped examining the status of other systems as their diagnosis of the failed system was carried out. Moray and Rotenberg also used scanning measures to identify problems associated with delayed feedback. After making a control adjustment to one system, operators switch their visual attention to the indicator where feedback for that response is expected. Their fixation often stays locked on to that indicator until it eventually reflects the control input. This can represent a substantial waste of visual attention if the delay is long. Bellenkes, Wickens, and Kramer (1997) found similar results in the cockpit. They found that novice pilots performing high-workload maneuvers tended to focus on the most important instrument (the attitude directional indicator) and failed to carefully monitor other instruments, even though the information displayed on those instruments was also important for keeping the aircraft on the desired flight path. Wikman, Niemeinen, and Summala (1998) made a similar observation that novice drivers tended to dwell for significantly longer periods than experts as they scanned head down to tune a radio or dial a cellular phone.

**Eye Movements in Target Search**

When the operator is looking for an object in the environment, such as a flaw in a piece of sheet metal or the presence of survivors in aircraft wreckage on the ground, the visual scan pattern tends to be far less structured than in the supervisory/control task. As a consequence, scanning is less amenable to optimal modeling. Nevertheless, a number of characteristics of visual search have emerged.

**Environmental Expectancies** Like supervisory/control scanning, target search is driven in part by cognitive factors related to the expectancy of where a target is likely to be found. These areas tend to be fixated first and most frequently. This characteristic of information-seeking and scanning behavior has been used to account for differences
between novices and experts. In football, the expert quarterback will know where to look for the open receiver with the highest probability (Abernethy, 1988; Walker & Fisk, 1995). Kundel and LaFollette (1972) have studied differences in the way that novice and expert radiologists scan x-ray plates in search of a tumor. The expert examines first and most closely those areas in which tumors are most likely to appear; the novice tends to search the whole plate evenly.

The role of information in visual scanning has also been used to explain how we scan pictures (Yarbus, 1967). People tend to fixate most on areas containing the most information (e.g., faces, contours, and other areas of high visual detail). Furthermore, a scan path over the same picture will change, depending on what information the viewer is expected to extract (Yarbus, 1967).

**Display Factors and Salience** Since visual search behavior is often internally driven by cognitive factors, there is no consistent pattern of display scanning (e.g., left-to-right or circular-clockwise) and no optimal scan pattern in search, beyond the fact that search should be guided by the expectancy of target location. Nevertheless, certain display factors tend to guide the allocation of visual attention.

Visual attention will be drawn to display items that are large, bright, colorful, and changing (e.g., blinking), a characteristic that can be exploited when locating visual warnings (see Chapter 13) but that may bias decision making (see Chapter 8). These salient items can be used to guide or direct visual attention, as discussed in more detail later in this chapter. An abrupt stimulus onset (e.g., a light turning on) also serves to attract attention, especially in the visual periphery (Remington, Johnston, & Yantis, 1992; Yantis & Jonides, 1984). Yantis and Hillstrom (1994) have collected evidence suggesting that this may be due to the visual system being extremely sensitive to new perceptual objects.

Visual search is also captured by the presence of unique stimuli, often called *singletons*. For example, Theeuwes (1992) found that subjects were slower in finding a target (a diamond among circles) when one of the distractors or nontargets was unique in color (e.g., a green circle when all other circles were red). Thus, the presence of the unique features of the singleton slowed detection of other targets. However, singletons are less likely to capture attention when the target is defined in a more complex manner (Bacon & Egeth, 1994). Presumably, this is because when the targets are complex, the searcher does not have a "set" for simply defined targets like a singleton.

There is evidence also that search behavior is sometimes guided by physical location in the display. For example, Megaw and Richardson (1979) found that when subjects exhibited a systematic scan pattern in searching for targets, they tended to start at the upper left. This fact may reflect eye movement in reading. A search also tends to be most concentrated toward the center regions of the visual field, avoiding the edges of a display, a pattern that Parasuraman (1986) dubbed the *edge effect*. Also, as in supervisory/control sampling, scans tend to be made most frequently between adjacent elements on a display, and horizontal or vertical scans are more common than those along the diagonal.

These display-driven search tendencies are usually dominated by conceptually or knowledge-driven scan strategies (Levy-Schoen, 1981). However, it seems reasonable that a knowledge of these tendencies should be employed in designing multi-element
Chapter 3  Attention in Perception and Display Space

displays to locate information of greatest importance (e.g., warning and hazard labels) in areas of greatest salience, an issue that we will return to in Chapter 8 in the discussion of the cues used for decision making.

Display-Driven and Conceptually Driven Processing  Display-driven and conceptually driven strategies commonly interact: Theeuwes and Godthelp (1995) noted that standardization of roadway and sign design helps drivers know when to expect certain events. Hence, the driver responds to the stimuli in the road environment (signs, signals, intersections, interchanges) and interprets these in terms of conceptual expectations (e.g., “I know that the distance signs are large and green and occur just after an interchange; to figure out how far I am from Springfield, I’ll look for a large green sign after the next interchange”). It is also important to forecast the unexpected event, a technique called positive guidance (Dewar, 1993). For example, in North America, left exits off a freeway should be signed well in advance (Wickens, Gordon, & Liu, 1998). Creating an expectancy for the user and then making the display or stimulus salient can be an effective combination in driving the scanning behavior of an observer.

Search Coverage and the Useful Field of View  How much visual area is covered in each visual fixation? Although we can sometimes take in information from peripheral vision (see Chapter 4), resolution of fine visual detail requires the highest acuity region of the fovea, an angle of no more than about 2 degrees surrounding the center of fixation. Mackworth (1976) addressed this uncertainty by defining the “useful field of view” (UFOV) as a circular area around the fixation point from which information necessary for the task can be extracted. The size of a UFOV can be estimated from the minimum distance between successive fixations in a search task, on the assumption that two adjacent UFOVs touch but do not overlap. The data collected by Mackworth and others suggest that the size of the UFOV varies from 1 to 4 degrees of visual angle.

Several factors affect the UFOV. The size appears to be determined by the density of information and by the discriminability of the target from the background. Thus, looking for a dark flaw on a clear background in glass inspection will lead to a larger UFOV than scanning for a misaligned connection in a circuit board or microchip. Aging tends to lead to a restricted UFOV (Ball, Beard, Roenker, Miller, & Griggs, 1988; Scialfa, Kline, & Lyman, 1987). Scialfa et al. proposed that older adults take smaller perceptual samples from the visual scene and scan the samples more slowly than do young adults. However, training can enlarge the UFOV, and the benefits of training are equal across age groups (Ball et al., 1988). Reduction in UFOV has serious implications for tasks having visual search as a component, such as driving. Ball and Rebok (1994) found that vehicular crash frequency was greater for people having smaller UFOVs. Finally, the UFOV is sensitive to task demands in the foveal region (Williams, 1989). Williams found that as a foveal task becomes more difficult, information at the periphery of the UFOV is processed less well.

The size of the UFOV and the maximum rate with which different fixations can be made (2–4 per second) limit the amount of area that can be searched in a given time. However, even in the absence of time limits, it is apparent that humans do not search in exhaustive fashion, blanketing an entire area with UFOVs, and inevitably locating a target. Stager and Angus (1978) studied airborne search and rescue experts who searched photographs for crash sites; the searches covered only 53 percent of the available terrain, a fact that led to less than perfect performance. In addition, targets may be fixated within a UFOV
and yet not detected (Abernethy, 1988; Kundel & Nodine, 1978; Stager & Angus, 1978), suggesting that potential targets are measured against some decision criterion (like beta in signal detection theory, Chapter 2) during the search process. The advantage of training described above may be due to optimizing the placement of the decision criterion.

**Fixation Dwells** We have said little about how long the eye rests at a given fixation. Since the eye extracts information over time, one might think that long dwells should be associated with greater information pickup. Indeed, the attitude directional indicator found in the cockpit produces longer dwells (Bellenkes, Wickens, & Kramer, 1997; Harris & Christhilf, 1980) and is fixated on most frequently (Fitts, Jones, & Milton, 1950). This is presumably due to its high information content. Harris and Christhilf also found that pilots fixated longer on critical instruments (showing information necessary to control the aircraft) than on those requiring a mere check to assure that they were "in bounds." In target search, Kundel and Nodine (1978) distinguished between short survey dwells, used to establish those regions more likely to contain a target, and longer examination dwells, used to provide a detailed examination of the region for an embedded target.

In addition to scanning and sampling strategies, fixation dwells are also governed by the difficulty of information extraction. Thus, displays that are less legible or contain denser information will be fixated on longer (Mackworth, 1976). In normal reading, longer dwells are made on less familiar words and while reading more difficult text (McConkie, 1983; see Chapter 6). When examining pictures, people fixate longer on objects that are unusual and out of context (Friedman & Liebelt, 1981). As we saw in Chapter 2, low familiarity, low frequency, and out of context messages have higher information content, suggesting that dwell time has some relation to the information content of a display. In addition, expertise affects the difficulty of information extraction and, therefore, fixation dwell times. For example, Bellenkes, Wickens, and Kramer (1997) found that novice pilots dwell nearly twice as long on the information-rich attitude directional indicator as experts, requiring more time to extract the more difficult information. As noted above, Wikman, Niemeinen, and Summala (1998) found that novice drivers had longer head-down dwells than experts.

**Conclusion** The discussion of visual scanning behavior yields two general conclusions. First, scanning tells us a good deal about the internal expectancies that drive selective attention. Second, the greatest usefulness of scanning research to engineering psychology is probably in the area of diagnostics. Frequently watched instruments can be seen as those that are most important to an operator’s task. This fact may lead to design decisions to place these instruments in prominent locations or close together (e.g., Elkind, Card, Hochberg, & Huey, 1990; Wickens, Vincow, Schopper & Lincoln, 1997). Differences between novice and expert fixation patterns can indicate how the mental model or the search strategy of the novice departs from that of the expert, and display items that require long dwells may indicate nonoptimal formatting. We will revisit the topic of visual scanning in Chapter 6, where we examine visual fixations in reading, a task that is neither search nor supervisory control but is of great importance in design.

