Attention Control

Introduction

Many real-world tasks—driving, flying, process management—require an operator to monitor multiple information sources over an extended period of time. In such cases, as is discussed in chapter 4, the operator typically learns to allocate attention in a way that is adaptive, shifting attention between various channels with a frequency determined by their relative importance and bandwidth. Individual shifts of attention, however, occur over the course of mere seconds or less. Moreover, even when the operator’s attentional scanning strategies are well tuned over the long run, it may still be useful to alert the operator to unexpected, infrequent, or high-priority events, interrupting the normal path of attentional scanning to ensure that important information is quickly encoded. Even an experienced and attentive driver, for example, might benefit from an alert that announces when the fuel gauge is approaching empty, from an alarm that signals an impending side collision, or from the bright reflective clothes of a nighttime jogger.

A designer may sometimes also anticipate users who bring little if any top-down knowledge-driven guidance to bear on their interactions with a product or system. Effective design in such instances requires cues to guide attention in a bottom-up manner toward crucial information. When designing instructional graphics or animations, for example, an educator may need to cue attention to important details that students, unfamiliar with the material, might otherwise overlook (Lowe 2003).

In such cases, the system designer is faced with a challenge of attentional control—of catching the operator’s attention and orienting it toward a timely piece of task-relevant information. Although the problem of cuing attention to an information channel seems straightforward, the designer may face difficulty in creating a cue that is noticeable but not disruptive. A visual cue that is readily detectable in a sparse or static display, for example, may be lost in a cluttered bank of blinking lights, spinning dials, and flickering monitors, particularly if it is visible only in the visual periphery or if the operator is engaged in another task (Nikolic, Orr, and Sarter 2004). Conversely, a cue that is conspicuous or persistent enough to guarantee detection in a busy environment may be overly distracting and subjectively annoying (Bartram, Ware, and Calvert 2003). The present chapter discusses several basic issues of attention capture and capture failures before turning to a discussion of
the most important applications of attention control: the role of automation in directing attention to critical events via alarms and alerts.

**Inattentio... and Change Blindness**

The issue of attentional control can be studied from the perspective of what captures attention. However, the issue is also relevant in cases where environmental events fail to draw attention, as in the phenomena of inattentional blindness (Carpenter 2002; Mack and Rock 1998) and change blindness (Rensink 2002; Simons and Levin 1997). Inattentional blindness, sometimes described as the *looked-but-failed-to-see* effect (Herslund and Jørgensen 2003), occurs when a lapse of attention causes an observer to overlook an object that is prominent in the visual field and is well above sensory threshold. Mack and Rock (1998), for instance, found that when subjects focused their attention intently on a shape in their central vision, they often failed to detect an unexpected probe object that appeared at the same time in the visual periphery. More surprisingly, if attention was directed to the visual periphery, many subjects failed to notice the unexpected probe even when it was presented in their central vision directly where their eyes were fixated. These effects were not simply the result of poor stimulus visibility; it is important to note, as the probes were easily detected when subjects expected them. Failures to notice the probes were attentional, not sensory. Looked-but-failed-to-see errors are easy to produce and sometimes startling in their strength (e.g., Simons and Chabris 1999) and have been implicated as a common cause of traffic accidents (Herslund and Jørgensen 2003).

Change blindness, another form of visual lapse, occurs when an observer fails to detect an event (e.g., discrete change) in the environment around him. Thus, whereas inattentional blindness is a failure to notice something here and now, change blindness is a failure to notice that something is different from what it was. For an online chat box user sitting back down at his computer, this could be a failure to notice that a new message has arrived; for a driver turning her eyes back to the road after a glance at the radio dial, it might be a failure to notice that the vehicle ahead now has its brake lights on (Pringle et al. 2001); for a pilot contending with the demands of a busy cockpit, it may be in a failure to notice the onset of a green light in the cockpit display that indicates a change in flight mode (Sarter, Mumaw, and Wickens 2007). In a face-to-face conversation, more remarkably, it can even mean that a speaker fails to notice that the person to whom he was speaking has disappeared and been replaced by someone new. In a study by Simons and Levin (1998), an experimenter approached pedestrians on a college campus to ask for directions. As the experimenter and a pedestrian talked, their conversation was interrupted by a pair of workers (confederates of the experimenter, in actuality) who barged between them carrying a wooden door. As the workers tromped through, the original experimenter snuck away behind the door, and a new experimenter was left standing in his place. In most
cases, the conversation carried on normally, with the pedestrian failing to notice that he was speaking to someone new.

How do such lapses occur? Very often, CB is a failure of attention. Data indicate that changes are likely to go unnoticed if they are not attended when they occur. Under most circumstances, a change in the visual environment produces a transient, a brief flutter of motion or flicker. These signals tend to attract attention and help ensure that the change is noticed. However, the change will often go undetected if it occurs while the observer is looking away, or if its accompanying transient happens to be hidden by an occluding object (Simons and Levin 1998), the flicker of a display (Rensink, O’Regan, and Clark 1997), egomotion (Wallis and Bulthoff 2000), other transients produced by concurrent events (O’Regan, Rensink, and Clark 1999), or even just an eye movement (Grimes 1996) or blink (O’Regan et al. 2000). In the terminology of Rensink (2002), a dynamic change, or a change in progress, is generally easy to notice, but a completed change, one that occurred while attention was elsewhere or vision is obscured, is not.

