CHAPTER SIX

Control of Visual Attention

Steven Yantis
Johns Hopkins University, Baltimore, USA

Vision allows organisms to know the contents and layout of their local environment. Its major function is object recognition. Experimental evidence accumulated over the last 30 years suggests that in many (but not all) cases, the visual system accomplishes object recognition by visually selecting a relevant or salient part of the visual image (e.g. the cluster of features constituting an object or located in a region of space) and operating only on that cluster, then selecting another part of the image, and so forth. This strategy reduces the complexity of object recognition by limiting it to only one or a small number of elements at a time. The mechanism that accomplishes selection is called visual attention, and this chapter is concerned with how visual attention is deployed within an image as a function of image properties and observer goals.

A major distinction that has guided research in this area, and one that will provide an organizing principle for this chapter, is whether attention is goal-driven, controlled in a "top-down" fashion, or stimulus-driven, controlled in a "bottom-up" fashion. This distinction has been recognized at least since William James first introduced it a century ago in his seminal book, *The principles of psychology* (1890). James characterized this distinction in terms of "active" and "passive" modes of attention, respectively.

Attention is said to be goal-driven (in James' words, "active") when it is controlled by the observer's deliberate strategies and intentions. For example, if one is looking for a particular type of cereal at the supermarket, and that cereal is known to have a yellow box, then yellow boxes in general are likely to be selected by attention and recognized. In contrast, attention is said to be stimulus-driven (or "passive") when it is controlled by some salient attribute of the image that is not necessarily relevant to the observer's perceptual goals. For example, if cereal boxes in one part of the aisle all tended to be blue, then a single red box among them might seem to "pop out" of the background and draw attention automatically.

Although the distinction between goal-driven and stimulus-driven control of attention is important, it is equally important to recognize that any given act of attention typically involves some combination of the two attentional modes. One of the themes of this chapter is how the two types of attentional control are coordinated to yield coherent visual performance.

Psychologists have devised a number of techniques for assessing how goal-driven and stimulus-driven selection occur. Perhaps the most revealing and widely used of these is *visual search* (see Wolfe, this volume). Visual search tasks require observers to view a display containing multiple elements and to find a prespecified *target* element among several *distractor* elements. Experimenters typically measure response time (RT) during visual search to assess the efficiency with which a target is detected or identified. In such tasks, it is useful to distinguish between the *defining attribute* of the target and its *reported attribute* (Duncan, 1985). The defining attribute is what the observer must search for during the task; it may be simple, such as the color or orientation of a bar, or complex, such as the conjunction of two or more features (e.g. large red vertical bar). The reported attribute is often simply the target's presence or absence, but in many instances it may be the name of the target or some attribute other than the defining one. When an observer is required to report the orientation of the red bar, the defining attribute is color and the reported attribute is orientation.

When experimental psychologists study attention, they must take certain precautions in order to ensure that the observed empirical effects are attentional and not due to some other, nonattentional, factors. For example, in everyday life it is standard to look at what we wish to attend to. The relationship between attention and eye movements is complex (see Hoffman, this volume); therefore in most of the experiments reviewed in this chapter subjects are required to maintain fixation and to attend to stimulus events "out of the corner of their eyes." When the maintenance of fixation is crucial to the interpretation of the experiment, eye movements are monitored. The stimuli are typically positioned so that variation in visual acuity with retinal eccentricity is minimized (for example, they might be arranged in a circle centered at the fixation point, so all the stimuli are equidistant from the center of gaze).

In the following sections, I will review evidence concerning goal-driven and stimulus-driven selection, respectively. In each case, evidence from visual search and other paradigms will reveal aspects of these two modes of attentional control. We will find that they are by no means independent of one another, and in a concluding section I will discuss how stimulus-driven and goal-driven selection complement one another in vision.

GOAL-DRIVEN VISUAL SELECTION

The modern study of visual attention began in the late 1960s (although several earlier studies will be noted) with investigations of how much control observers have in attending to relevant objects and ignoring irrelevant ones, and this was explored with a cuing paradigm in which relevant objects or locations were indicated visually. Later, interest developed in the distribution of attention in 2- and 3-dimensional space, and how attention is shifted from one location to another. In this section, each of these issues is addressed in turn.

Attention to Spatial Locations

The first well-known claim that observers can direct attention to locations in space other than their current point of fixation was made by Helmholtz (1866/1925). He performed an experiment in which a picture was placed in a light-tight box containing a pair of holes through which the picture could be viewed, and a pair of pinholes on the opposite wall that served as fixation points. The picture was illuminated briefly by a spark. This experiment was carried out in a series of studies of depth perception, but Helmholtz (1866/1925, p.455) made the following observation in passing:

> It is a curious fact, by the way, that the observer may be gazing steadily at the two pinholes and holding them in exact coincidence, and yet at the same time he can concentrate his attention on any part of the dark field he likes, so that when the spark comes, he will get an impression about objects in that particular region only. In this experiment the attention is entirely independent of the position and accommodation of the eyes, or indeed, of any known variations in or on the organ of vision. Thus it is possible, simply by a conscious and voluntary effort, to focus the attention on some definite spot in an absolutely dark and featureless field. In the development of a theory of the attention, this is one of the most striking experiments that can be made.

This observation was not pursued with rigorous experimentation until nearly a century later, and the earliest attempts to do so yielded negative results. For example, Mertens (1956) had observers view dim flashes of light under two conditions: "general attention" in which the flash could appear in any of four peripheral locations, and "special attention" in which the flash would appear in a single known location on each trial. Observers were required to press a button when they were able to see the light flash. The author found that the ratio of intensities for the general and special attention conditions required to achieve 25%, 50%, and 75% correct detections was 1.02, 1.11, and 1.18, respectively; that is, in all three cases the intensity required for a given level of performance in the general attention condition was greater than that in the special attention condition, implying an advantage for the latter. Nevertheless, Mertens (1956, p.1070) concluded that there was "practically no influence of the attention upon the probability of observation". He argued that the ratios for the two higher proportions, which seem to show clear attentional effects, should be ignored because of possible response bias effects. As van der Heijden (1992) and others have pointed out, however, Mertens's conclusion and the reasoning on which it is based are not entirely convincing.

A second failure to observe attentional effects was reported more than a decade later by Grindley and Townsend (1968). In their experiments, subjects were to report the orientation (up, down, left, or right) of a briefly flashed *T* appearing in one of four locations. They compared a condition in which the location of the *T* was known in advance to one in which it was not, and found essentially no advantage in accuracy when location was known in advance (at least when there was only a single item in the display; with multiple distracting characters present, foreknowledge did improve performance). Grindley and Townsend (1968) concluded that attention to location does not improve discriminability within single-item displays.

So the first two attempts to measure an attentional advantage after Helmholtz's original observation met with apparent failure. However, in many subsequent experiments, an attentional advantage was observed. Advantages were found both in how accurately subjects could detect or identify a briefly-presented object (e.g. Bashinski & Bacharach, 1980; Egly & Homa, 1984; Shaw & Shaw, 1977; Van der Heijden & Eerland, 1973), and in how rapidly subjects could detect or identify an above-threshold stimulus (e.g. Jonides, 1981; Posner, 1980; Posner, Snyder, & Davidson, 1980; Shaw, 1978).

Among the earliest studies to exploit Helmholtz's observation about one's ability to direct attention "by a conscious and voluntary effort" was a study concerned not with attention at all, but with perceptual memory. Sperling (1960) was interested in measuring the duration and capacity of visual memory, and so he invented the *partial report technique*. In a typical partial report experiment, an array of letters (say, three rows of four letters each) was illuminated very briefly, and this was followed by a tone (either immediately or after a delay). The tone could be high, medium, or low in pitch. Subjects were to report the contents of the top row if the high tone sounded, and so forth for the other two tones. The purpose of the procedure was to avoid requiring subjects to recall the contents of the entire array, which took time during which visual memory of the array might be decaying. In partial report, subjects had to report only part of the array, but their "readout" of the contents of the array had to come from visual memory, because the tone did not sound until the array had physically disappeared. A modification of this procedure was carried out by Averbach and Coriell (1961); in their experiments, a single letter was cued after the offset of the array with a circle surrounding the location that had been occupied by the cued letter. For our purposes, the interesting part of these experiments is that subjects could easily direct their attention to the relevant location in the remembered array and read out its contents.[1]

Criteria other than spatial location can also be used to direct attention in the partial report task, including differences in color or brightness (von Wright, 1968). For example, if an array containing half red items and half green items (randomly interspersed) is flashed very briefly, and a partial report cue is then presented (say, a high tone means report the red items and a low tone the green items), then subjects can successfully direct attention to the cued subset. However, more complex criteria, such as whether an item is a letter or a digit, cannot be used. This suggests that only "sensory" factors are useful in directing attention, but that more meaning-based factors, which require prior identification, cannot.

The earliest direct studies of the control of visual attention in the modern era were reported by C. W. Eriksen and his students in the early 1970s. They were concerned with an issue that remains unresolved today: to what extent are visual objects identified *before* attentional selection rather than after? This question is an extension of a debate that began in the early 1960s concerning the locus of selection in audition, the so-called early- vs late-selection debate. On one side, it was argued that only very simple stimulus attributes (e.g. pitch of a voice) can be detected preattentively and used to direct attention (Broadbent, 1958); on the other side, it was argued that stimuli are completely identified preattentively, and that attentional limitations arise only when a response must be selected (e.g. Deutsch & Deutsch, 1963). So the question as formulated by Eriksen and colleagues in the domain of vision was this: if an observer focuses attention on an object with the goal of identifying that object, to what extent will nearby but to-be-ignored objects interfere with the identification process?