**Visual Search Models**

Visual scanning is of course heavily involved in visual search. However, there are other aspects of search that cannot be revealed by scanning, including such aspects as the
uncertainty of target identification or differences in the physical makeup of targets (e.g., one-dimensional versus multidimensional). Furthermore, whereas scanning reveals details about the process of visual search, human factors engineers may often be interested in the product of that search: How long does it take to find a target? Or what is the probability that a target will be detected in a given period of time? Hence, engineering psychologists have been concerned with the development of visual search models that will allow these values to be predicted.

One such model was developed by Drury (1975, 1982) to predict the time it would take an industrial quality control inspector to detect a flaw in a product. Drury examined the inspection of sheet metal. The model has two stages. The first stage describes the target search and predicts that the probability of locating a target will increase with more search time. However, it will increase at a diminishing rate, as shown in Figure 3.2. This is not surprising, given that (1) a target may be fixated on more than once without being detected, and (2) search strategies do not usually cover the whole search field with UFOVs, even when adequate time is given. In a later “decision” stage, the operator uses the expectancy of flaws (the overall manufacturing quality) to set a decision criterion, as in signal detection theory (Chapter 2). If the expectancy of a flaw is high, the criterion will be set low.

The shape of the curve in Figure 3.2 has important implications for the designer of industrial inspection stations: There is an optimal amount of time that each product should be searched, given that one can specify a cost for inspection time (which increases linearly with longer time) and a cost for misses. If the operator searches for a longer time to achieve a higher detection rate, this leads to diminishing gains in inspection accuracy. Drury (1975, 1982) discusses how this optimal time could be established, given factors like the desired rate at which products should be inspected (often set by a manager), the probability of fault occurrence, and the desired overall level of inspection accuracy. Then industrial material to be inspected can be presented at a rate determined by the optimal time.

![Figure 3.2](image-url)

*Figure 3.2* Probability of detection as a function of time available for search.

*Source: Adapted from C. Drury, "Inspection of Sheet Metal: Model and Data," *Human Factors, 17* (1975). Reprinted with permission. Copyright 1975 by the Human Factors Society, Inc. All rights reserved.*
Another, more basic, approach has been to model the kinds of variables that affect search speed through a set of stimuli to locate a particular target. In these tasks, the operator searches through an array such as that shown in Figure 3.3 and might report:

1. the presence of a white X
2. the presence of a large T
3. the presence of a black target

Extensive research in this area reveals a number of general conclusions. First, in situations like Tasks 1 and 2, the number of elements to be searched has the dominant effect on search time (Drury & Clement, 1978; Treisman & Gelade, 1980). This is because the search is usually serial, as each item is inspected in turn. If there are more items, search times will increase. Many researchers have replicated this finding (e.g., Egeth & Dagenbach, 1991; Wolfe, Cave, & Franzel, 1989). The slope of the function for trials without the target reflects the average time required to scan each item in the array and is about 50 milliseconds per item for simple items like letters. The slope of the function for trials with the target is roughly half that of target-absent trials (e.g., Wolfe, Cave, & Franzel, 1989). This is what one would expect for a search that is serial and self-terminating. In a serial search, each item is inspected in turn. In a self-terminating search, the search stops when the target item is found. Thus, when a target is present it will be found, on the average, after about half of the items have been inspected.

Second, exceptions to the first conclusion regarding serial search occur when the target is defined by one level along one salient dimension (Treisman & Gelade, 1980). For example, performance of Task 3 above will be little affected by the number of items, since the target in Figure 3.3 is defined by a single level (black) of one dimension (color). It appears to "pop out" of the search field. That is, parallel search can occur when the target can be defined using a simple rule. Eye movements correlate with the performance data, showing greater search efficiency (fewer scans) for parallel than serial search (Williams, Reingold, Moscovitch, & Behrmann, 1997). Some visual search models (e.g., Treisman & Gelade, 1980; Wolfe, 1994) propose that parallel search of this type is preattentive (requiring few

![Figure 3.3 Stimuli for a typical experimental search task.](image-url)
attentional resources) and can be done across the entire visual field, whereas serial search
requires attentional resources, and can only be done over a limited portion of the visual
field (i.e., the UFOV).

Third, serial search is more likely when the target is difficult to discriminate from
distractors (Geisler & Chou, 1995). Nagy and Sanchez (1992) found that search times
increased with number of distractors when the luminance or color difference between
target and distractor was small, but search times did not increase when the difference
was large.

Fourth, exceptions to serial search also occur when the target is defined by having
a feature present rather than absent. For example, Treisman and Souther (1985) showed
that parallel search occurred when subjects searched for a Q among Os, but serial search
occurred when searching for an O among Qs. This is similar to the “target-present” ad-
vantage noted in the vigilance situation in Chapter 2 (e.g., Schoenfeld & Scerbo, 1997).
Again, this can be interpreted in terms of different discriminabilities of targets in the
two situations (Geisler & Chou, 1995).

Fifth, it matters relatively little if the elements are closely spaced, requiring little
scanning, or are widely dispersed (Drury & Clement, 1978; Teichner & Mocharnuk,
1979). The increased scanning that is required with wide spacing lengthens the search
time slightly. However, the high density of nontarget elements with closely spaced items
also lengthens search times slightly. Thus scanning distance and visual clutter trade off
with one another as target dispersion is varied.

Sixth, searching for any of several different target types is generally slower than
searching for only one (Craig, 1981). An example would be to “Search for a P or a Q”
in Figure 3.3. The exception occurs when the set of two (or more) targets can be dis-
criminated from all other nontargets by a single common feature (e.g., color). Varying
levels of training may be necessary for the perceptual system to tune in to this critical
discriminating feature. For example, in Figure 3.3, if the instructions were to “search
for an X and a K,” subjects might learn that given the particular set of nontarget stim-
uli used, X and K are the only letters that contain diagonal lines, and hence they will
be able to search efficiently for this unique shared feature (Neisser, Novick, & Lazar,
1964). Thus, there should be an advantage to training industrial inspectors to focus on
the set of unique and defining features common to all faults, distinguishing them from
normal items.

Seventh, the role of extensive training in target search can sometimes bring perform-
ance to a level of automaticity, when search time is unaffected by the number of targets
and presumably done in parallel (Fisk, Oransky, & Skedsvold, 1988; Schneider & Shiffrin,
1977). Generally speaking, automaticity results when, over repeated trials, targets never
appear as nontarget stimuli, a condition that Schneider and Shiffrin refer to as consistent
mapping. This is contrasted with varied mapping search, when a target may later appear
as a nontarget. We will discuss the concept of automaticity further in Chapter 6 in the
context of reading, in Chapter 7 in the context of training, and in Chapter 11 in the con-
text of time-sharing.

Although these studies were conducted in the laboratory, they have clear applica-
tion to a variety of work domains. For example, a vehicle dispatcher might need to scan
a computerized city map to locate a vehicle that is not in service and has a large carry-
ing capacity. The military commander must find a particular subset of symbols on an
electronic topographic map. Visual search is commonplace in many work domains, and the factors listed above are likely to play a role in the efficiency of those searches.

**Application: Symbol Coding** In the above examples, a symbol may be used to code multiple dimensions, so that its color represents one dimension, its size another, its shape a third, and so on. Imagine the operator is trying to find a particular target stimulus. When multiple levels of multiple dimensions define the target, serial search results, as noted previously (Treisman & Gelade, 1980). Hence, to determine if each symbol represents the target, the level of each dimension will be examined serially. This serial examination has two implications. First, with more coding dimensions, search times will increase. Second, if operators search the dimensions in a particular order, it implies that the discriminability of two symbols is not just a simple matter of the number of features in common and the number of unique features (e.g., Geiselman, Landee, & Christen, 1982; Tversky, 1977) but is determined by the specific order in which features are examined (Fisher & Tanner, 1992). To develop an optimal symbol set (i.e., to maximize symbol discriminability) the designer must take into account the order of the search through the dimensions. If this is known, the maximally discriminable symbol set can be determined from an algorithm developed by Fisher and Tanner.

**Structured Search**

**Basics** The model proposed by Drury (1975) describes a search in which a target could be located anywhere and there is little organization to guide the search (sometimes called free field search). Somewhat different is the process of structured search, in which information that may help guide the search is available in the display. For example, structured search might occur when a computer user wishes to locate a particular item on a menu or an airline passenger is scanning a TV monitor for information concerning a particular flight. When we perform a structured search, we examine each item in the set in a systematic order, making structured search more amenable to modeling than search in a free field. In the letter-search task developed by Neisser (1963), subjects scan a vertical column of random three- or five-letter sequences until they detect the target letter, as shown at the top of Figure 3.4. The researchers observed a linear relationship between the serial position of the letter in the list and the time needed to detect the target, as shown at the bottom of Figure 3.4. The slope of this function, which represents the time to process each letter in this structured search, is roughly the same as what was observed in free-field search (50–100 milliseconds per item).

**Application: Menus** One important application of structured search models is in the design of computer-based menu systems, a critical component in human–computer interaction. In the typical menu task, the user must locate a target word, symbol, or command. The user must scan the list until the item is located, and then press a key. Menus may be multilevel, in which case the target term may be reached only after a search through higher-level terms. Thus, a travel agent, searching for a flight from a particular city, may first access a menu of city names and then, after selecting an option within that name level of the menu, scan an embedded menu of all flights departing from that city.
Menu designers would like to structure a menu in such a way that target items are reached in the minimum average time, and the linear visual search model can serve as a useful guide. If menus are organized randomly, given the general tendency to search from the top downward (Somberg, 1987) and the linear search strategy, the target will be located after an average of \( NT/2 \) seconds, in a self-terminating search where \( N \) is the menu size and \( T \) is the time to read each item (Lee & MacGregor, 1985). Within each search, the time will be directly proportional to the distance of the item from the top of the menu. It is possible for designers, capitalizing on this linear search strategy, to reduce the expected search time if they know that some menu items will be searched for more often.
than others. These items can be positioned toward the top of the menu in proportion to their frequency of use. Using these assumptions, Lee and MacGregor (1985) have developed quantitative models that predict the expected time needed to locate a target item as a function of reading speed and computer response speed when there are embedded (multilevel) menus. Their model guidelines dictate that the optimal number of items per menu is between three and ten, depending on the reading speed and the computer response speed. Their data, consistent with others to be described in Chapter 9, argue against many embedded levels of short menus.