It is easy to envision real-world circumstances in which the difficulty of detecting completed changes may hinder task performance. Consider an operator monitoring a large dynamic display of air traffic. At any given time, most of the display will fall outside the momentary focus of attention. As a result, a large number of changes will be likely to escape notice. Indeed, in one simulation pilots monitoring such a traffic display detected fewer than half of such changes, including those such as changes in course and altitude of another aircraft with important implications for safety (Muthard and Wickens 2002). When display monitoring was coupled with an attention-demanding flight control task, change detection rates dropped to well under 25 percent (Steltzer and Wickens 2006).

Interestingly, recent work has indicated that despite their remarkable blindness to completed changes, people incorrectly tend to assume that such changes will be noticed easily, a phenomenon that has been termed change blindness (Levin et al. 2000). Thus, not only are operators performing a complex task likely to overlook much of the activity going on around them, but they are also likely to markedly overestimate the degree to which their knowledge of the world is current, a metacognitive error that may discourage them from seeking new information when appropriate. Theorists argue that the first step toward good situation awareness is to notice objects and events in the surrounding environment (Endsley 1995). The phenomena of IB and CB make clear that attention plays an indispensable role in this process (Wickens and Rose 2001).

**Covert versus Overt Attention**

How is visual attention allocated? Helmholtz (1867, 1962), describing an experiment in which he tested his ability to see images briefly illuminated by a spark of light, noted that the direction of visual attention was not strictly
locked to the direction of the eyes. Rather, he observed, it was possible to "keep our attention ... turned to any particular portion we please" of the visual field, "so as then, when the spark comes, to receive an impression only from such parts of the picture as lie in this region" (1925, p. 455) even while the eyes were fixed elsewhere. James (1890, p. 434) likewise noted a difference between "the accommodation or adjustment of the sensory organs" and "the anticipatory preparation of from within of the ideational centres concerned with the object to which attention is paid." Following Posner (1980), this difference is now described as the distinction between overt and covert visual orienting. An overt attention shift is one that involves a movement of the eyes, head, or body to bring the sensory surface into better alignment with the object of interest. In vision, this is most often a saccade: an abrupt, rapid movement with which the eyes dart from one point of fixation to another. A covert attention shift, conversely, is a change of mental focus in the absence of any physical movements. Thus, a computer user may orient overtly by moving her eyes to look directly at a text message that has just popped up on her screen, or she may notice the message covertly in her peripheral vision without turning to look directly at it. Likewise, a person in a noisy room may overtly turn his head and lean forward to listen more carefully to what a friend is saying, but may covertly listen more intently with no outer sign of effort while trying to eavesdrop on someone else's conversation.

To separate the effects of attentional processes from the effects of visual acuity, basic attention researchers often ask their subjects to make judgments of objects in the retinal periphery while holding the eyes fixed on a central point. In most real-world tasks, however, eye movements are common. In viewing a naturalistic scene, observers generally make two to three eye movements per second, pausing in between for fixations of 250–300 ms (Rayner 1998). Indeed, Moray and Rotenberg (1989, p. 1320) speculated that covert attention shifts are of little consequence in most applied contexts—that "the more a task is like a real industrial task, where the operators can move their head, eyes and bodies, can slouch or sit up, and can, in general, relate physically to the task in any way they please, the more 'attention' is limited by 'coarse' mechanisms such as eye ... movements." Nonetheless, information gathered from the study of covert attention is valuable to the applied attention researcher. Research has found that in vision, overt and covert movements are subserved by many common neural mechanisms (Corbetta 1998) and are functionally linked, though in an asymmetrical manner; it is possible to move attention covertly without also moving the eyes, but a saccade cannot be executed until covert attention has shifted to the target location (Deubel and Schneider 1996; Hoffman and

* Saccades themselves last only a few tens of milliseconds, and visual input is processed only during the intervening fixations. Visual processing is suspended while the eyes are in flight, leaving the viewer effectively blind for a fraction of a second. This effect, known as saccadic suppression, ensures that the viewer does not experience a smeared retinal image when a saccade occurs.
Subramaniam 1995; Kowler et al. 1995). Stimulus attributes that attract covert visual attention therefore tend to have similar, though less powerful, effects on saccades (Irwin et al. 2000; Theeuwes et al. 1998). Understanding the workings of covert attention tells us much about overt attention.

**The Spotlight of Visual Attention**

Covert visual attention is often described as analogous to a spotlight (Cave and Bichot 1999), enabling selective processing of the "illuminated" region of the visual field. While the spotlight model—and related conceptualizations of attention as a mental zoom lens (Eriksen and St. James 1986) or gradient of resources (Downing and Pinker 1985; Laberge and Brown 1989)—carries many implications that are at best unconfirmed (Cave and Bichot 1999) and in some cases have been disproven (e.g., Duncan 1984; Kramer and Hahn 1995), it remains popular and captures some aspects of selective visual processing.