In their experiments (e.g. Eriksen & Hoffman, 1973), subjects viewed a display containing multiple letters arranged in a circle so that all the letters were the same distance from the fixation point. At various moments before the letters appeared, a bar marker was presented. The bar marker indicated one of the display locations that would contain a letter. The subjects were

---

[1] The reader may be interested to know that Sperling (1960) and others (see Coltheart, 1980, for a comprehensive review) found that visual sensory memory has a very high capacity (that is, almost all of what is seen is remembered), but the contents of that memory decay rapidly, in less than one second.

required to identify the letter that was indicated by the bar marker (the *target*) and to ignore all the other letters in the display (*distractors*). Subjects were to push one switch if the target was (say) *A* or *U*, and another switch if the target was *H* or *M*. The distractors could appear at different distances from the target, and they could have various identities. Typically, most of the distractors were response-neutral (i.e. letters to which no response was ever required).

The critical comparison in this experiment was the time required to press the button when a nearby distractor's response assignment was compatible with that of the target (e.g. target *A*, nontarget *U* displays) vs when it was incompatible (e.g. target *A*, nontarget *H* displays). The effect of target compatibility could be measured as a function of the distance between the target and the critical distractor and of the duration between the onset of the bar marker and the onset of the display.

Eriksen and Hoffman (1973) made several observations. First, they found that the identity of the distractor did matter: responses were slower when adjacent distractors were incompatible with the target than when they were compatible. Second, the interference produced by incompatible distractors decreased as the distance between the distractor and the target decreased, and it also decreased as the time between the onset of the bar marker and the onset of the display decreased. These findings suggested that subjects require some time to localize the bar marker so as to focus attention on the indicated location, and that the attended region was limited in spatial extent. Yantis and Johnston (1990), using a variant of this paradigm, concluded that attention can be very efficiently focused, given optimal conditions, suggesting that early selection is at least possible under certain conditions.

In the late 1970s and early 1980s, Posner and colleagues (e.g. Posner, 1978, 1980; Posner et al., 1980) developed a technique for exploring top-down control over attention that extended the paradigm used by Eriksen. In Posner et al. (1980), for example, subjects were required to detect the onset of a light appearing in one of four horizontally arrayed positions, and to press a button to indicate their detection (simple detection and not identification was required). At the start of each trial, a digit appeared at fixation indicating which of the four locations (1, 2, 3, or 4) was likely to contain the light. Subjects were told that the light would appear in the cued location on 79% of the trials (these are referred to as *valid-cue* trials), and in each of the other locations on 7% of the trials (*invalid-cue* trials). On some trials, a plus sign appeared instead of a digit; this indicated that the light was equally likely to appear in all four positions (*neutral-cue* trials). Posner et al. (1980) found that responses were fastest when the light appeared in the expected location, slowest when it appeared in an unexpected location, and intermediate when all locations were equally likely. The magnitude of the RT benefit (i.e. the difference between the valid-cue RT and the neutral-cue RT) was approximately the same as the RT cost (i.e. the difference between the neutral-cue RT and the invalid-cue RT).[2]

In closing this section, let us return to the initial failures to observe attentional effects, the experiments of Mertens (1956) and Grindley and Townsend (1968). The many subsequent studies in which effects of directed attention *were* observed leave us with a puzzle: why did Mertens, and Grindley and Townsend fail to find evidence for directed attention? One possible answer lies in a critical methodological difference between these two studies and almost every other study discussed in this section. In both of these studies, the stimuli to be identified or detected were presented in an otherwise blank visual field. If the function of attention is to select relevant information from the visual field, then one might not expect to observe attentional effects in these special circumstances. A review of the literature by Shiu and Pashler (1994), together with their own targeted experiments, corroborate this idea. Attention effects are usually observed only when other, potentially competing, visual stimuli are present in the display.

Spatial Distribution of Attention

The experiments of Eriksen, Posner, and their colleagues established that attention could be directed to a spatial location, and that this location has a limited spatial extent. Several subsequent studies have explored the distribution of attention in space under different conditions. Among the earliest of these experiments was one reported by Engel (1971). Subjects viewed a blank display with a fixation point in it. A visual noise field of randomly oriented densely distributed line segments was flashed; somewhere in the noise was a randomly oriented L-shaped figure (or, in other experiments, a U-shape or a small square). Subjects were required to report the location and orientation of the L. Engel found that performance was good for targets near the point of fixation and declined with eccentricity. This is not too surprising, given the known variation in visual acuity with retinal eccentricity. He defined a "conspicuity area" within which performance was

---

[2] Shaw (1984) has pointed out a difficulty in the simple detection procedure used by Posner and others. In this task, subjects merely have to press a button when a dot or other simple change in the display is detected. Catch trials are included in which no stimulus appears to ensure that subjects respond to the stimulus and do not anticipate its appearance (subjects are instructed not to respond on catch trials). It is possible for subjects in a simple detection task to adjust their decision criterion so that they are willing to respond on the basis of less sensory information in the cued location (because they know that location is more likely to contain a target event), and this could account, at least in part, for their increased speed. This criterion shift problem is ameliorated (but not eliminated) when the task requires subjects to perform a discrimination (e.g. is the target a T or an L?), because errors in perception can then be observed directly. It is worth keeping this issue in mind while reading later in the chapter about other studies that use simple detection.

perfect with no foreknowledge of location and outside of which performance was imperfect; the conspicuity area was a horizontally oriented ellipse centered at fixation (see Fig. 6.1). He then provided subjects with perfect preknowledge about where the target stimulus would be presented; in this case, the observer only had to report the orientation of the stimulus. The region within which the target could be seen under these conditions, termed the "visibility area," was more than twice as large as the conspicuity area. Finally, Engel examined the effect of providing an "attention point" at some position in the visual field; subjects were informed that the target was most likely to occur at this point, but it could appear elsewhere. Observers were required to maintain fixation on the center of the display. Subjects' performance improved when the stimulus appeared near the attention point, and it declined only slightly when the target appeared in other locations within the previously defined visibility region. The conspicuity region simply bulged outward in the direction of the attention point.

Hoffman & Nelson (1981) provided additional evidence concerning the distribution of attention during visual search. Subjects were required to carry out two tasks simultaneously. The first task was to determine which of two target letters appeared in a sequence of briefly flashed four-letter displays. During one of the frames a small U-shaped character appeared in one of four orientations adjacent to one of the letters in one of the frames. The second task was to report the orientation of the U. In different conditions the relative importance of the letter task and the orientation task was varied, providing data to construct a *performance operating characteristic* (POC) which characterizes the performance trade-off between any two tasks (Navon & Gopher, 1979; Sperling & Melchner, 1978).

Hoffman and Nelson (1981) found that when the U was not adjacent to the target letter, there was a substantial performance trade-off between the tasks. That is, an emphasis on the letter task substantially reduced performance in the U task and vice versa. However, when the U was adjacent to the target position, the trade-off was more restricted; that is, subjects performed well on the U task even if the letter task was emphasized. Hoffman and Nelson concluded that in standard visual search tasks, attention is directed to the spatial region containing the target stimulus. To the extent that other relevant stimulus information is nearby, performance on tasks requiring that other information benefits.

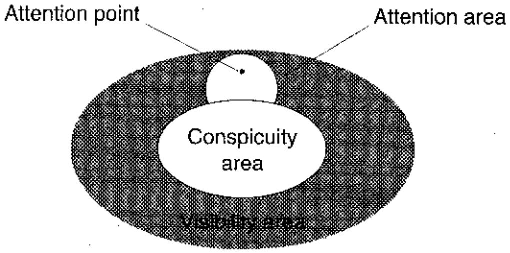

FIG. 6.1. The visibility, conspicuity, and attention areas as conceived by Engel (1971). In the conspicuity area, targets can be seen without foreknowledge of location. In the visibility area, targets can be seen with perfect locational foreknowledge, and the attention area represents a "bulge" in the conspicuity area when there is imperfect foreknowledge of location. Adapted from Vision Research, Vol 11, F.L. Engel, *Visual conspicuity, directed attention, and retinal locus*, pp.563–576. Copyright © 1971, with kind permission from Elsevier Science Ltd, The Boulevard, Langford Lane, Kidlington OX5 1GB, UK.

LaBerge (1983; LaBerge & Brown, 1986) attempted to assess the shape of the attended region. He required subjects to perform two tasks on each trial, one right after the other. The first task required them to determine whether a string of five letters briefly flashed at fixation was a male name or not (the word task) or, in a separate experiment, whether the middle letter in the letter string was a consonant or a vowel (the letter task). The letter string was followed on some trials by a probe dot appearing at one of the five locations previously occupied by the letter string. Subjects were required to press a button as quickly as possible when they detected the probe dot. The first (word or letter) task was intended to produce a specific distribution of attention, and the second (probe) task was then used to map out that distribution. LaBerge predicted that the word task would require subjects to distribute attention roughly evenly across the entire letter string, while the letter task would require subjects to focus attention on the central position in the letter string. He assumed that the distribution of attention produced by the letter or word task would be reflected in RT to detect the probe stimulus. It was: RT to the probe on word trials did not differ as a function of probe position; RT on letter trials, however, was fastest for probes occurring at the middle location and slowest for the extreme positions, producing a V-shaped function.