The Lee and MacGregor (1985) model ignores the effects of the similarity of the target item to the correct menu alternative, and the similarity of menu alternatives to each other. Pierce, Parkinson, and Sisson (1992) found that when the target item was highly similar to the correct alternative (e.g., the target item is Ballet, and the correct alternative is Dance), search was faster than when the target word was less similar (i.e., more generic; e.g., target item Ballet, correct alternative Art). When the various menu alternatives were made more similar, search was slower than when alternatives were less similar reflecting the third conclusion in visual search discussed above. A criterion-based model developed by Pierce, Sisson, and Parkinson (1992) accounts for the effects of similarity in menu search using a signal detection approach (see Chapter 2) in which the user evaluates a menu alternative in terms of its perceived similarity to the target item. There are two criteria in the model. If the alternative is seen as very different from the target item, then it falls below the lower criterion and is immediately rejected as a response. If the alternative is seen as highly similar to the target item, then it falls above the upper criterion, it is selected as a response, and the search terminates. If the alternative falls between the two criteria, search continues until an alternative falling above the upper criterion is found or the alternative with the highest similarity is selected. When menu alternatives are made more similar, the signal detection distributions representing correct and incorrect alternatives move together, increasing the likelihood of alternatives that fall between the two criteria. The model successfully accounted for the data collected by Pierce, Parkinson, and Sisson.

Such quantitative models are an important first step in understanding structured search with computer menus, and they hold up in a variety of situations. Nonetheless, people perform other tasks with menus than simply finding a target word, and the type of menu organization (e.g., alphabetic, semantic) can affect search effectiveness (Halgren & Cooke, 1993; Hollands & Merikle, 1987; Meleenbacher, Duffy, & Palmer, 1989; Smelcer & Walker, 1993). A comprehensive model of structured search with computer menus must account for such results.

In conclusion, we note that quantitative models of human visual search and scanning performance are fairly successful. Although they do not succeed in predicting exactly how an operator will accomplish a task or how long it will take for an item to be located, the answers they provide are at a more precise level than those offered by intuition. Visual search is only a small component of human performance, but it offers a success story in the domain of performance models.

**Directing Attention**

It is sometimes possible to advise an operator in advance where attention should be directed. An air traffic controller’s attention, for example, might be directed toward a pending conflict
if the symbols for the involved aircraft begin flashing. In the laboratory, this has been investigated by presenting a cue just before the onset of a faint target, at the same location as the target. Detecting the target becomes more accurate as the stimulus-onset asynchrony (SOA) between the warning (or cue) and the target increases (e.g., Eriksen & Collins, 1969). That is, if the cue appears 200 milliseconds before the target (SOA = 200 ms), it is more effective than if it appears 50 milliseconds before the target (SOA = 50 ms), allowing the subject more time to redirect attention to the cued location. But cueing helps (relative to no cue) even with a 50-millisecond SOA.

We can distinguish between situations where peripheral cues are used (i.e., cues at the pending target location, which is typically out of foveal vision), and where the cue is in some neutral foveal location but indicates the target location in some way (e.g., by using an arrow pointing in the target’s direction). This second type is called a central cue. Central cues are more effective with longer SOAs (e.g., 400 ms), and their benefits tend to be fairly long lasting; peripheral cues are typically more effective with short SOAs (Muller & Rabbitt, 1989) and have a more transient effect. Egeth and Yantis (1997) refer to peripheral cues as stimulus-driven and central cues as goal-directed, implying different mechanisms for the different types of cueing. The peripheral cues appear to be processed automatically, whereas central cues require controlled interpretation (Muller & Rabbitt, 1989). From a designer’s perspective, if a cue (e.g., a warning or prompt) cannot be presented until the last moment, a peripheral cue should be more effective, but a central cue is probably more effective otherwise, since its attention-directing effects are longer lasting.

It appears that cues can direct the spotlight of attention. One might suppose that the spotlight moves in analog fashion. Thus, as you switch attention from the pointer to a particular number on your speedometer, your attentional spotlight would move continuously as you make the switch. However, the evidence suggests otherwise. Eriksen and Webb (1989) failed to find a relation between the time to shift attention and the distance between elements when eye movements were not involved, a relation which one would expect if the attentional shift were continuous. The results are more consistent with attention moving in discrete, “all-or-none” fashion. This implies that intermediate elements in the display would not be attended as the switch was being made.

If cues are not perfectly valid indicators of a target, as may often occur in operational settings, a cost-benefit relationship results (Posner, 1986). Assume that the cue is 80 percent reliable in directing attention to the eventual location of the target. The observer can attend to the cue and be faster and make fewer errors on those trials when the cue is accurate but will suffer on the 20 percent of trials when the cue is inaccurate, taking longer and making more errors as a result (Posner, Nissen, & Ogden, 1978). Posner, Nissen, and Ogden found that this was not simply a result of eye fixations—the same results occurred in cases where there was no eye movement.

The question then arises as to whether the fewer errors made in response to an accurate cue are a result of lowering a criterion for detecting a signal at the target location (beta in signal detection theory, discussed in Chapter 2) or an increased sensitivity to the target location (Kinchla, 1992). If the change is in the response criterion only, this means that the observer is responding to the cue (i.e., their attention is directed to the
correct part of the display), but this does not increase their accuracy (i.e., sensitivity), since they make more false alarms at that location (i.e., a shift in beta). It appears that the typical result is a change in both $d'$ and beta (e.g., Downing, 1988).

**Attention in Depth** We have discussed how certain stimulus cues can be used to direct a person's attention to a particular location on a two-dimensional display. A person's attention might be directed in three-dimensional space analogously. For example, Atchley, Kramer, Andersen, and Theeuwes (1997) used a cue to indicate the approximate depth at which a signal was to occur (they used stereopsis information to produce the sensation of depth). Subjects took longer to respond to the signal when the cue was at a different depth from the signal than when both cue and signal were presented at the same depth. That is, the cue produced a focus of attention at its particular depth. However, this effect may be reduced or eliminated (i.e., the spotlight of attention may be “depth-blind”) if the target is difficult to detect or discriminate (Chiradelli & Folk, 1996).

There are also benefits to showing information at different depths. Chau and Yeh (1995) and Theeuwes, Atchley, and Kramer (1998) had observers detect a target that was separated in depth or not separated in depth from a background containing distractors. When the target was separated in depth from the distractors, search times were shorter.

These results imply that unusual distractor stimuli that typically slow search for a target will not slow search if they are at a different depth from the target. If a target and distractor stimuli are separated in depth, it might be worthwhile to preserve that depth information on a display screen so that an observer can more easily filter out distractors. For example, air and ground objects on a radar screen might be better displayed with a stereo vision facility in order to assist an observer in distinguishing an air target from ground objects.

**Applications** The topic of directing attention becomes of greater consequence when we consider that automated systems are being developed to provide *intelligent cueing* in various operational settings. This intelligent cueing directs the user's attention to certain target regions in the display or the world. Yeh, Wickens, and Seagull (1999) investigated the effectiveness of target cueing in the design of see-through helmet-mounted displays. They found that cueing lowered target detection response times for expected targets but made it more difficult to detect unexpected targets—targets of greater potential danger—both in terms of longer response times and more errors (i.e., more missed targets). This result echoes Posner's (1986) cost-benefit result described above. Conejo and Wickens (1997) cued pilots in a simulated air-ground targeting task. When the cue was unreliable, directing attention to an object that was similar to, but was not the designated target, pilots often chose the non-target, even when the correct target was visible on the display and the pilot knew what the target looked like. This result reflects the role of the response criterion (beta shift) in target cueing (Downing, 1988). Other researchers (e.g., Mosier, Skitka, Heers, & Burdick, 1998; Taylor, Finnie, & Hoy, 1997) have found similar results. In combination, these results suggest that cueing can be effective in directing attention for a variety of tasks, but people sometimes tend to follow and believe the cues indiscriminately—an example of excessive trust in automation, a topic to be discussed in Chapter 13.
In addition to cueing, attention may also be directed implicitly in a complex display by highlighting a selected subset of items that some agent infers should be attended (Hammer, 1999). For example, an intelligent filtering system that infers what would be of interest to a reader might highlight a set of document titles within a longer list. Or all aircraft on an air traffic controller display that lie within a certain, relevant, altitude range could be highlighted. Many different physical techniques can be employed to highlight "relevant" items and allow this subset to be easily scanned without distraction from the non-highlighted items (Fisher & Tan, 1989), such as color or intensity coding, boxing, underlining, flashing, or reverse video. The particular technique should be carefully chosen so that the features that may make a set of items stand out (and therefore be easily detected and discriminated from the nonhighlighted options) do not themselves disrupt the ability to read or interpret the items. For example, flashing words may be very difficult to read. Uniquely colored items do not appear to suffer this deficit (Fisher & Tan, 1989).

It is often difficult for the agent driving the highlighting to guarantee that every highlighted item is relevant, and that all "background" items are not relevant. For example, in the case of document search, some documents assumed to be relevant on the basis of keywords (and therefore highlighted) may not, in fact, be of any interest. This defines the issue of highlighting validity. Indeed, the extent to which a person uses highlighting to guide search (decreasing the effective number of items to be searched, since the background items can be easily ignored) is based on the user's expectancy that the highlighting is indeed valid; however, even validity that is considerably less than 1.0 will still enable users to search a highlighted subset for a target first, with the result that total search time will be reduced (Fisher, Coury, Tengs, & Duffy, 1989).