Evidence for the spotlight model originated in the work of C. W. Eriksen and colleagues (Eriksen and Hoffman, 1972, 1973) using what has become known as the flanker paradigm. In a typical flanker experiment, the subject’s job is to identify a target letter appearing at a particular location. Commonly, two or more target letters are linked to the same response. For example, the subject may be asked to press one button if the target is an H or an M and to press a different button if the target is an A or a U. The target is surrounded on either side by task-irrelevant flanker letters that are mapped to the same or a different response as the target. Subjects are asked to ignore the flankers and to respond solely on the identity of the target. Despite these instructions, the classical finding in the flanker paradigm is that response times (RTs) are influenced by the identity of the flanking stimuli, being longer when the flankers are response incompatible with the target than when they are response compatible. These effects, moreover, are strongest when the separation between the target flankers is small and often disappear when the separation between target and flankers exceeds approximately one degree of visual angle (Eriksen and Hoffman 1973; Yantis and Johnston 1990). Thus, flankers very near the target item appear to be processed to the point of recognition even when the subject wishes to ignore them, a failure of focused attention selectivity. Eriksen and Hoffman (1973) took their findings as evidence of a spatial attentional focus with a minimum size of about one degree. As is discussed more fully in chapter 6, this conclusion bears an important implication for display design: Displays and maps designed with items too closely together—even if they do not actually overlap—will hinder the focus of attention on any given item.

**Posner’s Cuing Paradigm**

In the real world, cues to shift attention may sometimes be inappropriate or just plain wrong. This issue was captured in an experimental procedure
known as the cuing paradigm, developed by Posner and colleagues (Posner, 1980; Posner, Snyder, and Davidson, 1980). In a typical version of Posner's cuing change to procedure the observer is asked to keep her eyes on a central fixation mark and to make a speeded detection judgment of a target signal that can appear in the visual periphery on either side. The observer's attention is manipulated by a cue that appears prior to signal onset. The columns of Figure 3.1 illustrate the events within two different types of experimental trials. Here, the cue is an arrow pointing toward one of the possible target locations. On valid cue trials, as illustrated in the left column, the target signal appears at the cued location. On invalid cue trials, as illustrated in the right column, the target signal appears at an uncued location. Generally, cue validity is above chance, such that target is more likely to appear at the cued than at the uncued location. The effects of attention are measured by comparing RTs for target detection following cues of different validity; an attentional cost obtains when RTs for invalid cue trials are longer than for control trials on which no location is cued, and an attention benefit obtains when RTs for valid cue trials are shorter than for control trials. Studies have confirmed that even for a task as simple as detecting the onset of an above-threshold spot of light, valid cues reduce RTs and invalid cues increase them (e.g., Posner, Snyder, and Davidson 1980). Similar effects obtain when a discrimination task is used instead of a detection task (e.g., ibid.), and when judgment accuracy is used as the dependent variable (e.g., Lu and Dosher 1998).
Posner, Snyder, and Davidson (1980) suggested that cues improve signal detection and discrimination by allowing subjects to align the mental spotlight of attention with the expected target location. What effect, more exactly, does this attentional orienting have? Although cuing costs and benefits are observed in simple uncluttered displays (e.g., Carrasco, Penpeci-Talgar, and Eckstein 2000), suggesting that cues can enhance the perceptual signal strength, data show that the benefits of attentional cuing are substantially larger when the target object is surrounded by easily confusable distractor objects (Shiu and Pashler 1994) or embedded in visual noise (Lu and Dosher 1998) than when it appears by itself against an open background. In real-world environments, these costs of invalid cueing can be quite large, in the order of several seconds (Yeh, Merlo, Wickens, and Brandenburg 2003). These results suggest that one effect of cuing is noise exclusion: the ability to ignore distracting or target degrading information.

Central versus Peripheral Cuing

As might be expected, characteristics of the cue stimulus itself also influence attentional performance. This was first demonstrated by Jonides (1981), who compared the effects of central and peripheral cues. A central, or endogenous, cue denotes the cued location symbolically but does not actually appear at that location. Typically, it is presented at a fixation point in the center of the visual field, though it need not always do so. A central cue may be an arrow, as in Figure 3.1, a word, or any other kind of symbol, so long as it does not directly mark the location to which attention should be shifted. A peripheral, or exogenous, cue, conversely, is a transient signal—usually a luminance change or burst of motion—that appears at the location to which attention is to be shifted. In Figure 3.1, for example, attention could have been exogenously cued to the target location by blinking or briefly brightening the box at the to-be-attended location. Jonides found that such peripheral cues, unlike central cues, tend to draw attention even if they do not predict the target location with accuracy better than chance, and tend to be effective even under relatively high levels of cognitive workload. For these reasons, peripheral cues are said to elicit reflexive attention shifts, whereas central cues are said to require voluntary orienting.

Other work in this area has found that reflexive attentional shifts are faster than voluntary shifts (Müller and Rabbitt 1989) but that the attentional benefits of reflexive orienting fade quickly if they are not maintained by voluntary processes. As such, reflexive visual attention shifts are sometimes described as transient and voluntary shifts as sustained (Nakayama and Mackeben 1989). As a real-world analog, we might consider a head-up display (HUD) that cues the driver’s attention to hazards in the roadway. Such a cue could be either a salient light on the windshield along the line of sight to the hazard (corresponding to a peripheral cue) or an arrow in the center of the driver’s forward view, pointing toward the hazard (central cue). Consistent with the
data from basic research, Yeh, Wickens, and Brandenburg (2003) obtained results suggesting that peripheral cues are more effective at least so long as they are reliable and do not obscure the actual target; these issues are addressed at the end of this chapter.