A related study was reported at about the same time by Downing and Pinker (1985). Subjects were required to detect a spot of light that appeared in one of 10 outline boxes arranged horizontally on a computer screen. Subjects were required to press a button as quickly as possible after one of the boxes was filled in (again, simple RT; see footnote 2). The digits 1–10 were displayed immediately above each box corresponding to the position of that box in the horizontal array. At the start of each trial, a numeral appeared at fixation indicating the likely location of the target that was about to appear (cued trials) or indicating that all positions were equally likely to contain the target (neutral trials). On 18% of the cued trials, no target appeared; these catch trials were included to ensure that subjects were responding to the stimulus and not anticipating its appearance. Of the remaining cued trials, the target appeared in the cued location 70% of the time and in one of the uncued locations 30% of the time. Subjects were encouraged to focus attention on the cued location so as to minimize RT to

Downing and Pinker measured the distribution of attention for a given cue condition by examining RT costs and benefits relative to the neutral baseline RT (see discussion of Posner et al., 1980, in the last section) so that differences in RT due to acuity would be factored out. The results revealed a distribution of attention that was similar to the one reported by LaBerge (1983): there was a V-shaped function surrounding the attended location such that targets appearing closer to the cued location were detected more rapidly than targets appearing farther from the cued location. Interestingly, Downing and Pinker (1985) found that the distribution was modulated in this way only within a hemifield; RT costs and benefits in the hemifield contralateral to the cued location did not vary with distance from the cued location. Furthermore, the attentional gradient was sharpest near to fixation and became shallower when a peripheral location was attended.

Shaw (1978; Shaw & Shaw, 1977) explored a parallel model of attention allocation to spatial locations. Shaw analyzed a visual search task in which there were n objects in the display; the probability that the target would appear in location j was p_j, $j = 1, 2, \ldots n$. According to her capacity allocation model (CAM), there exists a fixed total capacity for visual processing (denoted ϕ) that must be distributed over spatial locations, and the time required to identify a stimulus is a monotonically decreasing function of the capacity allocated to it (i.e. more capacity, faster identification). Shaw derived an optimal allocation function given a probability distribution (the set of p_js) and the total capacity ϕ. She then tested the assumptions of the CAM in an experiment with two different probability distributions and found almost identical estimates for the value of ϕ. She concluded that the CAM characterizes performance in visual search tasks of the sort she analyzed.

One can interpret Shaw's (1978) capacity allocation model as a mechanism that samples from the visual field according to the probability distribution of likely target locations. More samples per unit time are taken at likely locations, and fewer at unlikely locations. Only a limited number of samples may be taken per unit time. This analysis comes directly from the theory of optimal search developed during World War II in which limited search resources (e.g. aircraft) had to be allocated to searching for military assets (Koopman, 1957).

Attending to 3-D Space

The experiments reviewed so far suggest that observers can attend to locations in space according to task demands, and that the distribution of attention is at least somewhat graded. All the studies reviewed so far have dealt with stimuli that lie in the picture plane, which is not, of course, characteristic of real-world scenes. This led naturally to the question of

whether attention could also be directed selectively in 3-D space. The question has significant implications for determining the representational basis of visual selection: if selection is early, before a 3-D representation is constructed, then one might expect not to observe selectivity in 3-space. Several studies have explored the allocation of attention in 3-D space, with seemingly inconsistent results. Recently, however, a consensus has emerged about this elusive issue.

Among the earliest studies of attention in 3-D space was Experiment 1 of Downing and Pinker (1985), whose Experiment 2 in 2-D space was reviewed earlier. Downing and Pinker had subjects view a real 3-D scene (this was not a computer-generated simulation, which is a real break from common practice these days) containing four lights mounted on stalks and placed on a table. Viewing was monocular,[3] and the table was covered with textured cloth to enhance the monocular depth cues available in the scene. In any given block of trials, the lights were placed in four of eight defined locations on the table, four near locations and four far locations in two curved rows. The near and far locations were selected so that corresponding lights in the near and far rows fell on corresponding retinal locations (e.g. the leftmost light position in the near and far rows fell along the same line of sight). A typical arrangement for a block of trials would have lights in the near, far, near, and far rows, respectively, going from left to right. In the center of the display table was an LED that could display a digit. On each trial, subjects fixated the central LED, which then displayed a digit from 0 to 4, where 0 meant all positions were equally likely to contain a light (the neutral cue), and 1–4 meant that the corresponding light (numbered from left to right) was highly likely to be illuminated. Subjects were instructed to attend to the cued position (without moving their eyes) and to press a key when any of the lights was illuminated. The light appeared in the cued location on 79% of the trials, and in one of the uncued locations on 21% of the trials. The cue duration was randomly selected from the range 400–800 ms. No light appeared on 26% of the trials; these catch trials were meant to ensure that subjects did not anticipate the stimulus.

The experimental design permitted Downing and Pinker to assess detection RT for trials in which the light appeared in an uncued location as a function of the retinal distance between the lights and as a function of the distance in depth between the lights. In other words, if light 1 is cued, and a light appeared in location 2, does it matter whether the light in location 2 was at the same depth as light 1? To answer this question, Downing and Pinker computed the costs associated with detecting a light in an unattended

---

[3] The near and far lights were outside of Panum's Area (the region within which binocular images can be fused) and this would have caused double images if the lights were viewed binocularly.

location. They did this by subtracting the RT to detect a light in an unattended location from the RT to detect a light in that same location when they were not selectively attending anywhere (i.e. when they had received the neutral "0" cue). These costs were then plotted as a function of where the stimulus appeared relative to the cued location separately for same-depth and different-depth trials. If attention could be directed only to a retinal location and not to a specific depth, then these two functions should be superimposed.

That was not the case. First, as expected, costs increased with increases in 2-D separation. For example, when location 1 was cued, costs were greater if the stimulus appeared in location 4 than if it appeared in location 2. More importantly, costs for different-depth trials were greater than for same-depth trials. For example, if location 1-near was cued, costs for location 4-far were greater than costs for location 4-near. This result suggests that attention did rely on a representation containing depth information.

A more recent study by Gawryszewski, Tiggio, Rizzolati, and Umiltà (1987) was reported in which attention to locations in 3-D space was measured, and reliable costs and benefits to locations at different depths were also observed. It is important to note that the study of Gawryszewski et al. (1987), like that of Downing and Pinker (1985) employed real-world 3-D displays.

Several other studies, all using computer-generated stereoscopic displays, have been reported in which attention did not appear to be directed to locations in 3-D space. For example, Iavecchia and Folk (1996) found no effect of whether a target stimulus in an uncued location was in the same depth plane as the cued location.

Ghirardelli and Folk (1996) have also recently reported no attentional effect of the depth plane in which a stimulus appears. In their study, observers viewed a stereoscopic display in which one of two possible characters (e.g. A or H) was presented, and they had to decide as quickly as possible which character was presented. On each trial, two possible stimulus locations were defined by a pair of location markers; one of the location markers was brightened for 100 ms, and after an additional 50 ms, a target character appeared in one of the two marked locations. The target appeared in the cued (brightened) location on 80% of the trials and in the uncued location on 20% of the trials.

The location markers appeared either at the same depth (i.e. with zero disparity) but in different x–y positions (i.e. left and right of the fixation point), or at the same x–y position (i.e. at fixation) but in different depth positions (with crossed disparity, in front of the fixation point, or uncrossed disparity, behind the fixation point). The former condition is similar to standard spatial cuing procedures (e.g. Posner et al., 1980). The key question asked by Ghirardelli and Folk (1996) was whether observers could selectively attend to one of two positions that differed only in depth.

The answer was clear. When the two locations differed in x–y coordinates, there were significant benefits when the target appeared in the cued location and significant costs when it appeared in an uncued location (both relative to a neutral condition). This is consistent with many previous studies showing the same thing. However, when the two locations differed only in depth and not in x–y location, there were no costs or benefits at all. It was as if subjects could attend to both locations simultaneously. Ghirardelli and Folk (1996) concluded that observers cannot selectively attend to different locations in 3-D space if they do not also differ in 2-D retinal projection.

This leaves us with an apparent inconsistency: some studies (e.g. Downing & Pinker, 1985; Gawryszewski et al., 1987) have provided evidence that it is possible to attend to locations in 3-D space, while others (e.g. Ghirardelli & Folk, 1996; Iavecchia & Folk, 1995) suggest that attention cannot be directed selectively to locations in 3-D space. One difference between these two sets of studies is that the former involved real-world scenes (objects placed at different physical depths relative to the observer), while the latter involved computer-generated simulations of depth. However, a recent report by Hoffman and Mueller (1994) suggests that this is not the key difference. Instead, Hoffman and Mueller (1994) argue that what is crucial is whether there exist well-defined perceptual objects to which attention can be directed in advance. In their study, the to-be-attended locations were defined by placeholder objects, one of which transformed into the target object. In this case, observers were evidently able to attend selectively to a cued object that differed from the uncued object only in depth. Hoffman and Mueller (1994) concluded that object-based selection may operate on a representation that contains depth information.

In summary, the studies reviewed in this section suggest that, although attention cannot be directed to empty locations in 3-D space, attention to objects at different 3-D locations may be possible.

Shifts of Attention

The work reviewed so far has been concerned with determining how attention is distributed in space at a given moment in time. This leaves open the question of how attention is shifted from one location to another. Before reviewing some of the relevant studies, it will be useful to sketch the space of possibilities. One dimension along which models of attention shifts differ is whether the movement of attention is continuous or discrete. A continuous model holds that attention should be conceived as something like a spotlight: it "illuminates" a restricted convex region of the image, it permits information from the image to pass only from that region, and it moves continuously from one location to the next as a spotlight sweeps across the surfaces that it illuminates as it moves. This is contrasted with a "gradient" or "zoom-lens" model according to which the size but not the location of the attended region may change continuously; according to these models, changes in location occur discontinuously (for example, the "amount" of attention at one location might gradually decline as the "amount" at another location increases). One can imagine various hybrid possibilities as well.

Among the first to examine this question were Shulman, Remington, and McLean (1979). The experiment was designed to test a prediction of the class of spotlight models: When attention moves from one location to another, it "illuminates" intervening locations. Therefore, we might expect that responses to objects that lie on the presumed path of an attention movement should be facilitated by attention at some point during the course of the attention movement. Certain aspects of the data obtained by Shulman et al. were consistent with the analog spotlight hypothesis, and the authors concluded that the spotlight hypothesis was supported. However, Yantis (1988) and Eriksen and Murphy (1987) have pointed out that other aspects of their data are difficult to reconcile with the spotlight hypothesis. These inconsistencies have not yet been clarified empirically, and so evidence concerning the "attention to intervening locations" hypothesis remains uncertain.