PARALLEL PROCESSING AND DIVIDED ATTENTION

The first part of this chapter addressed those aspects of attention and perception that are often serial, as in the search or supervisory/control task. Yet even in this discussion, we alluded to situations where processing is parallel rather than serial. In models of scanning, we discussed the useful field of view, with the assumption that several items within that field might be processed together (in parallel). In reading, there is good evidence that when we fixate on a short word, all letters within that word are processed in parallel (see Chapter 6). We also noted that when a target is defined by one level along one salient dimension or by an automatically processed stimulus, search time did not depend on the number of elements, suggesting that the elements were processed in parallel.

In the last half of this chapter, we will focus on aspects of perceptual processing that operate in parallel. We speak of divided rather than selective attention in this case. Although divided attention and parallel processing are often good things for human performance—particularly in high-demand environments such as an air traffic control center or a busy office—it is sometimes impossible to narrow the focus of attention when needed and shut out unwanted inputs. This failure occurs when divided attention becomes mandatory rather than optional. In this case we speak of a failure of focused attention as being the downside of successful divided attention. In particular, many display principles that facilitate divided attention impair focused attention. For example, in the previous chapter, we saw that integral dimensions help when operators can divide their attention between two
redundant dimensions but hurt when they must focus attention on one while ignoring independent changes in the other. Because of this close and sometimes reciprocal relationship between divided and focused attention, our discussion will often treat the two topics in consort. We begin by considering parallel processing at the earliest phases of the visual information-processing sequence; we then consider the role of space, objectness, and color in attention; finally, we shift our discussion to parallel processing and focused and divided attention in the auditory modality.

**Preattentive Processing and Perceptual Organization**

Many psychologists have argued that the visual processing of a multiple-element world has two main phases: A *preattentive* phase is carried out automatically and organizes the visual world into objects and groups of objects; then we *selectively attend* to certain objects of the preattentive array for further elaboration (Kahneman, 1973; Neisser; 1967). These two processes might be associated with short-term sensory store and perception, respectively, in the model of information processing presented in Figure 1.3. Thus, distinguishing between figure and background is preattentive. So also is the grouping together of similar items on the display shown in Figure 3.5a. Gestalt psychologists (e.g., Wertheimer; see Palmer, 1992) made efforts to identify a number of basic principles that cause items to be preattentively grouped together on the display (e.g., proximity, similarity, common fate, good continuation, closure; see Palmer, 1992) Displays constructed according to these principles have high redundancy (Garner, 1974). That is, knowledge of where one display item is located will allow an accurate guess of the location of other items in a way that is impossible with the less organized display shown in Figure 3.5b. Indeed, Tullis (1988) and Palmiter and Elkerton (1987) have developed a set of information-theory-based measures of display organization that can be used to quantify the organization of alphanumeric and analog displays, respectively. Because all items

![Figure 3.5](image-url)
of an organized display must be processed together to reveal the organization, such parallel processing is sometimes called global or holistic processing, in contrast to the local processing of a single object within the display.

Two examples illustrate the differences between global and local processing. One example, shown in Figure 3.6, is a stimulus presented to subjects by Navon (1977). Figure 3.6a shows a large F made up of a series of small T’s. When subjects are asked to report the name of the large letter, there is a conflict. The small letters perceived by local processing lead one to respond T, whereas the large letter requiring global processing leads one to respond F. This response conflict is not present in Figure 3.6b. But this interference is asymmetric. When asked to report the large letters, there is little interference from the incompatibility of the small. Thus, the global aspects of the stimulus appear to be automatically processed in a way that makes them immune to the local aspects, for which more focused attention is required. This phenomenon is known as global precedence.

A second example is the texture segregation shown in Figure 3.7. At the top of the figure, the vertical T’s appear more different from the slanted T’s than from the L’s on

\[
\begin{array}{cc}
TTTT & FFFF \\
TTT & F \\
TTT & FFFF \\
TT & F \\
T & F \\
T & F \\
\end{array}
\]

(a) (b)

**Figure 3.6** Global and local perception. (a) Global and local letters are incompatible; (b) global and local are compatible.

\[
\begin{array}{cc}
LLLTTT & TTTTTT \\
LLLTTT & TTTTTT \\
LLLTTT & TTTTTT \\
LLLTTT & TTTTTT \\
LLLTTT & TTTTTT \\
LLLTTT & TTTTTT \\
\end{array}
\]

**Figure 3.7** Global versus local perception. On the top (global perception), contrast the L’s (left) with the T’s (center) with the slanted T’s (right). The distinction between the T’s and slanted T’s is greater. However, in the bottom (local perception), the distinction between the L’s and T’s is at least as great as between the T’s and slanted T’s.
Chapter 3  Attention in Perception and Display Space  

The concepts of global and local processing are closely related to the *emergent features* concept discussed in Chapter 2. An emergent feature is a global property of a set of stimuli (or displays) not evident as each is seen in isolation. Consider the two sets of gauges shown in Figure 3.8, in which the normal setting of each gauge is vertical. The vertical alignment of the gauges on the top set allows more rapid detection of the divergent reading because of the emergent feature—a long vertical line—present in the top set but not in the bottom (Dashevsky, 1964).

Because global or holistic processing tends to be preattentive and automatic, it might reduce attentional demands as an operator processes a multielement display. But this savings is only realized under two conditions: First, the Gestalt principles based on information theory (e.g., redundancy) should be used to produce groupings or emergent features. Second, the organization formed by the spatial proximity of different elements on the display panel must be *compatible* with task demands. Thus, for example, in Figure 3.5a, the organization of the displays will not be helpful, and may even be harmful, if the task performed by the operator requires constant comparison of information presented in dials in the top-left with the bottom-right groups. We refer to this as a violation of *compatibility* between the display and task requirements. Some nuclear power consoles were designed with the panels for two reactors lying side by side, one the mirror image of the other. This configuration provided wonderful symmetry, which at a global level provided organization, but it made it difficult for the operator to switch between panels.

![Figure 3.8](image)

**Figure 3.8**  Global perception in the detection of misalignment.
Spatial Proximity

Overlapping Views: The Head-Up Display The previous discussion suggests that spatial proximity, or closeness in space, should also enable parallel processing (and therefore help divided attention). Although one cannot simultaneously look at the speedometer and look out the windshield at the road, a display that could superimpose a view of the speedometer on a view of the road should facilitate divided attention or parallel processing between the two channels (Goesch, 1990; Tufano, 1997).

However, although spatial proximity will allow parallel processing, it certainly will not guarantee it. For example, in an experiment by Neisser and Becklen (1975), subjects watched a video display on which two games were presented simultaneously, one superimposed over the other. One showed distant figures tossing a ball, the other showed two pairs of hands playing a clapping game. One game was designated as relevant, and critical elements were to be monitored and detected. Neisser and Becklen found that while monitoring one game, subjects failed to see events in the other game, even when these were unusual or novel (e.g., the ball tossers paused to shake hands). They also had a difficult time when detecting events in two games at once. These results suggest that separation may be defined not only in terms of differences in visual or retinal location but also in terms of the nature of the perceived activity.

Neisser and Becklen's (1975) display has a counterpart in aviation, the head-up display (HUD) (Newman, 1995; Weintraub & Ensing, 1992; Wickens, 1997), which shows critical instrument readings on the glass windscreen superimposed on the forward view, as shown in Figure 3.9. Similar displays are being introduced into the automobile (Goesch, 1990; Tufano, 1997). The HUD was designed to ensure that information inside and outside an aircraft could be processed simultaneously without visual scanning. Neisser and Becklen's results suggest that this may not occur. A pilot may treat the two information sources as different attentional channels and become engrossed in processing instrument information on the HUD while ignoring critical cues from outside the aircraft, a phenomenon observed in experiments by Fischer, Haines, and Price (1980) and Larish and Wickens (1991). In studies using pilots as subjects, Wickens and Long (1995) found that an unexpected obstacle, an airplane crossing the runway, was detected more poorly with the HUD than with the head-down configuration. This airplane may be seen, poised to "move out," in Figure 3.9c. However, the HUD does have its advantages (Fadden, Ververs, & Wickens, 1998). Wickens and Long (1995) showed that a HUD could improve control of position during landing, both when in view and when the runway was obscured by clouds. Sojourner and Antin (1990) compared driver performance with HUDs and head-down displays and found a HUD advantage for detecting cues presented in the road scene. Other studies have also found HUD advantages relative to head-down presentation of the same information (e.g., Martin-Emerson & Wickens, 1997; Ververs & Wickens, 1998).

These apparently contradictory results appear to hinge on the expectations of the observer. The HUD appears to facilitate parallel processing of scene and symbology when the pilot expects the stimulus (e.g., the appearance of a runway during landing, objects that occur repeatedly during driving) and interferes when the stimulus is quite unexpected (e.g., a small airplane crossing the runway). A second factor affecting the costs and benefits of overlapping imagery is the conformal nature of the symbology itself, to be discussed below in the context of object displays.
Figure 3.9 Head-up display used in aviation. (a) Head-up Guidance System (HGS). Courtesy of Flight Dynamics, Inc., Portland, Oregon. (b) Head-up display similar to that used by Wickens and Long (1995) and Martin-Emerson and Wickens (1997) with conformal imagery. Note the runway overlay. (c) Head-up display with nonconformal imagery. Note the airplane on the ground at the far left.

Visual Confusion, Conflict, and Focused Attention Although close proximity in space may sometimes allow more successful divided attention, it appears that it may increase confusion between those items that are momentarily the desired focus of attention and those that are not—that is, a failure to focus attention. Several pieces of evidence support this claim.

First, as we saw earlier in this chapter, in visual scanning, the spatial density of the objects has little effect on visual search time. With a high-density field, any advantages that may be realized in terms of more items per fixation will be negated by the increased clutter. Second, in a study in which subjects monitored several display indicators, Wickens and Andre (1990) found that the most critical variable in predicting performance is the degree of spatial separation of relevant from irrelevant items, not the spatial separation between the relevant items themselves. Third, a study by Holahan, Culler, and Wilcox (1978) found that the ability to locate and respond to a stop sign in a cluttered display is directly inhibited by the proximity of other irrelevant signs in the field of view.