As the previous example suggests, the distinction between exogenous and endogenous orienting can help display designers know how to create signals or cues that most effectively guide attention. In light of this, researchers have invested great effort to determine precisely what stimulus properties are capable of eliciting reflexive attention shifts. Recent findings have in fact begun to blur the line between voluntary and reflexive processes. Contrary to Jonides's (1981) early findings, a number of researchers have documented instances in which symbolic cues that are highly overlearned or otherwise very easy to interpret seem to elicit reflexive attentional movements. Familiar words such as left and right, for instance, or social cues such as a face in which the eyes are looking in a particular direction, may produce reflexive attentional orienting even if they are not predictive of the likely target location (e.g., Gibson and Kingston 2006).

Nonetheless, some conclusions about the relative power of various cues to grab attention—of interest both to researchers and designer—can be drawn. At least three factors help determine the likelihood with which a visual signal captures and holds attention. The first is stimulus salience, where salience is a signal-to-noise measure of the feature contrast (e.g., color, motion, luminance contrast) between the target and the surrounding stimuli relative to the feature variability among the surrounding (Itti and Koch 2000; Nothdurft 1992; Theeuwes 1994). Motion and color contrast, which provide strong cues for object segmentation in the natural visual environment (Gibson, 1979; Regan, 2000), may be particularly salient.

Thus, for instance, a tilted gauge among a display of vertical gauges can attract attention readily, whereas a tilted gauge within a display of randomly oriented gauges may not. These issues are discussed more in the next two chapters.

The second factor that improves attention capture is stimulus newness. In Jonides's (1981) original study of endogenous and exogenous orienting, as noted, peripheral cues were arrows that flashed on-screen briefly at the cued location. Evidence has since suggested that such abrupt onsets are especially potent in capturing visual attention. More specifically, studies by Yantis and colleagues (Jonides and Yantis 1988; Yantis and Hillstrom 1994; Yantis and Jonides 1984, 1990) produced evidence that the appearance of something new within a scene or display triggers an attention shift. Under this hypothesis, it is the status of a stimulus as something new that captures attention rather than simply a luminance change or other visual transient. Consistent with this possibility, data indicate that a luminance change by itself is relatively less likely to capture attention if it does not signal the appearance of a new object (Irwin et al. 2000; Jonides and Yantis 1988). Echoing these findings from the basic research lab, a study of battlefield display monitoring
found that completed changes indicating the appearance of an enemy unit were reported at an 84 percent rate, whereas those involving the removal of a unit were only detected 44 percent of the time (Wickens, Thomas, and Young 2000).

The third factor modulating the likelihood of capture is the observer's attentional set. Any visual signal, no matter how salient, is more likely to attract and hold attention if the observer is searching for it or is otherwise prepared to let it grab attention (Folk, Remington, and Johnston 1992; Most and Astur 2007). The ability of a signal to grab and hold attention is therefore strongly modulated by its validity—the degree to which it predicts the occurrence and location of an impending target stimulus (Folk, Remington, and Johnston 1992; Yantis and Egret 1999). The importance of attentional set in applied domains is obvious in the next section, which discusses the effects of alert and alarm reliability.

**Auditory and Cross-Modal Cuing**

Although attentional control has been studied most intensely in the visual modality, researchers have adapted Posner's (1990) cuing task to explore auditory and tactile attentional processes as well. In a series of experiments by Spence and Driver (1994), for example, subjects were asked to make speeded judgments of auditory signals that could originate from either the left or right of the subjects' midline. The target signal on each trial was preceded by an auditory cue that came from either the same or the opposite side at varying intervals prior to target onset. In some experiments the cue was unpredictable of the target stimulus location (i.e., the target was equally likely to appear on the cued and uncued sides), and in other cases it was predictive with 75 percent validity. Consistent with data from studies of visual attention, judgments were faster when the target stimulus originated on the cued side than when it came from the uncued side. The effects of auditory cuing were qualified in two important ways, however. First, cuing benefits were larger and longer lasting when auditory cues were predictive of the target than when they were unpredictable. Thus, as in vision, exogenous attentional processes were most effective when aided by endogenous processes. Second, cues speeded auditory discrimination judgments (e.g., Was the pitch high or low?) but had no effect on the speed of auditory detection. The reason for this, the authors speculated, was that target detection did not require focused auditory attention information. This finding points to the particular value of auditory stimuli as alerts or alarms in applied settings.