Most of the remaining evidence about the mechanisms for shifts of attention concerns the timecourse of attentional shifts. One prediction of a continuous spotlight hypothesis is that if the spotlight of attention moves at a constant rate, then it should take longer to shift attention a long distance than to shift it a short distance.[4] This was the focus of an experiment reported by Remington and Pierce (1984). Subjects were required to detect the onset of a light appearing at one of two locations to the right or left of fixation; the eccentricity was either 2° or 10° on any given trial. At various moments in time before the light flashed, an arrowhead appeared at the fixation point indicating the side on which the light was likely to appear. On some trials, instead of an arrowhead, a neutral cross, providing no spatial information, appeared instead. The purpose of the neutral cross cue was to provide subjects with a warning about when the target flash would appear without also providing them with spatial information about where to direct attention. In this way, Remington and Pierce could separately estimate the timecourse of the spatially specific attentional effect uncontaminated by changes in readiness or arousal that are also known to accompany a cuing signal (e.g. Bertelson, 1967; Posner & Boies, 1971). Subtracting the nonspatial warning effect from the spatially specific attentional effect, Remington and Pierce (1984) found that response times improved to a minimum, asymptotic, level as the cuing interval increased, and that the asymptotic level occurred at the same cuing interval for different distances. They concluded that the time required for a shift of attention does not depend on the distance over which attention must be shifted.[5] This result undermines a simple attentional spotlight hypothesis that assumes a constant-velocity trajectory.

Several additional experiments have corroborated the findings of Remington and Pierce (1984). Kwak, Dagenbach, and Egeth (1991) presented displays consisting of two letters and required subjects to judge whether they were the same or different. The two letters appeared on the circumference of an imaginary circle with radius 4.5°; the distance between the letters varied from one trial to the next. In their Experiment 2, the letters were rotated Ts and Ls; evidence from Experiment 3 of Kwak et al. (1991) and elsewhere (e.g. Egeth & Dagenbach, 1991) strongly indicates that these letters are processed strictly sequentially, which makes them useful for assessing movements of attention. If attention moves with a fixed velocity, then responses should be slower when the letters are far apart than when they are close together. The data revealed no such variation with distance, however. Kwak et al. (1991) concluded in favor of time-invariant shifts of visual attention. Kröse and Julesz (1989) drew similar conclusions based on data from a paradigm in which response accuracy, rather than response time, was the dependent variable.

These experiments are based on the notion that attention must be directed to a location in space with a fixed "attentional aperture." Eriksen and St. James (1986) proposed an alternative "zoom-lens" model of attentional focus. According to the zoom-lens model, attention is directed to a region of space the size and location of which changes with task demands. If a subject is required to focus attention on a single location indicated by a bar marker, the locus of attention begins in a diffuse state, so that the entire display is attended (in preparation for the appearance of the bar marker), and then it "zooms in" on the indicated location over time. If multiple contiguous locations are relevant, attention can be distributed over the larger region containing those locations, but at some loss in speed. Eriksen and St. James (1986)—using a procedure similar to that of Eriksen and Hoffman (1973), described earlier—probed the interference produced by incompatible distractors at various moments in time after the onset of the

---

[4] Of course, movements of attention could move at a variable velocity (e.g. attention could move faster the farther it has to go) and still be continuous. However, the constant velocity assumption is a reasonable starting point.

[5] An earlier study by Tsal (1983) had similar logic but came to a different conclusion. However, Tsal's experiments failed to include the critical neutral-cue control condition. See Yantis (1988) for a discussion.

bar marker, and found that the spatial extent over which interference occurred decreased over time as predicted by the zoom-lens model.

A consistent but incomplete picture emerges, then, concerning shifts of visual attention. They clearly do not take more time as the distance to be traversed increases. This result casts some doubt on a simple spotlight metaphor. However, little enough is known about the principles governing shifts of visual attention, so that not much more can be said with confidence. This difficult issue requires more theoretical and empirical analysis.

STIMULUS-DRIVEN CONTROL OF ATTENTION

The evidence reviewed in the previous section shows that people are able to deploy attention to relevant objects or regions of space when they wish to do so. We now turn to the evidence concerning how stimulus events that are not explicitly represented in the observer's state of attentional readiness may capture or otherwise modulate attention. At issue is the extent to which stimulus events can control the distribution of attention independently of the goals and intentions of the observer.

As I stated earlier, William James wrote about attention in his classic *Principles of psychology* a century ago, and he noted that certain kinds of events are likely to automatically draw attention. Referring to this variety of attention as passive and immediate, he said that certain stimuli have a "directly exciting quality;" examples include "strange things, moving things, wild animals, bright things, pretty things, metallic things, words, blows, blood, etc., etc., etc." (James, 1890, pp.416–417). In this section, we will examine recent empirical evidence that clarifies what sorts of stimulus events do capture attention in a stimulus-driven fashion.

Attentional Capture by Spatial Cues

As we saw earlier, attention can be directed voluntarily to spatial locations by informative visual cues (e.g. Eriksen & Hoffman, 1973; Posner et al., 1980). It was initially assumed that a salient display change in the visual periphery would capture attention automatically; Jonides (1981) set out to test this assumption empirically. Jonides used a visual search task in which subjects viewed an array of eight letters arranged in a circle so that each letter in the array was equidistant from the fixation point. Every array contained either an *L* or an *R*, and subjects were to determine which of these targets was present by pressing the corresponding left or right button. Shortly before the letter array appeared, an arrowhead cue appeared indicating one of the display locations. On some trials the cue was valid and indicated the location in which the target letter eventually appeared, and on others it was invalid. The arrowhead cue appeared either at fixation (a central cue) or near the letter location that it indicated (a peripheral cue).

The question of interest was whether these two different cue types had different effects on the deployment of attention. Jonides speculated that the peripheral cue would capture attention "automatically", whereas the central cue would require voluntary effort to direct attention to the indicated location. Three criteria for automaticity were used to assess the automaticity of the central and peripheral cue. An automatic process does not require significant mental capacity, it is resistant to suppression, and it does not require deliberate intent on the part of the observer to have its effect (Shiffrin & Schneider, 1977).

Jonides examined these criteria in three experiments; I will describe his Experiment 2 in which resistance to suppression was measured. The cue validity in this experiment was 12.5%; because there were eight letters in each array, the cue provided no predictive information about the likely location of the target (i.e. 1/8 = .125). One group of subjects was instructed to attend deliberately to the cued location on each trial (the "attend" group), and a second group was explicitly told that the cue was uninformative and should therefore be ignored (the "ignore" group).

According to the automaticity hypothesis, the central cue should affect the distribution of attention only when subjects deliberately adopt a strategy to use the cue, but the peripheral cue should affect the distribution of attention whether subjects choose to use it or not. That is exactly what happened. In the "attend" group, there was a substantial effect of cue validity for both central and peripheral cues—although the magnitude of the difference between invalid cue RT and valid cue RT was larger for peripheral cues (95 ms) than for central cues (61 ms). This shows that subjects can choose to use either type of cue if they wish. In the "ignore" group, however, the effect of cue validity was very different for central and peripheral cues. There was no difference in response time for valid and invalid central cues (2 ms *slower* for valid central cues), but a substantial validity effect for peripheral cues (98 ms). These results showed that while central cues are effective only when subjects adopt a strategy to use the cues, peripheral cues capture attention even when subjects deliberately attempt to ignore the cues. Experiments 1 and 3 of Jonides (1981) provide converging evidence for the automaticity of peripheral cues.

A decade later, Remington, Johnston, and Yantis (1992) set out to define the limits of the conclusions drawn by Jonides (1981). They asked how effectively such peripheral display changes capture attention when there is a substantial incentive actively to ignore the cue. In the critical conditions of their experiments, the cue *never* appeared in the target location, and this negative contingency was emphasized in the instructions. Subjects were to identify a target character appearing in one of four locations arranged in a diamond configuration above, below, and to the left and right of fixation. On each trial a "cue" appeared around one of the four possible target locations (here I use quotation marks around the word cue to remind the reader that the peripheral stimulus did not always provide information about the upcoming target location and in those cases would not be considered a cue in standard usage). The cue consisted of a highly salient array of four small crosses surrounding the potential target location. In *SAME* blocks, the cue always indicated the target location, and subjects were encouraged to use the cue to direct attention to the indicated location; in *DIF* blocks, the "cue" never indicated the target location (in fact, the cue told subjects that the target was certain to appear in one of the three *other* locations), and subjects were exhorted to ignore the cue and attend elsewhere if possible. In addition, there were blocks of trials in which all locations were cued (*ALL* blocks) and blocks in which none were cued (*NONE* blocks); in both of these conditions there was no information about where the target might appear.

This manipulation tests a stronger version of the automaticity hypothesis than that tested by Jonides (1981): here, subjects were actively attempting to suppress the cued location; in Jonides (1981), the target could sometimes appear in the cued location, so although subjects may not have been actively using the cues, they were probably not actively attempting to suppress them, either. Remington et al. (1992) found that even when the cues were explicitly to be ignored, subjects were unable to do so. Response time in DIF blocks were always slower than in the SAME, NONE, and ALL blocks; this is what one would expect if attention was drawn to the cue even in the DIF blocks when the cue was explicitly to be ignored.