The fourth piece of evidence is found in a classic study by Eriksen and Eriksen (1974), which will be discussed in more detail because it sets the stage for the discussion of object displays. In this experiment, subjects moved a lever to the right if the letter H appeared and to the left if the letter F appeared. Reaction time (RT) was measured in this control condition. In other conditions, the central target was closely flanked by two adjacent letters, which were irrelevant to the subjects' task and were therefore to be ignored (e.g., UHF). The presence of these flanking letters slowed RT relative to the control condition. This is the result of perceptual competition, a failure of focused attention caused by the competition for processing resources between close objects in space.

In the particular case in which the flanking letters are mapped to the opposite response (i.e., an H flanked by F's: FHF), RT is slowed still further. There is now an added cost to processing, which Eriksen and Eriksen (1974) describe as response conflict, a concept that we introduced in the context of Navon's (1977) experiment on global and local processing. It illustrates more clearly the failure of focused attention. It is as if the navigator sitting next to the automobile driver were saying, "Turn left," while a passenger in the back seat, engaged in a different conversation, says, "Yeah, right." Only when the flanking letters were identical (i.e., an H surrounded by two other H's: HHH) were RTs faster than in the control condition. This is another example of redundancy gain.

Response conflict and redundancy gain are thus two sides of the same coin. If two perceptual channels are close together, they will both be processed, even if only one is desired. This processing will inevitably lead to some competition (intrusion or distraction) at a perceptual level. If they have common implications for action, the perceptual competition is overcome because both channels activate the same response. If, however, their implications for action are incompatible, the amount of competition is magnified.

In real-world displays, perceptual competition and redundancy gain effects are more likely to be observed with greater display clutter. Flanking letters interfere most if they are close to the target (e.g., about 1 degree of visual angle), as if there is a minimum diameter of the spotlight of attention that guarantees some parallel processing (Broadbent, 1982). Parallel processing is less if flankers are placed farther from the target (e.g., 2–3 degrees) (Murphy & Eriksen, 1987). However, if the observer cannot be certain about the location of the target, then the interfering effect of flankers can occur at these greater distances (Murphy & Eriksen, 1987). Yantis and Johnston (1990) found that flanker effects could
almost be eliminated by cueing subjects about target position. One can account for such results by using the spotlight metaphor. When the target location is uncertain, the observer must broaden the spotlight, which means that the flankers are more likely to be processed. When the target location is certain, the observer can narrow the spotlight so that the flankers have less effect (Kinclha, 1992). This effect is also sometimes described by the metaphor of a "zoom lens." In terms of interface design, there appears to be a penalty for not using a constant location for items of interest (e.g., changing the order of items on a menu in different contexts) in that the broader attentional spotlight produced by target uncertainty will ensure that irrelevant items are processed.

Mori and Hayashi (1995) found evidence for perceptual competition from adjacent windows in a computer display. When Mori and Hayashi had observers perform a visual search task in a main window, they found that increasing the number of peripheral windows increased the interference. When the search task target was nearer to the peripheral windows, performance degraded. Having overlapping windows also increased interference, as did dynamic peripheral windows. Thus, the perceptual competition described by Eriksen and Eriksen (1974) appears to play a part in multiwindow environments common in today's graphical user interfaces.

**Object-Based Proximity** In Eriksen and Eriksen's (1974) study, similarity effects were reduced when flanking letters were moved away from the central letter. We might expect that the observed effects of response conflict and redundancy gain would be amplified even further if the different sources of information represented different attributes of a single stimulus object at one spatial location. Indeed several studies have shown that this is the case. Many of these investigations have employed some variation of the **Stroop task** (Stroop, 1935), in which the subject is asked to report the color of a series of stimuli as rapidly as possible. In a typical control condition (e.g., Keele, 1972), the stimuli consist of colored symbols—for example, a row of four X's in the same color. In the critical **conflict** condition, the stimuli are color names that do not match the color of ink in which they are printed (e.g., the word **red** printed in blue ink). We can consider the word as an object having two attributes relevant to the task: its meaning and its ink color. The results are dramatic: Reporting ink color is slow and error prone relative to the control condition, as the semantic attribute of the stimulus (**red**) activates a response incompatible with information the subject must process (the color blue). The mouth cannot say the words **red** and **blue** at the same time, yet both are called for by different attributes of the single stimulus. Redundancy gain effects have also been observed when the color of the ink matches the semantic content of the word (e.g., Keele, 1972).

Similar examples of redundancy gain and response competition have been reported with various kinds of stimuli. Clark and Brownell (1975) observed that judgments of an arrow's direction (up or down) were influenced by the arrow's location within the display. "Up" judgments were made faster when the arrow was higher in the display. "Down" judgments were made faster when it was low. Similarly, Rogers (1979) found that the time it took to decide if a word was **left** or **right** was influenced by whether the word was to the right or left of the display, and Algom, Dekel, and Pansky (1996) found that the time to classify a number as large or small was affected by the size of the numeral used to portray it.
The Stroop effect suggests that multiple dimensions belonging to a single object are likely to be processed in parallel (Logan, 1980; MacLeod, 1991; Kahneman & Treisman, 1984), which will help performance if parallel processing is required but will disrupt performance if one dimension is irrelevant and to be ignored, particularly if it triggers an automatic and incompatible response. Since objects are more likely to define integral dimensions, this finding is consistent with results reviewed in Chapter 2. That is, integral dimensions produce a cost for a filtering task and a benefit when dimensions are redundant.

We have discussed an attentional spotlight that allows concurrent processing of elements lying close together in space (a space-based model of attention). In contrast, an object-based model proposes that concurrent processing occurs when elements lie within a single object, independent of its spatial dimensions. Indeed, several researchers have shown that judgments made about two parts of the same object are faster than judgments made about parts of different objects, even when the distance between parts is held constant (e.g., Behrmann, Zemel, & Mozer, 1998; Egly, Driver, & Rafal, 1994). Other studies have separated objects from their locations using motion (e.g., Kahneman, Treisman, & Gibbs, 1992). In these studies, subjects were shown a pair of simple geometric shapes (e.g., triangle and square) each with a letter inside (a cue). The letters disappeared, and the objects moved to new locations. One of the two letters was presented in either the appropriate or inappropriate object, and the subject’s task was to name this target letter. Subjects were faster when the target letter was consistent with the cue letter for that object than when it was not. Here, attention allocated to an object helped the subsequent perception of its properties, even though its location changed.

Object-oriented attention has also been shown in the Eriksen and Eriksen (1974) paradigm described above (identification of a target letter H is impaired when it is flanked with P’s). Kramer and Jacobson (1991) showed that the effect of flanking elements was enhanced when lines were drawn connecting the flankers to the target letter (i.e., creating a single object), and reduced when the lines connected the flankers to other display objects. Baylis and Driver (1992) showed that flanking letters matching the target letter in color had greater interference and facilitation effects than did flanking letters of a different color, even when the different-color flankers were closer to the target. When display elements are arranged to form part of an object, they are perceived and attended to differently than when they are not, and in consequence we must give serious consideration to how display elements are combined to form objects in display design. Consider, for example, how menu items for a web site are sometimes placed on various parts of an iconic object.

Applications of Object-Based Processing

It is not a simple thing to define what an object is. In cognitive psychology, an object is typically said to have three features: (1) surrounding contours or connectedness between parts, (2) rigidity of motion of the parts (relative to other elements in the scene), and (3) familiarity. None of these are truly defining features, but the more of these features a stimulus has, the more “object-like” it becomes, and the more it can benefit from object-based attention.

We shall discuss the benefit of objects in two contexts. The first concerns the mapping of display objects to real-world objects using conformal symbology. The second involves the construction of object displays in which multiple information sources are encoded as the stimulus dimensions of a single object.
Conformal Symbology and Augmented Reality  Earlier we mentioned a study by Wickens and Long (1995) that showed that the head-up display could improve control of aircraft position during landing. However, when the runway was in view, this result was true only when the HUD symbology was conformal: that is, the position of HUD objects corresponded to the position of objects they represented in the scene (e.g., the HUD runway superimposed on the physical runway as shown in Figure 3.9b). Some have referred to this conformal symbology as a form of augmented reality in that the reality of the far domain scene is augmented by computer generated imagery, projected on a near display (e.g., Drascic & Milgram, 1996). The Wickens and Long result is consistent with the object-based theories of attention discussed above: Having two components superimposed (the actual and HUD runways) to form one rigidly moving object using conformal symbology helped the pilot divide attention between the display and the world beyond, align the display object to the real object, and reduce tracking error (Martin-Emerson & Wickens, 1997).

Despite its utility in the aviation context, conformal imagery for automobile HUDs appears problematic. The number of objects a driver must see and keep track of is typically high, and the distances of the various objects from the driver varies considerably (Tufano, 1997). To present all this information at one distance (as HUDs typically do) may both clutter the display and distort the driver's perception of object distances in the scene. Nevertheless, some conformal imagery in vehicles has been considered to enhance visibility of the roadway ahead with synthetic imagery at night or in fog (Bossi, Ward, Parkes, & Howarth, 1997). Furthermore, the fusion of near and far, via augmented reality, has applications in many other domains.

Object Displays  Designers have also capitalized on the parallel processing of object features to create multidimensional object displays. In these displays, multiple information sources are encoded as the stimulus dimensions of a single object. Figure 3.10 illustrates four such examples. Figure 3.10a shows an attitude directional indicator, a two-dimensional object display used in aircraft control, which we discussed earlier in this chapter. The vertical location of the aircraft symbol relative to the horizon line indicates aircraft pitch (nose up or nose down), and the angle between the symbol and the rotating horizon represents the bank, or roll, of the aircraft. In addition to its objectness, this display is configured in a way that represents the aircraft, and is therefore familiar to the pilot. Figure 3.10b shows the safety parameter display for nuclear power reactor operators designed by Westinghouse, in which the values of eight key parameters are indicated by the length of imaginary "spokes" extending from the center of the display and connected by line segments to form a polygon (Woods, Wise, & Hanes, 1981 see Chapter 13). In addition to its objectness, a potential advantage of this display is that each type of system problem produces a unique shape or configuration of the polygon, as seen on the right of the figure, resulting in an emergent feature.