An additional characteristic that makes auditory stimuli useful as alarms is that they summon not only auditory attention but also visual attention. Indeed, data suggest that visual, auditory, and even tactile attention are all spatially linked (Driver and Spence 2004; Spence, McDonald, and Driver 2004), such that attentional cuing in one perceptual modality can produce RT benefits for targets in a different modality. A cue that directs auditory
attention to one side of the observer’s midline, for example, will speed judgments of visual targets on that side as well; conversely, a cue that directs visual attention to one side of the visual field will also facilitate judgments of auditory stimuli from that side (Spence and Driver 1996). Auditory and visual attentional also appear to travel together in tasks that required sustained monitoring of information channels. For example, a study by Spence and Read (2003) asked subjects to monitor a stream of speech while they performed a simulated driving task and found that performance was better when the auditory stream came from a speaker directly in front of the driver, near the focus of visual attention, than when it came from a speaker to the driver’s side. Spatial links are also seen between visual and tactile attention (e.g., cuing visual attention to a position near one of the hands facilitates judgments of vibrotactile targets with the same hand) and between auditory and tactile attention (Lloyd, Merat, McGlone, and Spence 2003; Sarter 2007; Spence, Pavani, and Driver 2000).

Interestingly, cross-modal cuing effects do not always require that display channels in different modalities be physically coincident. Rather, it is sufficient that there be a mental correspondence between locations. For example, an auditory or tactile cue that comes from behind a driver facilitates the detection of an impending rear collision more than one that originates from the front—even if the only way for the driver to detect the collision is by checking the rearview mirror in the forward field of view. This is true even if the attentional cues are purely exogenous. In other words, drivers’ reflexive attention-shifting processes recognize the well-learned spatial correspondence between the location of the nonvisual cue behind them and the visual information depicted in the rearview mirror (Ho and Spence 2005; Ho, Tan, and Spence 2005).

Two additional characteristics of the spatial links among auditory, visual, and tactile attention are also noteworthy to the engineering psychologist (Spence and also Driver 1997). First, these links are not simply the results of the observer’s task strategy but are obligatory. Thus, even when it might be beneficial to divide auditory, visual, or tactile attention between different locations—either to monitor spatially separated streams of information (Driver and Spence 1994) or to filter away an irrelevant stream in one modality—people find it difficult do so (Spence and Walton 2005; Spence, Ranson, and Driver 2000). Second, though cross-modal interactions are seen even with purely covert attention shifts (Driver and Spence 2004), they also affect eye movements. For example, observers are capable of targeting saccades toward a sound source quite accurately (Frens and Van Opstal 1995) and are speeded in executing a saccade when a sound or vibrotactile stimulus coincides with the visual saccade target (Amlôt, Walker, Spence, and Driver 2003).

The implications of these facts for display design are straightforward. When displays in multiple perceptual modalities present information that is to be compared or integrated—for example, when a speech stream is used
to provide directions to a driver whose vehicle is displayed on a dashboard-mounted electronic map, or when an auditory alert is used to cue a pilot's visual attention to a display indicating a system error—the display channels in the various modalities should be arranged near one another. Conversely, when an operator is expected to attend to a single channel while filtering away distracting information in another channel, the to-be-attended channel should be spatially separated from the distractor channel even if the two channels are in different perceptual modalities. Further implications of attentional control for display are discussed in the next section.

**Applied Implications of Attention Capture Research:**

**Alarms and Attention Guidance**

Designers have used what is known about attention control to create displays that capture and then direct attention to important events and information channels in real-world systems. Of course, advertisers have long known about many of these factors, taking advantage of them to draw attention toward products on shelves, in the Yellow Pages, and as pop-up items on Web pages. The following sections discuss two sorts of applications: spatial cuing systems and alarms and alerts (Pritchett 2001; Stanton 1994; Wogalter and Laughery 2006). Both have in common the goal of capturing attention, often in a busy multitask environment, though they differ in that alarms and alerts tend to direct attention implicitly (e.g., the fire alarm will lead a person to attend to sounds or smells in the hallway), whereas attention guidance systems, by definition, do so explicitly. Both are forms of automation in which a computer system decides that the human should redirect attention. As such, both have the potential to err and, in the language of signal detection theory, to do so in either of two different ways: (1) by committing a false alarm, directing attention to a problem that does not exist (Breznitz 1983); or (2) by committing a miss, failing to direct attention to a problem that is real. More regarding the issue of highlighting to achieve attentional direction is discussed in chapter 5.

A framework for understanding these systems is provided in Figure 3.2, which depicts the automation system (top) and the human (bottom) monitoring in parallel a domain of raw data in which events—typically hazards—occur (Getty et al. 1995; Sorkin and Woods 1985). Performance of both the automation system and the human perceiving automation's output in parallel with the raw data and integrating the two can be represented in the context of the signal detection theory matrix, shown at the top of the figure. Such a matrix considers the two types of responses—signal present or absent—that can be made in the face of the two states of the world (signal or no signal), and the matrix then yields the four types of decision events that can result.

Following is discussed the research bearing on four key factors that influence the effectiveness of such systems in appropriately capturing and guiding
**Figure 3.2** Parallel human and automation alerting system. The classic signal detection theory matrix is shown in the insets at the top, representing performance of the automated alert system alone and of the human relying both on the automated system and direct perception of the raw data. Later in the chapter, a discussion is given of the research bearing on four key factors that influence the effectiveness of such systems in appropriately capturing and guiding attention: (1) physical salience (bottom-up conspicuity); (2) reliability; (2) response bias; and (4) the frequency of the critical events that occur in the raw data of the alerted domain.