Several studies have pursued the difference between attentional capture by cues adjacent to the object to be attended versus deliberate attentional deployment based on cues that indicate a location other than themselves; one distinguishing characteristic of these two types of cuing is the timecourse with which attention is deployed. It has generally been found that peripheral cues tend to draw attention rapidly, but release it rather quickly; in contrast, deliberate attentional deployment based on central or indirect cues takes longer to occur but can persist for several hundred ms. For example, Müller and Rabbitt (1989) asked subjects to identify and localize a briefly flashed target letter. Preceding the appearance of the target at various points in time, one of two cuing events occurred: an arrowhead appeared at fixation pointing toward the target's location, or a box surrounding the target location brightened briefly. Like Jonides (1981), Müller and Rabbitt (1989) found that (compared to the central arrowhead) the peripheral cue was resistant to suppression even when it was known to be irrelevant. Furthermore, they found that the timecourse of attentional deployment for the two cues was also different: RT to peripherally cued targets reached a minimum at relatively short cuing intervals (i.e. the interval from the onset of the cue to the onset of the target), and then increased again; in contrast, RT to centrally cued targets improved more slowly with cuing interval, reaching an asymptote later and remaining at the faster level for some time.

An interesting violation of this pattern was reported by Warner, Juola, and Koshino (1990). Subjects practiced a cued letter identification task for many sessions. Early in practice, the standard pattern was observed: identification of targets in uncued location was slower than targets in cued locations, even if the uncued location was the likely location of the target. However, after about 4500 trials, subjects were able to direct attention to a location opposite the cue very efficiently. Somehow practice was able to overcome the natural tendency for attention to be captured by a peripheral cue.

The distinction between attention to a peripheral cue and a central indicator has been reported by other labs using somewhat different paradigms. Nakayama and Mackeben (1989) required subjects to search for a complex target in an array of as many as 64 elements. The array was presented briefly and then masked. At various moments in time before array onset, a square appeared at the location that would eventually contain the target. Nakayama and Mackeben recorded detection accuracy as a function of the cuing interval and found that accuracy started at a level just above chance for very short cuing intervals, then increased to a peak at cuing intervals of approximately 150 ms, and then fell to an asymptotic but well above chance level. A similar pattern of results was reported by Cheal and Lyon (1991).

The results of all three of these studies (Cheal & Lyon, 1991; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989) led the authors to conclude that there exist two separate attentional mechanisms. One mechanism responds to peripheral events, draws attention automatically and rapidly, and then dissipates rapidly. The second mechanism is voluntary, requires effort, and has a slower timecourse. Overall performance in the experiments is thought to reflect the superposition of both mechanisms.

Not all investigators subscribe to this view. However, there is good evidence that attentional deployment based on central (symbolic, indirect) cues is different in potency and timecourse than attentional capture by peripheral (direct) cues. At the end of the chapter, I will return to this issue in a discussion of the interaction between top-down and bottom-up control of attention.

Attentional Capture and Visual Salience

Certain visual stimuli are subjectively salient. The clearest examples are objects that differ substantially from their surround in some simple visual feature. For example, a red element will stand out from a background of uniformly green elements. It is not enough that the element differs from the background elements, however: a unique red element will not be subjectively salient if the background consists of multicolored elements, even if none of them is red. So salience requires two conditions: a stimulus that differs from its immediate surround in some dimension, and a surround that is reasonably homogeneous in that dimension (Duncan & Humphreys, 1989). Stimuli that satisfy these criteria are termed *feature singletons*.

Singletons are easy to find in visual search. For example, Egeth, Jonides, and Wall (1972) found that visual search for a *4* in a background of *C*s or vice-versa was very efficient: RT to find the target did not increase as the number of nontargets in the array was increased. Indeed, this is one of the central facts to be explained by theories of visual search (Treisman & Gelade, 1980; Wolfe, 1984; see Wolfe, this volume, for further details).

The efficiency of visual search for a feature singleton led to the assumption that such stimuli capture attention. One manifestation of this assumption is the terminology that is sometimes applied to feature search: feature singletons are said to "pop out" of the display. One must keep in mind, however, that in most feature search tasks, the subject is actively seeking the feature in question; one cannot conclude that a feature for which the observer is deliberately searching captures attention because of the likely involvement of top-down strategies and mechanisms. In this section, several studies are reviewed that cast light on the circumstances under which feature singletons do capture attention. The studies will be divided into two types: those suggesting that feature singletons typically do not capture attention when they are known to be irrelevant to the search task, and those suggesting that when one type of feature singleton (say, a color singleton) is relevant to the search task, then another, irrelevant, singleton (say an orientation singleton) may capture attention despite intentions to the contrary.

Feature Singletons Do Not Always Capture Attention. Only a handful of studies has been carried out to test directly whether feature singletons capture attention when they are completely irrelevant to the task at hand. However, the results have been clear and consistent. Jonides and Yantis (1988) had subjects search through an array of letters for a prespecified target letter; the total number of letters in each display was varied from trial to trial. Each trial contained one letter that differed from all the rest in some way. For example, in one condition, one of the letters was red and the rest were green (and for half the subjects the reverse was true). In another condition, one letter was bright and the rest were dim. In all cases, the position of the unique element was uncorrelated with the position of the target. For example, during trials in which there were seven letters, the target was the bright element on one-seventh of the trials and it was one of the dim elements on the remaining six-sevenths of the trials. The experimental design thereby included no inducement to deliberately attend to the unique element.

The critical observation was whether RT depended on whether the target happened to be the unique element or not. If feature singletons capture attention, then they should do so even when they are uninformative about the location of the upcoming target. (The logic is the same as that used by Jonides, 1981, when he found that peripheral cues capture attention even when they are known to be uninformative about the target's location.) Jonides and Yantis (1988) found that feature singletons, unlike peripheral cues, do not capture attention when they are irrelevant. RT to unique (e.g. bright) targets did not differ from RT to the nonunique (e.g. dim) targets. Hillstrom and Yantis (1994) performed a similar experiment in which the unique feature was motion. Subjects searched for a target letter among nontargets; one letter in each display exhibited motion (they used five different forms of motion, including oscillation, moving internal texture, and small moving elements revolving around the critical letter). Hillstrom and Yantis (1994) first ensured that the features they used were indeed salient by running a version of the visual search experiment in which the moving element was *always* the target, and subjects were encouraged to guide attention to the moving element to speed search. In this case, all five types of motion yielded very efficient search with RTs that did not increase with the number of letters in the array. They then ran a version of the experiment in which the location of the moving element was uncorrelated with the target; in this case, the time required to find the target was no greater when it happened to be the moving element than when it was one of the stationary elements.

Yantis and Egeth (1994) recently reported a similar result using simpler stimuli (oriented colored bars). They extended the findings of Jonides and Yantis (1988) and Hillstrom and Yantis (1994) by showing that response times to singleton targets are highly sensitive to the informativeness of the singleton. The correlation between the location of the singleton element and the target varied for several different groups of subjects. In addition to a group with no correlation and a group with a perfect correction (the results for which were the same as in the studies summarized earlier), they included groups with three intermediate degrees of correlation. The visual search function for singleton targets reflected decreasing influence of the number of nontargets as the correlation between the locations of the singleton and target letters increased (and hence, as the informativeness of the singleton increased). This suggests that subjects were able to use the contingencies contained within an experimental design to guide attention deliberately and in some sense optimally, even with salient feature singletons as the carrier of that information.

Folk and Annette (1994) speculated that the reason Jonides and Yantis (1988) did not observe attentional capture by color or brightness singletons was that the feature density in their displays was too low, leading to reduced feature contrast (the same objection could also be leveled at the studies of Hillstrom & Yantis, 1994, and Yantis & Egeth, 1994.) To test this idea, Folk and Annette performed an experiment similar to those of Jonides and Yantis (1988) but they increased the feature density by including between the letters in the search array randomly positioned dots whose color matched the nonunique color in the display. In this way, the uniquely colored element was much closer to elements with a contrasting color, yielding a sharper color contrast than in the earlier studies. Folk and Annette found, however, that this did not change the results: irrelevant color singletons still did not capture attention. Together, the results described in this section provide strong evidence that feature singletons do not capture attention in a purely bottom-up fashion.

Attentional Control in Singleton Search. A somewhat different story about the attentional effects of feature singletons emerges when observers are actively searching for a singleton. One of the first studies to address this issue directly was reported by Pashler (1988). Subjects were required to search for a circle target in an array of tilted lines (or vice-versa). On any given trial, they did not know whether the target would be a circle in a background of tilted lines or a tilted line in a background of circles. The subjects' primary task, then, was to find and report the location of the form singleton. Two of the elements in each display were colored red, and the rest were green; subjects were informed that the color singletons were irrelevant to their task and should be ignored (the form singletons were never red). Pashler found that localization accuracy for the form singleton decreased in the presence of the two irrelevant color singletons. Evidently Pashler's subjects could not ignore the color singletons when searching for form singletons.

A corroborating result was reported by Theeuwes (1992). Observers were required to report the orientation of a line segment (vertical or horizontal) that was located within one of several colored shapes (red and green diamonds and circles) arranged on the circumference of an imaginary circle centered at fixation (Fig. 6.2, top). The target element was always contained within a form that differed from all the other forms in shape or color. Within a given block of trials, for example, the target would always appear within a circle; all the other forms in each trial were diamonds. This induced subjects to direct attention to the shape singleton on each trial and then make a horizontal/vertical judgment about the line contained therein. On some trials, all the forms had the same color (e.g. green; represented by the solid lines in Fig. 6.2); on other trials, one form had a unique color (e.g. red; represented by the dashed lines in Fig. 6.2). The color singleton, when present, was always in a location that differed from the location of the shape singleton. Color variation was known to be irrelevant to the task and subjects were strongly encouraged to ignore the color singleton if it occurred. Theeuwes (1992) found that orientation responses were always substantially slower in the presence of the color singleton than in its absence (Fig. 6.2, bottom), strongly suggesting that subjects could not ignore the color singleton in this task. (It is worth noting that this finding was not symmetrical: responses to color-defined targets were not slowed by the presence of a shape singleton, a finding Theeuwes attributed to differing degrees of salience in the shape and color domains.) Several other studies have yielded similar results and conclusions (e.g. Theeuwes, 1991a; Folk, Remington, & Johnston, 1992, Experiment 4).