The display developed by Cole (1986) for medical applications, in Figure 3.10c, illustrates another example of an emergent feature. The rectangular display represents the oxygen exchange between patient and respirator. The width represents the rate of breathing, and the height represents the depth of breathing. Hence, the area of the rectangle—an emergent feature—signals the total amount of oxygen exchanged, a critical variable to be monitored. This correspondence holds true because oxygen amount = rate × depth, and rectangle area = width × height.
Figure 3.10  Four examples of object displays: (a) aircraft attitude display indicator, (b) safety parameter display, (c) medical display of oxygen exchange, (d) decision aid display.

One might ask, however, if the critical parameter is oxygen amount, why not display that quantity directly rather than have the operator infer it from the rectangle's area? The reason is that it is sometimes necessary to focus attention on one of the variables (rate and depth) contributing to the amount. This raises a question: Will the close proximity created by the object display disrupt the ability to focus attention on one of its dimensions, as when the operator must check one of the values being integrated? We address this question in the next section.

The Proximity Compatibility Principle

In the previous sections, we have discussed three ways in which multiple display channels can be integrated: through configuration to create emergent features, through spatial proximity, and through object integration. The issue of whether different tasks are served differently by more or less integrated displays is represented explicitly in the proximity compatibility principle (Barnett & Wickens, 1988; Wickens & Andre, 1990; Wickens & Carswell, 1995). To understand this principle, we must distinguish between display and processing proximity. Display proximity defines how close together two display components are. The distance between the components can be defined in spatial terms (i.e., the components are 1 cm apart) or in terms of object-based properties (the components are displayed as part of the same object, as in the display by Cole, 1986). In addition, Wickens and Carswell note that display proximity can be increased by other factors, such as using a common color or coding each variable using a common dimension; e.g., the two variables are both represented by lengths vs. one represented as a length and the other represented by an angle or digit. Processing proximity defines the extent to which two information sources are used within the same task (e.g., compared or integrated). A task with high processing proximity might be to estimate whether there is an increasing or decreasing trend in a scatterplot—many information sources (data points) must be considered. A task with low processing proximity might be to estimate the y-axis value of one data point in the scatterplot. The proximity compatibility principle can be summarized as follows:

If a task requires high processing proximity, there should be high display proximity.
If a task requires low processing proximity, there should be low display proximity.

Hence, to the extent that information sources must be integrated, there will be a benefit to presenting those sources either close together, in an objectlike format, or by configuring them to create emergent features. To the extent that information sources must be treated separately, the benefit of the high-proximity object display will be reduced, if not sometimes reversed. The advantage of object displays for information integration results from two factors. The first is that object dimensions can be processed in parallel (Treisman & Kahneman, 1984). Hence, the two sources, coded by the two dimensions, will more rapidly gain access to central processing. The second is that an object will be more likely to produce an emergent feature (e.g., area or shape) that can directly serve integration task requirements (Bennett & Flach, 1992). This is especially true when the dimensions are of the same type (e.g., all measures of extent, like the height and width of a rectangle) (Carswell, 1990). However, when a task requires focused attention on one dimension, the very
emergent feature that helped integration can make the focused attention task more difficult, just as closeness in space can create clutter which disrupts focusing.

The predictions of the proximity compatibility principle have been investigated and generally supported in many different contexts (e.g., Goettl, Wickens, & Kramer, 1991; Haskell & Wickens, 1993; Liu & Wickens, 1992, Mort & Hayashi, 1995; Vincow & Wickens, 1993). The situation illustrated in Figure 3.10d is representative. It shows an intelligent airborne decision aid that advises a pilot about whether to continue the current mission. Each recommendation is based on an information source (e.g., a weather advisory or an engine fuel check). The two dimensions of the display represent the two characteristics of each source: its reliability (how much it can be trusted) and its diagnosticity (its relevance to the decision at hand). As we will see in Chapter 8, these two features combine to indicate the total information worth of the source. Barnett and Wickens (1988) found that tasks requiring the integration of the two dimensions to evaluate total worth were served better by a rectangle display than by a bar graph display. As with Cole's (1986) experiment, the emergent feature of the area directly revealed the quantity (information worth) to be inferred. But when attention had to be focused on one variable to the exclusion of the other (e.g., the question "What was the reliability of weather information?"), the advantage for the rectangle over the bar graph disappeared. The proximity compatibility principle has also been successfully applied to the design of statistical graphs (e.g., Carswell & Wickens, 1988; Hollands & Spence, 1992), to be discussed further in Chapter 4.

Emergent features need not be created exclusively by single objects (Buttigieg & Sanderson, 1991; Sanderson, Flach, Buttigieg, & Casey, 1989). Indeed, we saw the emergent feature created by the vertical array of dials at the top of Figure 3.8. An array of parallel bar graphs representing, for example, four engine parameters, such as that in Figure 3.11, also creates an emergent feature—a horizontal line across the top—when all engines are running at the same level. In this case, display proximity is defined not by belonging to a common object but by the identical form of representation (vertical extent) of all the indicators. Displays that portray higher-order information using emergent features are referred to as configural displays. An alternative approach put forth by Elvers and Dolar (1995) proposes that separated displays (e.g., bar graphs) can be augmented to

![Figure 3.11](image_url)  
**Figure 3.11** Emergent features in engine parameter display.
show directly that information important for an integration task. In either case, it is clear that one can design spatially separated displays that are effective for integrated tasks as long as they use similar attributes to convey information (e.g., spatial extent) and the emergent features are constructed to support the task demands (Bennett & Flach, 1992).

Some researchers have advocated the use of *Chernoff face displays* as a way of integrating information for portraying complex multivariate data (e.g., Jacob, Egeth, & Bevan, 1976). These displays arrange simple geometric symbols in the shape of a face. The size or shape of each symbol varies with the magnitude of the variable it represents. The face may be considered an object in that it is both highly familiar and is enclosed by a single contour, the head. Chernoff face displays tend to be effective for integration tasks (since the expression and appearance of the face change considerably with changes in the data). Suzuki and Cavanagh (1993) had subjects perform integrated and focused search tasks on face displays and a random arrangement of the same features and found that subjects performed better with the face display on the integrated task, but they performed better with the random feature arrangement on the focused task. These results are consistent with the predictions of the proximity compatibility principle.

The identical color of two objects on a display, like the integrality of the dimensions of an object, also creates display proximity that serves processing proximity (Wickens & Andre, 1990). That is, two items on a cluttered display will be more easily integrated or compared if they share the same color (different from the clutter), but the shared identity of color may disrupt the ability to focus attention on one while ignoring the other. A unique color code helps this focusing process, just as it disrupts the integration process. This appears related to the effect of flanker color similarity in the Eriksen and Eriksen (1974) paradigm mentioned earlier (Kramer & Jacobson, 1991).

The proximity compatibility principle also applies to spatial distance in a cluttered display. Two pieces of information that need to be integrated on a cluttered display should be placed in close spatial proximity, as long as this proximity does not also move them too close to irrelevant clutter (Wickens & Andre, 1990). For example, Sweller, Chandler, Tierney, and Cooper (1990) found that visual materials lead to better learning if graphic material and related text (two information sources with close mental proximity because they need to be integrated) are adjacent to one another on the page. Bettman, Payne, and Staelin (1986) discuss the importance of spatial proximity between related items (costs and benefits of a product) in the design of product warning labels. Milroy and Poulton (1978) point to the importance of close proximity between graphed lines and their labels. That is, labels should be set next to the lines, not in a legend below. Weinstein, Gillingham, and Ercoine (1994) found that when a symbol showing vertical velocity was integrated into a circular altimeter, it produced better performance than other arrangements where vertical velocity information was further away from the altimeter. This result supports the proximity compatibility principle because a pilot must often integrate vertical velocity and altitude information.

With computers, an advantage of window-based systems is that they allow simultaneous and adjacent positioning of different information sources that need to be compared (e.g., a version of a document with and without critical comments). This is clearly superior to earlier screen arrangements where comparing two screens meant remembering the information on one while viewing the other. When the windows are placed
in adjacent positions, corresponding parts of the documents are located in similar locations and can therefore be compared more easily than otherwise, also helping to reduce attentional demands. By allowing the user to view both windows simultaneously, and by placing the windows adjacent to each other, we are allowing high display proximity when our task requires high processing proximity. The proximity compatibility principle can therefore be used to help reduce attentional demands when comparing information. As noted above, however, closely-spaced irrelevant windows can hurt the focus of attention (Mori & Hayashi, 1995).

An illustration of a violation of the principle is found in the design of the radar display on the USS Vincennes. An Iranian passenger airplane was mistaken for an attacking fighter plane by the Vincennes and inadvertently shot down in the Persian Gulf (U.S. Navy, 1988; Klein, 1996). In the radar display, the symbol signifying the location of the aircraft was in a separate location from information describing the vertical actions of the approaching aircraft (labeled the "range gate"). It is likely that the lack of close spatial proximity prevented the operators from integrating the two pieces of information correctly. Hence, the aircraft was classified as a descending, attacking fighter rather than a climbing, commercial air carrier.

In conclusion, moving multiple displayed elements close together, providing them with a common representation (e.g., color or format), or integrating them as dimensions of a single object) has the following effects:

1. This close proximity will increase the possibility of parallel processing by moving both dimensions into foveal vision. Parallel processing will be most likely to occur if they are integrated as dimensions of a single object.
2. Both close spatial proximity and objectness can create useful emergent features such as symmetry, shape or order if the display dimensions are the same (e.g., the length of two lines can create an emergent feature; the length of a line and its color cannot). These emergent features can help information integration if they are mapped into key variables of the task. (The mapping calls for creativity and ingenuity by the display designer.) The emergent features can hurt performance if they are not mapped into the task.
3. The close proximity, enhanced by objectness, can create unwanted clutter (sometimes in the form of emergent features) or response conflict. Both response conflict and emergent features will be troublesome to the extent that the task calls for focused attention on one of the variables combined in the display.