Attention: (1) physical salience; (2) reliability; (3) response bias; and (4) the frequency of the critical events that occur in the raw data of the alerted domain.

### Alert Salience

Drawing from consideration of the fundamental properties of the human senses (Proctor and Proctor 2006), it becomes apparent that auditory stimuli, or auditory onsets, are the most reliable attention grabbers for alarm systems. This is because the detection of auditory events is omnidirectional. That is, a sound is nearly equally salient no matter how the head is oriented. This of course is not the case with vision, where salience falls off rapidly with visual angle away from the fovea and visual events tend to be invisible at angles beyond around 90 degrees. Recently research has also focused on the tactile modality as being nearly as efficient and omnidirectional as the auditory modality in capturing attention (Sarter 2007; Sklar and Sarter 1999; van Erp 2006; van Erp, Veltman, and van Vern 2003).

Despite the auditory superiority in direct alerting, auditory alerts can also present problems because of that very attribute of intrusiveness. They may well interrupt ongoing tasks that could be of even higher priority (Ho, Tan,
Chapter three: Attention Control

and Spence 2004; Wickens and Colcombe 2007a,b), a phenomenon known as auditory preemption that is discussed again in chapters 8 and 9. It is in part for this reason that designers have often considered less disruptive visual alerts, particularly for those events that may be less critical for safety (Latorella 1996). In this regard we can again draw on guidance from more basic research.

The most direct generalizations from basic research on attention capture to alarm and alert design discussed already is that onsets tend to capture attention. Hence, the most effective visual alarm will be the flashing or blinking signal, because the flashing entails repeated onsets, any of which may eventually be noticed as the eyes are busy scanning the environment (Wickens and Rose 2001). A second generalization is that unique colors can be effective as alarms and alerts (e.g., a red light in a swarm of green). However, the ability of a singleton to capture attention will be reduced if nearby stimuli are also color coded; as discussed in Chapter 5, the salience of a uniquely colored item is decreased by background stimuli of heterogeneous color.

The most severe limitation to generalization of the conclusions drawn from basic attention capture research is the fact that in most environments where warnings are presented, the visual space that the operator is required to monitor is much larger than the typical computer screen—where, for example, onsets are sometimes guaranteed to capture attention (Yantis 1993). In this light, we must acknowledge that there are many circumstances in which singletons and even onsets do not invariably capture attention. A perfect example here is the failure of the green box onset in automated aircraft to capture the pilot’s attention to changes in automation mode (Nikolic, Orr, and Sarter 2004; Sarter, Mumaw, and Wickens 2007). Because important alarms are typically auditory, considerable effort has been taken to define the physical parameters that can convey different levels of urgency in attention capture (Edworthy, Loxley, and Dennis 1991; Marshall, Lee, and Austria 2007). Here, however, caution must be taken because increasing urgency may also lead to increasing listener annoyance (Marshall, Lee, and Austria 2007)—a real problem if the alarms are false, as discussed following.

One final concern when designing visual alerts for attention guidance is the possibility that highly salient exogenous cues may mask the raw data beneath because of their close spatial proximity to those data (i.e., less than one degree). For example, consider the soldier whose attention is cued by the onset of a box surrounding the location of a potential enemy target on a head mounted display (Maltz and Shinar 2003; Yeh and Wickens 2001a; Yeh et al. 2003; Yeh, Wickens, and Seagull 1999). Here, if the raw stimulus information (i.e., the image of the enemy target) is faint, the soldier may not be able to easily discern whether or not a target is actually present behind the visual cue or, worse yet, whether the target is actually an enemy. If the automation is always correct in designating location and identity, then this perceptual challenge will present no problem. But if the automation is less than perfectly reliable in making its classification, the consequences of degrading the
view of the raw data image can be severe. Good human factors design would therefore suggest that the concern of masking can be addressed by using an arrow adjacent to and pointing toward the target rather than a shape surrounding the target object. This leads to the second important feature of attention-guiding automation: its reliability.

**Alert Reliability**

When alert systems are asked to detect dangerous events based on uncertain data or are asked to predict events (e.g., mid-air collision, a hurricane track intersecting a city) in a world that is inherently probabilistic, they will sometimes be wrong, driving their reliability below the perfect value of 1.0. In particular, with predictive alert systems, the longer the look-ahead time or span of prediction, the lower that reliability will typically be (Thomas and Rantanen 2006). Automation reliability, mimicking the concept of cue validity previously discussed (Posner, Snyder, and Davidson 1980) will affect overall system performance as well as two related but distinct attributes of user cognition and behavior: trust and dependence (Lee and Moray 1994; Lee and See 2004; Parasuraman and Riley 1997). Trust represents a subjective belief that automation will perform as expected. Dependence, often correlated with trust, is the actual behavioral tendency of the human to do what automation prescribes (e.g., look where the automation says to look). Clearly, both of these measures will drop as automation reliability declines. But it is easy to envision systems when trust may be fairly low (i.e., the human expects automation to make errors) but when dependence nevertheless remains high, either because (1) the automation, though imperfect, is adequate, or (2) the human operator’s workload is high, so attentional resources are needed elsewhere. Because the major attentional implications of imperfect automation are related to behavioral dependence rather than trust, the focus here is on the former—behavioral measure.