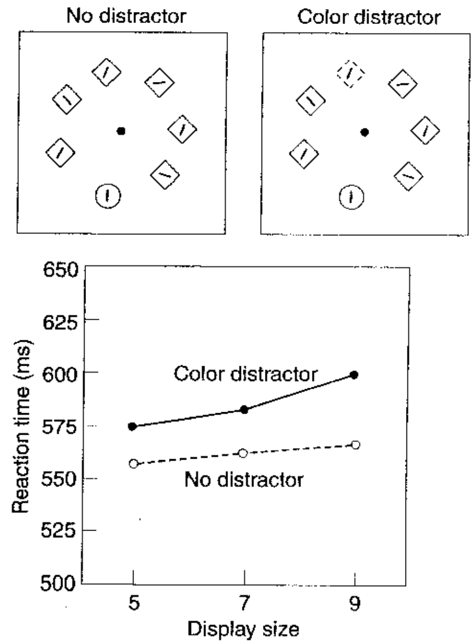

FIG. 6.2. Stimuli and data from Theeuwes (1992) (adapted with permission). Top: A vertical or horizontal bar appeared within a green circle (the other bars were all oblique). On the left, all the diamonds are green (solid lines); on the right, one of the diamonds is red (dashed lines). Subjects were to report the orientation of the bar in the circle. Bottom: The presence of the irrelevant color singleton slowed responses to the bar. Reprinted with the kind permission of the Psychonomic Society, Inc.

On the face of it, the results described here, suggesting that singletons do capture attention and cannot be ignored, are at odds with those described in the last section suggesting that singletons can be ignored easily. How can this apparent discrepancy be explained? The critical difference between the experiments described in the last section and those in this section is this: in all the former studies, singletons were completely irrelevant to the task (the targets were difficult-to-discriminate stimuli that required scrutiny to find), while in the latter studies, the targets were defined as feature singletons. Bacon and Egeth (1994) have argued persuasively that subjects in experiments like those of Theeuwes (1992) and Pashler (1988) adopt a strategy they call *singleton-detection mode*. This strategy relies on a feature-contrast detector to guide attention (see Nothdurft, 1993, for relevant evidence). The feature-contrast detector provides information about the locations in the display with high feature contrast (i.e. locations in which neighboring elements differ in some dimension like color) but it does not provide information about what dimension is producing the contrast. In singleton-detection mode, subjects direct attention to the most salient feature contrast in the display; this strategy is presumed to be very efficient when the target is a singleton. However, when there are one or more other, irrelevant, singletons in the display, attention may sometimes be directed to the wrong location.[6]

To test this idea, Bacon and Egeth (1994) performed several experiments modeled after those of Theeuwes (1992). They first replicated the findings reported by Theeuwes (1992). They then changed the design slightly so that singleton-detection mode would no longer be an effective strategy. In one experiment, for example, they did this by including three instances of the target feature in the display; now the target was not a singleton (i.e. its feature contrast with the rest of the display was low), so singleton-detection mode was inefficient. In this case, the presence of a potentially distracting singleton in another dimension had no discernable effect. Bacon and Egeth concluded that when subjects adopt a different strategy, irrelevant singletons no longer capture attention.

Folk et al. (1992) proposed a framework for conceptualizing the interaction between a subject's strategies and the extent to which stimulus features capture attention. According to their proposal, subjects adopt an *attentional control setting* that determines what stimulus features will control the deployment of attention in any given task. Bacon and Egeth's (1994) singleton detection mode might be considered an instance of an attentional control setting. The idea is that when an observer adopts a deliberate state of attentional readiness for some feature or feature set, then stimuli that are consistent with that feature will tend to receive attention, and stimuli that are not consistent with it will not.

Folk et al. (1992) carried out several experiments to test their idea. The experiments required subjects to identify a target that was defined as possessing a particular feature (e.g. the target might be the only red element in an array of white elements, or the target might be the only element in the display with an abrupt onset). Shortly before the target appeared, a distracting "cue" was presented at a different location.[7] The cue either matched the target feature (if the target was uniquely red, so was the cue) or it did not (if the target was uniquely red, the cue might be a unique abrupt onset). Folk et al. found that the cue disrupted performance (i.e. slowed responses) only when it matched the target's defining feature (and hence, by assumption, the subject's attentional control setting). When the cue did not match the attentional control setting, it did not disrupt performance.

Folk et al. (1992) also found that the disruptive effects of a cue extended to feature values other than ones that were identical to the target's defining feature. For example, if the target was defined as uniquely red, than a uniquely green cue also captured attention. This finding corroborates and extends the finding of Theeuwes (1992) described earlier, and led Folk et al. to speculate that there might be categories of features (e.g. all color singletons) that could serve as a joint attentional control setting, and any singleton within that category would control the deployment of attention. This is a refinement of Bacon and Egeth's (1994) singleton-detection mode according to which any singleton will capture attention with a potency that is proportional to the salience of that feature.

The attentional control setting idea of Folk et al. (1992) can account for an impressive range of results. Some of the specific claims made by Folk et al. (1992) have been contested, however. For example, Folk et al. argued (see Folk, Remington, & Wright, 1994, for additional evidence) that there is a basic distinction between static discontinuities (e.g. shape and color singletons) and dynamic discontinuities (e.g. motion and onset singletons). Theeuwes (1994) reported experiments using a visual search task without spatial cues in which search for a target defined by a static discontinuity was disrupted by the presence of a dynamic discontinuity and vice versa. This finding casts doubt on the static/dynamic distinction proposed by Folk et al. (1992).

Another contested claim made by Folk et al. is that all instances of attentional capture are mediated by an attentional control setting. Yantis (1993b) objected that attentional capture by abrupt visual onset occurs even without a corresponding state of attentional readiness (see the next section for details). Folk et al. (1993) responded by speculating that, in the absence of a simple attentional set, there may exist a "default" attentional control

---

[6] Of course, this raises the question of why subjects continue to use singleton detection mode when irrelevant singletons may appear. It is possible that search is less effortful in this case even though attention may sometimes be misdirected.

[7] As before, I enclose the word "cue" in quotation marks because the term is not used by Folk et al. in the usual way. The "cue" serves only a distracting function (and subjects were clearly informed of this).

setting for abrupt onset, and the default setting may control the deployment of attention in certain tasks.

These objections notwithstanding, Folk et al. (1992) have contributed to an understanding of attentional control by offering a framework for characterizing the interaction between deliberate attentional strategies and nondeliberate attentional capture. I will return to this important issue at the end of the chapter.

Attentional Capture by Abrupt Visual Onsets

In a 1976 article, Breitmeyer and Ganz discussed the role of transient and sustained visual channels in visual pattern masking and information processing. Late in the article, they speculated that one role for the transient channels (i.e. visual channels specialized for the detection of abrupt luminance change over time) might be to direct attention to regions or events in the image that exhibit change; they argued that such events are likely to be significant for behaving organisms because they may require rapid identification and response. Shortly thereafter, Todd and Van Gelder (1979) tested a version of this hypothesis by measuring eye movement latencies to visual stimuli that either did or did not contain abrupt visual onsets. They found that latencies to stimuli with abrupt onsets were faster than to those without abrupt onset, and argued that the speculation of Breitmeyer and Ganz (1976) was consistent with this finding.

Abrupt Visual Onsets Capture Attention. Yantis and Jonides (1984) extended this idea to explore a specifically attentional version of the hypothesis. They used a visual search task to ask whether an abrupt onset stimulus captures attention automatically. The logic was similar to that used by Jonides (1981): to what extent will an abrupt onset stimulus enjoy an advantage in visual search even when there is no incentive for subjects to deliberately attend to it?

The stimuli used by Yantis and Jonides (1984) are illustrated in Fig. 6.3 (top). Each trial began with the presentation of the target letter for that trial (not shown), followed by a set of figure-eight placeholders similar to the seven-segment characters in a digital clock. The placeholders were present for one second, and were replaced by an array of letters. The letters were formed from a subset of the seven segments. All but one of the letters appeared in locations that previously were occupied by figure-eight placeholders; two line-segments were removed to reveal letters in those locations. These letters were termed *no-onset* stimuli. One letter appeared in a previously blank location; this letter was called an *onset* stimulus.

Subjects were required to press one button if the target letter was present in the array, and another button if it was absent. Display sizes of two and

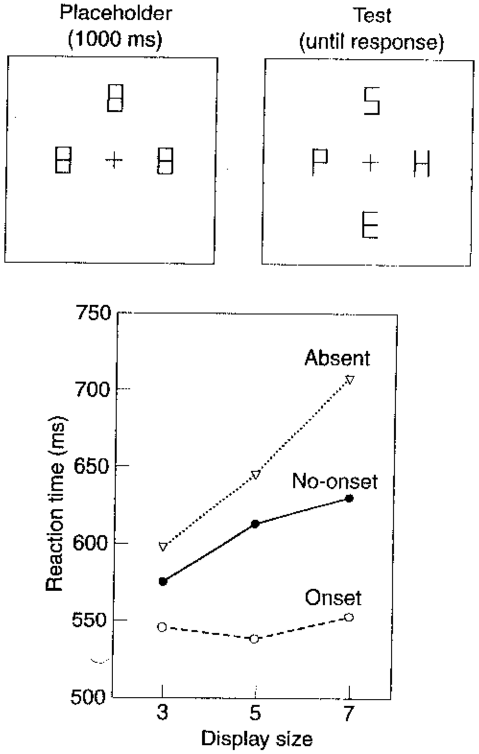

FIG. 6.3. Stimuli and data from Yantis and Jonides (1984; Jonides & Yantis, 1988). Top: An array of figure-eight placeholders appears for 1000 ms, and is followed by an array of letters. One letter appears in a previously blank location (the onset letter), while the remaining letters appeared in locations previously occupied by figure-eight placeholders. These no-onset letters are revealed by the removal of camouflaging line-segments. Bottom: Response times for trials in which the target is the onset letter, one of the no-onset letters, or is absent. Data are from Jonides and Yantis (1988): (Adapted with permission). Reprinted with the kind permission of the Psychonomic Society, Inc.

four letters were used in the experiment. As Wolfe (this volume) states, when the target–distractor discrimination is difficult, then response time typically increases with display size, reflecting an attentionally demanding search. The letters in the experiments of Yantis and Jonides (1984) were designed to be difficult to discriminate, providing a context within which attentional capture might be observed.