**Color Coding**

Discussions of target search and visual attention must include a brief treatment of the specific effects of color coding in displays, although color coding relates to a number of other topics in this book. For the 97 percent of the population who are not color-blind (roughly 7 percent of males cannot adequately discriminate certain wavelengths of light), differences in color are processed more or less automatically and in parallel with characteristics of shape and motion (Treisman, 1986). Although there are costs, several benefits result from the automatic characteristics of color processing.
1. Color stands out from a monochrome background. Therefore, color-coded targets are rapidly and easily noticed. As suggested in our discussions of visual search, search time for a uniquely color-coded target in a cluttered field is independent of the size of the field. Therefore, color coding of targets or critical elements in a display is quite effective for rapid localization (Christ, 1975). Color is effective as a means of highlighting an important item on a menu, for example (Fisher & Tan, 1989).

2. Certain colors have well established symbolic meaning within a population (e.g., in America, red signals danger, hostility, the order to stop; green signals go, safety), and therefore, color coding can capitalize on these population stereotypes, a concept to be discussed further in Chapter 9.

3. Color coding can tie together spatially separated display elements. As noted, this characteristic will be most useful if the commonly colored items also need to be integrated as part of the task (Wickens & Andre, 1990). Thus, for example, there will be an advantage to color coding different regions on a weather map according to temperature. Regions having similar temperatures can be perceptually grouped in parallel.

4. The automaticity with which color is coded enhances its value as a redundant coding device in combination with shape, size, or location. As noted in Chapter 2, the traffic light is an example of redundant coding of color with location. Both Kopala (1979) and Hughes and Creed (1994) found that redundant use of color improved search times in aircraft displays. Backs and Walrath (1992) found similar results, especially when the task involved identifying a set of multiple targets in the display. Backs and Walrath also found that fewer eye movements were necessary when redundant color coding, suggesting larger UPOVs in this situation. To realize the full benefits of redundant coding was used, however, it may be necessary to cue users as to the existence of the redundancy (Backs & Walrath, 1995).

Because of its aesthetic appeal, color coding has become prevalent in many displays. However, we note several subtle limitations that may be critical for system design.

1. Like other sensory continua, color is subject to the limits of absolute judgment (see Chapter 2). To guarantee that the value (and therefore meaning) of a color will not be misidentified (i.e., no errors), the system designer should use no more than five or six colors in a display (Carter & Cahill, 1979), although maximum information throughput occurs using about ten colors (Flavell & Heath, 1992). Furthermore, if colors are to be perceived under conditions of glare or changing or low illumination (e.g. the dashboard or cockpit), failures of absolute judgment will be even more prevalent, because color perception is affected by ambient light; for example, red may be confused for brown (Stokes, Wickens, & Kite, 1990).

2. Color does not naturally define an ordered continuum. If people are asked, for example, to place five colors in an order from “least” to “most,” there will be a great divergence of opinion about the appropriate ordering. Even the rainbow spectrum is far from universally recognized as a continuum. Since color or hue
ordering does not have a strong population stereotype, it is generally ineffective to use color coding to represent an ordered variable like speed or density. Instead quantitative variables should be coded using saturation or brightness rather than—or redundantly with—hue (Kaufmann & Glavin, 1990). For example, to code altitude on a map, a single hue (e.g., brown) is used, but greater levels of the variable being coded are shown in more saturated color (e.g., higher altitudes are shown in darker brown, or greater ocean depths are shown in a darker blue) (Tufte, 1990).

3. Population stereotypes can produce poor design if a color-coding scheme associates a color with a conflicting meaning. For example, suppose a temperature-coding scheme is designed in which green represents low temperature, but in the system very low temperatures signal an unsafe operating condition. Hence, the population stereotype of green with “safe” or “go” is not the one that should be inferred by the operator.

4. Given the automaticity with which it is processed, irrelevant color coding can be distracting. When different colors are used to highlight different areas or items, it is important that the distinction made by the colors is compatible with relevant cognitive distinctions that are intended to be interpreted by the viewer. This issue of display-cognitive compatibility was discussed in the context of spatial organization and will emerge again in Chapter 4 in discussions of display motion.

Given the reduced costs of color VDT displays, color has become a more viable option for the display designer. Although operators usually express a preference for color over monochrome displays, caution should be exercised before deciding on its implementation and careful consideration should be given to the limitations and constraints, as described.

**ATTENTION IN THE AUDITORY MODALITY**

The auditory modality is different from the visual in two respects relevant to attention. First, the auditory sense can take input from any direction, and thus, there is no analog to visual scanning as an index of selective attention. Second, most auditory input is transient. A word or tone is heard and then it ends, in contrast to most visual input, which tends to be more continuously available. For example, the printed word usually remains on the page. Hence, the preattentive characteristics of auditory processing—those required to “hold on” to a stimulus before it is gone—are more critical in audition than in vision. As discussed briefly in Chapter 1, short-term auditory store is longer than short-term visual store.

There is a long history of research in auditory selective attention, which will not be discussed here (see Moray, 1969; Wickens, 1984). Much of this research is based on the dichotic listening task, in which the listener hears two independent channels of sounds, words, or sentences, one in each ear. Usually the subject attends to only one channel and the other is ignored. Interest has focused on the physical and semantic characteristics of messages that lead to successes and failures in these divided and focused attention tasks (Broadbent, 1958; Cherry, 1953; Moray, 1969; Treisman, 1969).
Auditory Divided Attention

A general model of auditory attention (see Norman, 1968; Keele, 1973) proposes that an unattended channel of auditory input remains in preattentive short-term auditory store for 3-6 seconds (see Chapter 6). The contents of this store can be “examined” if a conscious switch of attention is made. Thus, if your attention wanders while someone is talking to you, it is possible to switch back and “hear” the last few words the person spoke, even if you were not attending to them when they were uttered.

Even in the absence of a conscious attention switch, information in unattended channels may make contact with long-term memory. That is, words in the unattended channel are not just meaningless “blobs” of sound, but their meaning is analyzed at a preattentive level. If the unattended material is sufficiently pertinent, it will often become the focus of attention (i.e., attention will be switched to the unattended channel). For example, a loud sound will almost always grab our attention, as it may signal a sudden environmental change that must be dealt with. Our own name also has a continued pertinence, and so we will sometimes shift attention to it when spoken, even if we are listening to another speaker (Moray, 1959; Wood & Cowan, 1995). So also does material semantically related to the topic that is the current focus of attention (Treisman, 1964a).

What of the fate of the words or sounds that never receive our attention—either because their pertinence is not high enough or because we do not voluntarily choose to listen to them? As Dr. Seuss says, “Oh, their future is dreary” (Seuss, 1971). There is little evidence that this material makes any impact on long-term memory, beyond the brief, transient activation of the semantic unit. Hence, the idea of learning without awareness, whether in one’s sleep or through techniques of “subliminal perception,” has received little empirical validation (Swets & Druckman, 1988).

Information presented in an unattended channel is temporarily inhibited for several seconds following presentation, demonstrating a phenomenon called negative priming (Banks, Roberts, & Ciranni, 1995). On some trials, Banks et al. presented information in the attended channel that had been presented to the unattended channel on the previous trial. When subjects shadowed this information, they were slower relative to a control condition in which the information was new, demonstrating negative priming. The negative priming was the same whether both presentations were to the same ear or different ears, suggesting that the inhibition of the unattended information adheres to the content and not the position in space. Negative priming has also been demonstrated in the visual modality (e.g., Fox, 1995).

In our discussion of visual attention, we saw that close proximity, particularly as defined by objectness, was a key to supporting the successful division of attention necessary in an information integration task. We also saw that the same manipulations of proximity that allowed success in divided attention were responsible for the failure of focused attention. These manipulations and observations have analogies in audition.

It is possible to think of an “auditory object” as a sound (or series of sounds) with several dimensions, which seem to enjoy the same benefits of parallel processing as do the dimensions of a visual object. For example, we can attend to both the words and melody of a song and to the meaning and voice inflections of a spoken sentence. Moore and Massaro (1973) found that subjects were able to judge the quality and pitch of a tone
simultaneously as well as either dimension could be judged alone. Auditory warning alerts have been designed to capitalize on our parallel processing ability using redundant dimensions like pitch, timbre, and interruption rate in various combinations (Edworthy & Loxley, 1990; Sorkin, 1987).

**Focusing Auditory Attention**

In vision, we saw that using close proximity to facilitate parallel processing was a double-edged sword because it disrupted the ability to focus attention. In the auditory modality, too, we find that focused attention on one channel is disrupted when two messages have similar spatial locations (e.g., Egan, Carterette, & Thwing, 1954; Treisman, 1964b). For example, in monaural listening, two messages are presented by headphones with equal relative intensity to both ears. This is similar to what you would experience when listening to two speakers both directly in front of you. In dichotic listening, the headphones deliver one message to the left ear, and the other to the right, and you hear one voice in each ear. Egan, Carterette, and Thwing (1954) found that there are large benefits of dichotic over monaural listening in terms of the operator’s ability to filter out the unwanted channel. However, we become less able to perform this selective filtering task as we age (Barr & Giambra, 1990).

We can also attend selectively to auditory messages even from similar locations. The *cocktail party effect* describes our ability to attend to one speaker at a noisy party and selectively filter out other conversations coming from similar spatial locations (with varying degrees of success). In this case, we must be able to use dimensions other than location to focus attention selectively. One such dimension for selection is defined by pitch. It is easier to attend to one voice in the presence of a second if the second voice is of the opposite sex (and thereby having different pitch) than if the two voices are of the same sex (Treisman, 1964b). Intensity may also serve as a dimension of selection. It is easy to attend to a loud message and tune out a soft one. Semantic properties can also serve as a cue for selection, so that it is easier to focus attention on one auditory message if a concurrent message has a very different semantic content (deals with a different topic) than if the content is similar (Treisman, 1964a).