It is clear that the total human–system performance of the parallel system shown in Figure 3.2 will improve as diagnostic automation reliability approaches 1.0. However, the relationship between automation reliability and system performance is not perfect. When the automation alone is better than the human alone, combined system performance will generally be better than unaided human performance but worse than the automation by itself (Parasuraman 1987; Wickens and Dixon 2007). Thus, the human typically depends somewhat, but not fully, on automation. However, there is some evidence that when reliability of the automation is very low (e.g., less than around 0.70), the human may continue to partially depend on it even when his or her own unaided performance is superior to that of the automation. Wickens and Dixon analogized automation in such circumstances to a concrete life preserver: the human would be better off letting go of the automation rather than clinging to it. Recently, for example, researchers have found that physicians depending on automated devices to detect tumors
Figure 3.3 Monitoring imperfect automation within a dual-task context.

on mammograms do no better—and in some respects worse (more false alarms)—than physicians without such aids (Fenton et al. 2007). Of course, the other side of this story is that an automated attention guidance system will generally assist human performance, even if it is less than fully reliable, so long as that reliability is above about 0.80 (Dixon and Wickens 2006; Xu, Wickens, and Rantanen 2007).

Automation reliability has particularly important effects on attention within the multitask contexts, in which automation (e.g., a warning or alert) is likely to be most valuable, because attention is usually needed for purposes other than monitoring the raw data (Figure 3.3). A quantitative analysis of the literature indicates that, on the one hand, dependence on automation is greatest when workload is high, such as in a dual-task context (Wickens and Dixon 2007). On the other hand, this literature suggests that it is the automation-supported detection task rather than the concurrent tasks that suffers most as automation reliability declines, as if the human continues dedicating necessary attention to the concurrent task to preserve its performance—an issue of resource allocation discussed in chapters 7 and 8.

In studying the effects of automation detection and diagnosis imperfection, it is always important to distinguish between the effect of a first failure, when the operator encounters the initial failure of a system that has heretofore in his or her experience performed perfectly, and the effect of subsequent failures, which occur at a long-term rate inversely proportional to the automation’s reliability level (Molloy and Parasuraman 1996; Yeh and Wickens 2001a; Yeh et al. 2003). The consequences of a first failure are often large and potentially catastrophic, as the operator may have become complacent, assuming a perfect system and therefore allocating minimal, if any, attention to monitoring the raw data (Parasuraman, Molloy, and Singh 1993). Subsequent performance, however, becomes more stable as the operator...
adjusts his or her attentional strategies to the long-term statistical quality of the system. The next section turns to these.

The Nature of Automation Errors: Error Salience and the Alert Threshold

When automation reliability declines, automation errors become more common. It turns out that it is not only the number of errors but also the kind of errors that influence human trust and dependence. One particularly noteworthy factor that bears directly on attention is the obviousness or salience of the errors that automation does make. If the automated system's errors are made on judgments that seem easy to the operator (Madhavan, Weigmann, and Lacson 2006) or if the consequence of such errors is large, those errors will be more obvious to the operator and hence will undermine trust and dependence to a greater extent.

Whether errors are obvious or subtle, given an imperfect alerting system alarm designers—and often users—typically have the option of adjusting the threshold of the system to trade off between the two kinds of automation errors within the context of signal detection theory shown in Figures 3.2 and 3.3, choosing anywhere between a low threshold (i.e., hair trigger), which ensures that misses are rare but also allows frequent false alarms, and a high threshold, which reduces the risks of false alarms by increasing the probability of misses. In many predictive alerts, such as collision warnings, such misses are actually late alerts (Maltz and Shinar 2004). In the terminology of signal detection theory, this is an adjustment of the alarm's response criterion (beta). In most real world systems, the costs of misses outweigh those of false alarms, and so the threshold is set low (Getty et al. 1995); consider the difference in consequences of a fire alarm that fails to detect a true fire versus one that produces a false alarm.

It turns out that this threshold adjustment will influence two different cognitive states of the human user—reliance and compliance—which have direct impacts on attention and via this exert influences on performance (Dixon and Wickens 2006; Maltz and Shinar 2003; Meyer 2001, 2004), as depicted in Figure 3.3. Reliance is the cognitive state that allows the operator to feel confident that there really is no hazard at the times when the alert is silent. Compliance, conversely, is the cognitive state that allows the operator to act confidently in response to an alarm when it occurs. Both of these states can be thought of as subcategories of automation dependence discussed already. The system, prone to a high miss rate, resulting from setting a high alert threshold, will lower the human's reliance on the automation to issue an alert if there really is a dangerous event. As a consequence, as shown in Figure 3.3, the human will be more likely to reallocate attentional resources or residual attention away from the concurrent tasks to monitor the raw data of the automated task to pick up the events that automation misses. As a
result of this allocation, concurrent tasks will suffer (Wickens and Colcombe 2007b; Dixon and Wickens 2006), but the (now frequent) failures of automation to detect those events will actually be easily caught by the operator, who expects those automation misses to occur. We may think of this as a drop in complacency—a healthy skepticism (Maltz and Shinar 2003). In contrast, as shown at the top of Figure 3.3, the false alarm (FA)-prone system, resulting from a low alert threshold, will induce a loss of compliance. Suspecting a false alarm, the human may delay taking an action in response to the sounding of an alarm—whether true or false—or may even decline to respond at all. This is the cry wolf effect (Breznitz 1983; Sorkin 1989). As shown in Figure 3.3, in terms of attentional effects, low compliance reflects a slowing of on the speed of attentional switching to any alarm, whether true or false.