A critical feature of the experimental design was that the onset letter was no more likely to be the target than was any other letter in the display. In particular, the target happened to coincide with the onset letter only randomly (i.e. on $1/d$ of the trials, where d is display size). This design minimized subjects' incentive to attend deliberately to the onset location.[8] Thus it was possible to assess the extent to which the onset letter captured attention in a purely stimulus-driven fashion.

Mean RT to correctly detect the target letter for the various conditions of the experiment is shown in the bottom of Fig. 6.3. Four relevant observations can be made. First, when the target was an abrupt-onset stimulus, RT did not increase with display size. Second, when the target was one of the no-onset letters, RT increased significantly with display size. Third, when the target was absent, RT increased even more steeply with display size. Finally, the slope of the target-absent function is approximately twice that of the target-present function for no-onset targets. The last three observations are consistent with a serial self-terminating search strategy for the no-onset letters according to which the letters are scanned one at a time and a response is made when the target is found or after the display is scanned exhaustively and no target is found. The flat display-size function for onset targets strongly suggests that the onset stimulus was identified first during the search, and a response was made after that single identification if the target was the onset letter. This is the prediction of attentional capture by abrupt onset.

New Perceptual Objects Capture Attention. The speculation by Breitmeyer and Ganz (1976) that motivated the experiments of Yantis and Jonides (1984) was that any stimulus that strongly activated the transient visual channels would capture attention. According to this account, the presentation of a luminance increment is critical in drawing attention; this is called the *luminance-increment account*. Yantis and Hillstrom (1994) considered an alternative account for attentional capture by abrupt onsets, motivated by recent object-based theories of visual selection. According to object-based theories, the representational basis for visual selection is not a specific region of space, but a perceptual object (Duncan, 1984; Kahneman, Treisman, & Gibbs, 1992; Kanwisher & Driver, 1992; see also Wolfe, this volume). According to the *new-object account*, the appearance of a new perceptual object requires the creation of an object representation, and this in turn triggers a shift of attention to the new object.

The luminance-increment account and the new-object accounts are both consistent with the results of Yantis and Jonides (1984) because the abrupt onset stimuli were new objects that exhibited luminance increments. In order to distinguish between the two hypotheses, Yantis and Hillstrom (1994) performed experiments that were very similar in design to those of Yantis and Jonides (1984) except that they used objects that were defined by equiluminant discontinuities in motion, depth, or texture. For example, the motion-defined letters were constructed as follows: the screen was filled with randomly placed dots, and windows were defined in the shape of the desired letters (e.g. figure-eights). The dots within the windows moved at a constant speed to the left, and the background dots remained stationary. The contours of the letters were thus defined by a motion discontinuity, but there was no difference in the mean luminance of the letters and the background. A trial started with figure-eight placeholders for one second followed by an array of letters, one appearing in a background location and the others revealed by removing subsets of the line segments making up the figure-eights. Thus when a new object appeared, there was no localized change in the luminance at that location.

The luminance-increment account predicts that such a stimulus will not capture attention, whereas the new-object account predicts that it will. The experiments revealed that new objects do capture attention even without a luminance increment. Yantis and Hillstrom (1994) concluded that a luminance increment is not necessary to capture visual attention.

Overall, the experiments we have carried out lead us to conclude that the appearance of a new perceptual object is an important perceptual event that has significant consequences for the deployment of attention. It is important to emphasize that new objects do not have absolute control over attention (e.g. Koshino, Warner, & Juola, 1992; Theeuwes, 1991b, 1995; Yantis & Jonides, 1990). Nevertheless, the visual system appears to be predisposed to attend to objects that require the creation of a new perceptual object representation.

INTERACTION OF GOAL-DRIVEN AND STIMULUS-DRIVEN ATTENTION

The experiments reviewed in this chapter reveal some of the properties that characterize both goal-driven and stimulus-driven attentional control. Attention can be directed to locations in space "by a conscious and voluntary effort," as Helmholtz suggested. It can also be captured by abrupt onset and other stimulus events. Perhaps more importantly, the experiments show that attentional control results from an interaction between the observer's intentions and the properties of the image.

A complete understanding of control requires a specification of the nature of that interaction. We have touched on several varieties of interaction in this chapter. One is exemplified by the experiments of Pashler (1988), Theeuwes (1992), Folk et al. (1992), and Bacon and Egeth (1994). When an observer is searching for a featural singleton, then even to-be-

---

[8] Indeed, as shown earlier, when a highly salient stimulus (e.g. a red element among green ones) is known by subjects to be uninformative about target location, it is routinely ignored by subjects (e.g. Jonides & Yantis, 1988; Yantis & Egeth, 1994). This provides converging evidence that this experimental design effectively eliminates any top-down intention to direct attention to the unique (in this case, onset) location.

ignored singletons will capture attention. This suggests that subjects adopt a perceptual set to attend to any feature discontinuity (Nothdurft, 1993). However, when an observer searches for a more complex object (e.g. a rotated T in a background of rotated Ls), then irrelevant featural singletons will not capture attention. When the target of search cannot be localized on the basis of contrast in a single feature, then subjects do not enter singleton detection mode.

When an observer directs attention to a spatial location in advance of a display, then visual events that would otherwise capture attention will generally fail to do so. Evidence for this fact was provided by Yantis and Jonides (1990), Koshino et al. (1992), and Theeuwes (1991b, 1995). In these studies, observers were cued to attend to a location that was likely to contain a target stimulus; the appearance of an abrupt onset stimulus in another (to-be-ignored) location did not capture attention. However, that same stimulus did capture attention (and thereby affected performance) when attention could not be focused elsewhere in advance.

Stimulus-driven attentional control is both faster and more potent than goal-driven attentional control (e.g. Jonides, 1981; Nakayama & Mackeben, 1989). A likely reason for this difference is the necessary translation or decision process that is required in goal-driven control: if attention is to be voluntarily directed to a location, then the cue or instruction indicating the to-be-attended location must be perceived and decoded to determine which location it indicates. Sometimes the decoding process is straightforward (e.g. an arrowhead indicating a location) but sometimes it is more symbolic (e.g. a tone or a numeral indicating a location). When the stimulus itself captures attention, however, no such translation is required; attention is directly and immediately deployed.

The interactions between goal-driven and stimulus-driven aspects of attentional control are well characterized by the notion of attentional control settings as proposed by Folk et al. (1992; Folk et al., 1994). It is safe to assume that any alert observer viewing a natural scene has a complex constellation of goals and expectations about what they are about to see. When searching for a specific object, the attentional set is likely to include properties of the desired object. But even when walking down the street with no particular search goal in mind, the observer will seek to avoid obstacles and to walk toward his or her destination. He or she may have a mild interest in, say, architecture. In addition, there appear to be some "default" or "hard-wired" control settings, such as the predisposition to attend to new perceptual objects. All of these internal aspects of the observer will contribute to a multidimensional attentional control setting that will in part determine what aspects of the scene are selected for identification and possible memory storage. For example, image features whose motion trajectories are likely to intersect with the observer's own will receive attention, as will (perhaps) the gargoyles on the building across the street and the dog that appears from behind a mailbox. The deployment of attention in an image is determined by an interaction between the properties of the image and the observer's set of attentional goals.

ACKNOWLEDGEMENT

Preparation of this chapter was supported by a grant from the National Institute of Mental Health (R01-MH43924).

REFERENCES

Averbach, E., & Coriell, A.S. (1961). Short-term memory in vision. *Bell Systems Technical Journal*, *40*, 309–328.

Bacon, W.F., & Egeth, H.E. (1994). Overriding stimulus-driven attentional capture. *Perception and Psychophysics*, *55*, 485–496.

Bashinski, H.S., & Bacharach, V.R. (1980). Enhancement of perceptual selectivity attention to spatial locations. *Perception and Psychophysics*, *28*, 241–248.

Bertelson, P. (1967). The time course of preparation. *Quarterly Journal of Experimental Psychology*, *19*, 272–279.

Breitmeyer, B.G., & Ganz, L. (1976) Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychological Review*, *83*, 1–36.

Broadbent, D.E. (1958) *Perception and communication*. Oxford: Pergamon Press.

Cheal, M.L., & Lyon, R.D. (1991). Central and peripheral precuing of forced-choice discrimination. *Quarterly Journal of Experimental Psychology*, *43*A, 859–880.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception and Psychophysics*, *27*, 183–228.

Deutsch, J.A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*, 80–90.

Downing, C.J., & Pinker, S. (1985) The spatial structure of visual attention. In M. Posner & O. Martin (Eds.), *Attention and performance XI* (pp.171–187). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*, 501–517.

Duncan, J. (1985). Visual search and visual attention. In M.I. Posner & O.S. Marin (Eds.), *Attention and performance XI* (pp.85–106). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Duncan, J., & Humphreys, G.W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458.