By moving the eyes to a location, our visual system can selectively attend to the information at that location and ignore other information sources. Although the auditory modality does not have an “earball” that can rotate like an eyeball, it appears that auditory attention can be directed by cueing (e.g., Ward, 1994), just as visual attention can be directed without movement of the eye. Mondor and Zatorre (1995) found that auditory attention can be shifted to a specific location in response to an auditory spatial cue, and that the distance of the shift does not affect the time required for the shift of attention (i.e., the advantage of the cue was no greater when the target was at midline than when fully left or right). As with visual attention (when eye movement is not involved), it appears that the attentional spotlight is moved in discrete fashion rather than continuously as auditory attention is shifted from one spatial position to another.

To summarize, auditory messages differ from one another in terms of many dimensions such as pitch, location, loudness, and semantic content. The greater the
difference between two messages along a given dimension and the greater the number of dimensions of difference, the easier it will be to focus on one message and ignore the other. When a message is ignored, its perception is subsequently inhibited for a brief period. Finally, auditory attention can be shifted to a particular location using an auditory cue.

**Practical Implications**

The characteristics of auditory attention have practical implications for system design, some of which have already been discussed. For example, we noted the concept of an auditory object and that system designers can capitalize on the parallel processing of several dimensions of an object to provide more redundancy or information in a given auditory alert. This will be discussed further when we consider detection of failures in process control (Chapter 13).

The auditory display designer wants to know what features of an alert will grab attention, so that it will be processed (Sorkin, 1987). As described in Chapters 12 and 13, although loud tones call attention to themselves, they can annoy and startle, and their intensity can increase stress, leading to poor information processing. Designers might capitalize on the operator’s tendency to switch attention to contextually pertinent material (that is not necessarily loud) to design less noxious alerts. If a pilot is landing an airplane, for example, it may not be necessary to have loud alerts for those operations relevant to landing. However, loud alerts may still be necessary to indicate other changes in the status of the airplane (e.g., a drop in pressure in the passenger cabin). Since one has a low attentional threshold for one’s own name, personalized alerts prefixed with the operator’s name may also attract attention without high volume. These attention-grabbing but quieter auditory warnings have been called *attensors* (Hawkins & Orlady, 1993).

As noted above, the auditory modality does not have a directional “earball.” Hence, greater concern must be given to determining those auditory display features that allow different auditory channels to be distinguished and discriminated. For example, how can the automobile designer ensure that an auditory warning will not be confused with a radio channel, engine noise, or ongoing conversation? The spatial dimension can be employed to some degree. An experiment by Darwin, Turvey, and Crowder (1972) suggests that three “spatial” channels may be processed without distraction if one is presented to each ear and a third is presented with equal intensity to both ears, thereby appearing to originate from the midplane of the head.

In this manner, airplane pilots might have available three distinct audio channels—for example, one for messages from the copilot, one for messages from air traffic control, and a third for messages from other aircraft or for synthesized voice warnings from their own aircraft. They could not process the three in parallel since all would call for common semantic analysis, which we saw was impossible, but they could at least focus on one with less intrusion from the others. The definition of channels in terms of the pitch dimension suggests that additional separation might be obtained by distinguishing the three spatial channels redundantly, through variation in pitch quality. Thus, the center message that is most likely to be confused with the other
two could be presented at a substantially different pitch (or with a different speaker's voice) than the others.

The fact that there is no direction-sensitive auditory "earball" has its advantages as well. It can cue the user to locations in space in the full 360° volume of space. Thus, one can use spatial audio to assist a pilot (or potentially, a car driver) to identify targets of interest in the environment (e.g., Begault, 1993; Bronkhorst, Veltman, & van Breda, 1996). Begault (1993) was interested in whether 3D audio information presented redundantly with a visual target would help a pilot locate a target. The 3D audio system (both ears) that auditorily presented the location of the target shortened acquisition times to capture the target relative to an audio signal presented to one ear only. Begault and Pittman (1996) found that 3D audio alone produced shorter target acquisition times than a combined visual-auditory display in which a warning was presented auditorially and the location was presented visually. It is also important to use broadband signals (signals that have a range of different frequencies) in order for pilots to accurately localize signals (King & Oldfield, 1997). This would seem even more necessary if pilots are to distinguish different pitch qualities.

**Cross-Modality Attention**

The discussion up to this point has focused exclusively on attention within a modality. But we are often confronted with parallel inputs across modalities. Consider our van driver in Chapter 1, who needed to attend to visual information (the map, the stalled car ahead), and auditory (the truck horn) simultaneously. Consider also other everyday situations, as when we drive and our passenger gives us verbal directions, or when the pilot landing an aircraft monitors the visual environment while listening to the copilot's spoken messages regarding key velocities. Advances in multimedia technology make it possible to view text or pictures and hear audio information simultaneously when we visit a web site. The construction of virtual environments, to be discussed in Chapter 5, also requires the proper integration of visual and auditory information.

There are advantages to using multiple modalities. Redundantly coding a target across modalities (e.g., a visual warning is coupled with an auditory beep) speeds processing (Miller, 1991). We saw this in the vigilance situation in Chapter 2. In addition, Miller also found redundancy gain effects depended on how levels of stimuli were paired. High spatial locations were responded to more quickly when they were paired with high-frequency tones than with low, and the opposite was true for low spatial locations, an example of configurable stimulus dimensions as described in Chapter 2, and also similar to the Stroop effects described in this chapter. In Chapter 11, we will discuss some experiments suggesting that dividing attention between modalities may be better than dividing attention within a modality.

It is commonly found that when input from vision and other modalities is put in conflict, the phenomenon of **visual dominance** results. Examples of visual dominance over auditory or proprioceptive modalities are abundant. For example, Colovita (1971) required subjects to respond as fast as possible to either a light (with one hand) or a tone (with the other hand). On infrequent occasions, both stimuli were presented simultaneously. When this occurred, subjects responded to the light and did not notice the tone. Jordan (1972) found that reaction time to a compound stimulus consisting of
a light and displacement of a limb was slower than reaction time to the proprioceptive stimulus alone. This result suggests that the light captured attention and slowed down processing of the proprioceptive information. Different examples of visual dominance are observed when vision and proprioception are placed in conflict through prismatically distorted lenses (Rock, 1975). Behavior in these situations suggests that the subject responds to the visual information and disregards that provided by other modalities.

Some time-sharing situations described in Chapter 11 also show a form of visual dominance when auditory and visual tasks are performed concurrently. In these circumstances, the auditory task tends to be hurt more by the division of attention than the visual task (e.g., Massaro & Warner, 1977).

There are circumstances in which visual dominance can lead to nonadaptive behavior. Illusions of movement provide an example. When the visual system gives ambiguous cues concerning the state of motion, the correct information provided by proprioceptive, vestibular, or “seat of the pants” cues is often misinterpreted and distorted. For example, while sitting in a car at an intersection with another car beside, passengers may experience the illusion that their car is moving backward, while in fact their vehicle is stationary and the adjacent car is moving forward. The passengers have discounted the proprioceptive evidence from the seat of the pants that no inertial forces are operating.

Visual dominance can be moderated in some cases. Ward (1994) measured response times for visual or auditory targets (an “x” appearing on the left or right of fixation, or a sound occurring to the left or right, respectively), with no cue, a visual cue, an auditory cue, or both types of cues. He found that when the visual cue conflicted with the auditory cue, visual cues dominated if the target was visual, but auditory cues dominated if the target was auditory. Heller (1992) found that visual dominance was eliminated and haptic dominance was shown when observers identified letter shapes by haptic exploration and vision simultaneously. It may be that in most studies of visual dominance, the fundamental task is visual, but visual dominance is not universal across tasks.

The phenomenon of visual dominance appears to oppose our natural tendency to switch attention to stimuli in the auditory and tactile modalities. These stimuli are intrusive, and the peripheral receptors have no natural way to shut out auditory or tactile information. We cannot close our “earlids,” nor can we move our earball away. As a consequence, auditory devices are generally preferred to visual signals as warnings (Simpson & Williams, 1980; Sorkin, 1987). The truck horn effectively warned our van driver of the truck’s presence in the Chapter 1 vignette, perhaps saving the driver’s life.

In summary, when an abrupt auditory stimulus intrudes on a background of ongoing visual activity, it will probably call attention to itself and alert the operator. However, if visual stimuli are appearing at the same frequency and providing information of the same general type or importance as auditory or proprioceptive stimuli, biases toward the visual source at the expense of the other two is likely if the task is visual in nature.

**TRANSITION**

In this chapter, we have described attention as a filter to the environment. Sometimes the filter narrows to decrease irrelevant visual or auditory input, and sometimes the filter
broadens to take in parallel streams of environmental information. The effective breadth of the filter is dictated by the limits of our senses (e.g., foveal vision), the differences and similarities between stimulus channels, and the strategies and understanding of the human operator. What happens, then, when material passes through the filter of attention? We saw in Chapter 2 that material may be provided with a simple yes-no classification (signal detection) or categorized into a level on a continuum (absolute judgment). More often, however, the material is given a more sophisticated and complex interpretation. This interpretation is the subject of several subsequent chapters.

It is convenient to distinguish between two kinds of perceptual interpretation. The first are analog-spatial interpretations, whose relevance is defined by continuous spatial dimensions. Judgments about how far away things are, where they are, how big they are, and how they are oriented involve this type of interpretation. The judgments that a driver makes about the state of a vehicle, or that a pilot makes about the state of an aircraft, are of this form. So also is the reading of a dial, a graph, or the mercury level in a thermometer. The interpretation is directly analogous to the physical form. The second class of interpretations consists of those that are verbal and symbolic. The meaning of these stimuli is not directly embodied in their physical form (location, shape, or orientation). Rather, this meaning is interpreted by decoding some symbolic representation, a written or spoken word or alphanumeric story or geometric symbol. Hence, this form of perception is heavily language based.

Attention to these stimulus sources, whether analog or symbolic, is necessary but not sufficient to properly interpret the state of the world on which to base future actions. In Chapters 4 and 5, we will discuss the perception and interpretation of analog material; in Chapter 6, we will discuss symbolic verbal material. We will revisit the concept of attention in Chapter 11, in the context of dividing attention among tasks rather than perceptual channels.

REFERENCES


international symposium on aviation psychology (pp. 415–421). Columbus, OH: Dept. of Aviation, Ohio State University.