It is noteworthy that even though the states of reliance and compliance are logically independent, in practice they often appear to vary in tandem (Dixon and Wickens 2006). Thus, although a low threshold should leave the human confident that all true events will be alerted—and hence should allow full attention allocation to concurrent tasks—FA-prone systems in fact are often as disruptive of concurrent tasks as miss-prone systems, if not more so. This disruption is seen both in the laboratory (Dixon, Wickens, and McCarley 2007) and in the real world of alerting systems (Bliss 2003). Two explanations suggest themselves. First, as noted, every alert will usually be responded to in some fashion by the operator, whether true or false, and the added frequency of false alerts will thereby cause an added frequency of attention shifts away from the concurrent tasks. This effect alone, however, cannot fully account for the detrimental effects of automation false alarms on concurrent task performance, since this performance suffers even on trials where the automation commits no false alarm (Dixon, Wickens, and McCarley 2007).

Second, false alerts are often more perceptually obvious and cognitively disruptive than automation misses—which may not be noticed at all, if their consequences are not serious—and this greater salience of automation false alarms than misses may lead the human to perceive the automation as overall less reliable, leading to a general decline in automation dependence with negative implications for reliance as well as compliance (Dixon, Wickens, and McCarley 2007).

**Event Base Rate**

Finally, the influence of the base rate of to-be-alerted events is noted (Getty et al. 1995; Parasuraman, Hancock, and Olofinbaba 1998). If the need to avoid misses (the cost of misses) remains constant, even when the events to be detected are extremely rare, then a low-threshold setting designed to avoid misses will produce extremely high false alarm rates. That is, the probability that an alarm will sound given that there is no event in the real world will be very high. Using a different statistic, the positive predictive value (PPV) of
the alert (i.e., the probability that an alert will actually signal the dangerous event) can be quite low—well under 0.50 (Getty et al. 1995; Krois 1999). Most alerts will be false. As a consequence, it is easy to understand why people frequently turn their alarms off when those alarm systems generate false alarms to events that occur quite rarely (Parasuraman and Riley 1997; Sorkin 1989).

**Solutions**

The most obvious solution to the issues of human performance costs to imperfect alerting automation is to increase the reliability in discriminating dangerous from safe events. Though this often entails a purely engineering solution (i.e., designing better algorithms), one human factors approach that can sometimes be taken is to reduce the look-ahead time of predictive automation such as that involved in forecasting events or collision warnings. Of course, the look-ahead time should be no shorter than the time necessary for the human to respond appropriately to the alerted state. For example, it makes no sense to have a look-ahead time for a pilot collision warning of five seconds. Though this guarantees high accuracy, five seconds is too little time to maneuver and avoid collision (Thomas and Rantanen 2006).

A second solution is to adopt likelihood alarms, whereby the alert system itself signals its own degree of uncertainty in classifying events as a dangerous signal versus a safe noise (Sorkin and Woods 1985; Xu, Wickens, and Rantanen 2006). However, the operational success of this approach in a dual-task context remains ambiguous (Sorkin, Kantowitz, and Kantowitz 1988; St. John and Mannes 2002; Wickens and Colcombe 2007a,b). A third approach, which is discussed more in chapter 9, is based on the concept of preattentive referencing (Woods 1995). Here the human is given access to continuous information about the evolving state of the alert domain (e.g., the raw data), often in nonfocal sensory channels such as peripheral vision or sonification, which might be used, for example, to represent the continuous sound of a heart beat in an intensive-care monitoring workstation (Watson and Sanderson 2004). Finally, solutions in training can be suggested, typically in training different aspects of attention allocation to be calibrated with actual system reliability as well as training the alarm user to understanding of the inevitable nature of the trade-off between misses and FAs as the threshold is varied and therefore to better tolerate the high FA-rate that must be expected when low base rate events are coupled with imperfect diagnostic automation.

**Conclusion**

This chapter has discussed properties of events that capture or grab attention, in both the theoretical laboratory-based context and that of the real world, where alarms were the most important application. However, other attention forms of cueing devices, such as highlighting, can be equally relevant to basic research on cueing. In both the theoretical and applied environments, the
reliability with which the cue actually directed attention to a meaningful target or event proved to be a critical concept.

We also discussed the characteristics of events and the observer that failed to capture attention, in the context of change blindness. In both successes and failures of attention capture, this chapter's focus was on environmental properties and hence on factors that guide attention in a so-called bottom-up event-driven fashion. Yet we know also that our mind can choose where to attend and often can override these bottom-up effects of salience using top-down or knowledge-driven processes. Thus, the following chapter integrates all the factors responsible for guiding our selection, whether bottom up or top down, and emphasizes the very prominent role of visual scanning in selective attention.