Egeth, H.E., & Dagenbach, D. (1991). Parallel versus serial processing in visual search: Further evidence from subadditive effects of visual quality. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 550–559.

Egeth, H., Jonides, J., & Wall, S. (1972). Parallel processing of multielement displays. *Cognitive Psychology*, *3*, 674–698.

Egly, R., & Homa, D. (1984). Sensitization of the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 778–793.

Engel, F.L. (1971). Visual conspicuity, directed attention and retinal locus. *Vision Research*, *11*, 563–576.

Eriksen, C.W., & Hoffman, J.E. (1973). The extent of processing noise elements during selective encoding from visual displays. *Perception and Psychophysics*, *14*, 155–160.

Eriksen, C.W., & Murphy, T.D. (1987). Movement of attentional focus across the visual field: A critical look at the evidence. *Perception and Psychophysics*, *42*, 299–305.

Eriksen, C.W., & St. James, J.D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics*, *40*, 225–240.

Folk, C.L., & Annett, S. (1994). Do locally defined feature discontinuities capture attention? *Perception and Psychophysics*, *56*, 277–287.

Folk, C.L., Remington, R.W., & Johnston, J.C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1030–1044.

Folk, C.L., Remington, R.W., & Johnston, J.C. (1993). Contingent attentional capture: A reply to Yantis (1993). *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 682–685.

Folk, C.L., Remington, R.W., & Wright, J.H. (1994). The structure of attentional control: Contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 317–329.

Gawryszewski, L.D.G., Riggio, L., Rizzolatti, G., & Umiltà, C. (1987). Movements of attention in the three spatial dimensions and the meaning of "neutral" cues. *Neuropsychologia*, *25*, 19–29.

Ghirardelli, T.G., & Folk, C.L. (1996). Spatial cuing in a stereoscopic display: Evidence for a "depth-blind" attentional spotlight. *Psychonomic Bulletin and Review*, *3*, 81–86.

Grindley, G.C., & Townsend, V. (1968). Voluntary attention in peripheral vision and its effects on acuity and differential thresholds. *Quarterly Journal of Experimental Psychology*, *20*, 11–19.

Helmholtz, H. von (1925). *Treatise on physiological optics* (3rd edn., Vol. III, J.P.C. Southhall, Ed. & Trans.). Washington, DC: The Optical Society of America. [Original work published 1866.]

Hillstrom, A.P., & Yantis, S. (1994). Visual motion and attentional capture. *Perception and Psychophysics*, *55*, 399–411.

Hoffman, J.E., & Mueller, S. (1994, November). *An in depth look at visual attention.* Paper presented at the 35th Annual Meeting of the Psychonomic Society, St. Louis.

Hoffman, J.E., & Nelson, B. (1981). Spatial selectivity in visual search. *Perception and Psychophysics*, *30*, 283–290.

Iavecchia, H.P., & Folk, C.L. (1995). Shifting visual attention in sterographic displays: A minecourse analysis. *Human Factors*, *36*, 606–618.

James, W. (1890). *The principles of psychology* (Vol. 1). New York: Henry Holt & Co.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J.B. Long & A.D. Baddeley (Eds.), *Attention and performance IX* (pp.187–203). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception and Psychophysics*, *43*, 346–354.

Kahneman, D., Treisman, A., & Gibbs, B.J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175–219.

Kanwisher, N., & Driver, J. (1992). Objects, attributes, and visual attention: Which, what and where. *Current Directions in Psychological Science*, *1*, 26–31.

Koopman, B.O. (1957). The theory of search: III. The optimum distribution of searching effort. *Operations Research*, *5*, 613–626.

Koshino, H., Warner, C.B., & Juola, J.F. (1992). Relative effectiveness of central, peripheral, and abrupt-onset cues in visual search. *Quarterly Journal of Experimental Psychology*, *45A*, 609–631.

Kröse, B.J.A., & Julesz, B. (1989). The control and speed of shifts of attention. *Vision Research*, *29*, 1607–1619.

Kwak, H.-W., Dagenbach, D., & Egeth, H. (1991). Further evidence for a time-independent shift of the focus of attention. *Perception and Psychophysics*, *49*, 473–480.

LaBerge, D. (1983). The spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 371–379.

LaBerge, D., & Brown, V. (1986). Variations in size of the visual field in which targets are presented: An attentional range effect. *Perception and Psychophysics*, *40*, 188–200.

Mertens, J.J. (1956) Influence of knowledge of target location upon the probability of observation of peripherally observable test flashes. *Journal of the Optical Society of America*, *46*, 1069–1070.

Müller, H.J., & Rabbitt, P.M.A. (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 315–330.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research*, *29*, 1631–1647.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, *86*, 214–255.

Nothdurft, H.C. (1993). Saliency effects across dimensions in visual search. *Vision Research*, *33*, 839–844.

Pashler, H. (1988). Cross-dimensional interaction and texture segregation. *Perception and Psychophysics*, *43*, 307–318.

Posner, M.I. (1978). *Chronometric explorations of mind.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Posner, M.I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.

Posner, M.I., & Boies, S.J. (1971). Components of attention. *Psychological Review*, *78*, 391–408.

Posner, M.I., Snyder, C.R.R., & Davidson, B.J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*, 160–174.

Remington, R.W., Johnston, J.C., & Yantis, S. (1992). Involuntary attentional capture by abrupt onsets. *Perception and Psychophysics*, *51*, 279–290.

Remington, R.W., & Pierce, L. (1984). Moving attention: Evidence for time-invariant shifts of visual selection attention. *Perception and Psychophysics*, *35*, 393–399.

Shaw, M., & Shaw, P. (1977). Optimal allocation of cognitive resources to spatial locations. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 201–211.

Shaw, M.L. (1978). A capacity allocation model for reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 586–598.

Shaw, M.L. (1984). Division of attention among spatial locations: A fundamental difference between detection of letters and detections of luminance increments. In H. Bouma & D.G. Bonwhuis (Eds.), *Attention and performance X* (pp.109–120). Hove, UK: Lawrence Erlbaum Associates Ltd.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127–190.

Shiu, L.-P., & Pashler, H. (1994). Negligible Effects of spatial precuing on identification of single digits. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1037-1054.

Shulman, G.L., Remington, R.W., & McLean, J.P. (1979). Moving attention through visual space. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 522–526.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74*, 1–29.

Sperling, G., & Melchner, M.J. (1978). The attention operating characteristic: Examples from visual search. *Science, 202*, 315–318.
Todd, J.T., & Van Gelder, P. (1979). Implications of a transient-sustained dichotomy for the measurement of human performance. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 625–636.
Theeuwes, J. (1991a). Cross-dimensional perceptual selectivity. *Perception and Psychophysics, 50*, 184–193.
Theeuwes, J. (1991b). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception and Psychophysics, 49*, 83–90.
Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception and Psychophysics, 51*, 599–606.
Theeuwes, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 799–806.
Theeuwes, J. (1995). Temporal and spatial characteristics of preattentive and attentive processing. *Visual Cognition, 2*, 221–233.
Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136.
Tsal, Y. (1983). Movements of attention across the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 523–530.
Van der Heijden, A.H.C. (1992). *Selective attention in vision*. New York: Routledge, Chapman, & Hall.
Van der Heijden, A.H.C., & Eerland, E. (1973). The effects of cuing in a visual signal detection task. *Quarterly Journal of Experimental Psychology, 25*, 496–503.
von Wright, J.M. (1968). Selection in immediate visual memory. *Quarterly Journal of Experimental Psychology, 20*, 62–68.
Warner, C.B., Juola, J.F., & Koshino, H. (1990). Voluntary allocation versus automatic capture of visual attention. *Perception and Psychophysics, 48*, 243–251.
Wolfe, J.M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1*, 202–238.
Yantis, S. (1988). On analog movements of visual attention. *Perception and Psychophysics, 43*, 203–206.
Yantis, S. (1993a). Stimulus-driven attentional capture. *Current Directions in Psychological Science, 2*, 156–161.
Yantis, S. (1993b). Stimulus-driven attentional capture and attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 676–681.
Yantis, S., & Egeth, H.E. (1994). Visual salience and stimulus-driven attentional capture. *Investigative Ophthalmology and Visual Science, 35*, 1619.
Yantis, S., & Hillstrom, A.P. (1994). Stimulus-driven attentional capture: Evidence from equiluminant visual objects. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 95–107.
Yantis, S., & Johnston, J.C. (1990). On the locus of visual selection: Evidence from focused attention tasks. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 135–149.
Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 601–621.
Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 121–134.

CHAPTER SEVEN

Neurophysiology of Selective Attention

Steven J. Luck
University of Iowa, USA

INTRODUCTION

Why Neurophysiology?

At any given moment, the human brain is confronted with a multitude of inputs: sounds from many sources; tactile sensations from the entire surface of the body; visual inputs from a huge array of retinal receptors; a variety of smells and tastes; and internally generated thoughts, emotions, memories, and images. Most of these inputs are irrelevant at any given moment, however, and it is therefore more efficient to focus our limited cognitive processing resources on a subset of the available information and ignore the rest; this is the primary role of selective attention in information processing. Although this selection process is usually conceptualized as a cognitive phenomenon that falls exclusively within the domain of psychology, neuroscientists have also become interested in the topic of selective attention in recent years. Before I begin describing their discoveries, however, I would like to consider why neuroscientists have become interested in attention and why psychologists are becoming interested in neurophysiological studies of attention.

To understand the recent interest of neuroscientists in attention, it is useful first to review some important developments in the areas of sensory anatomy and physiology. During the past 20 years, scientists have discovered over 30 separate areas of primate visual cortex that contribute to different aspects of visual perception, including depth perception, motion