*Annu. Rev. Psychol. 1997. 48:269–97*

# VISUAL ATTENTION: Control, Representation, and Time Course

*Howard E. Egeth and Steven Yantis*

Department of Psychology, The Johns Hopkins University, Baltimore, Maryland 21218

KEY WORDS: attention, cognition, human subjects, perception, psychophysics, reaction time, vision, visual search

## ABSTRACT

Three central problems in the recent literature on visual attention are reviewed. The first concerns the control of attention by top-down (or goal-directed) and bottom-up (or stimulus-driven) processes. The second concerns the representational basis for visual selection, including how much attention can be said to be location- or object-based. Finally, we consider the time course of attention as it is directed to one stimulus after another.

## CONTENTS

## INTRODUCTION

We review the literature on three fundamental aspects of attention that have been the focus of much recent research. The first concerns attentional control, or the extent to which the deployment of attention is a result of the observer's

0084-6570/97/0201-0269$08.00

deliberate state of attentional readiness (called top-down or goal-directed control), or whether attention is captured by certain aspects of the image independently of current perceptual goals (called bottom-up or stimulus-driven control). We emphasize stimulus-driven control and the interaction between the two attentional control modes. The second topic concerns the representational basis for visual selection. In particular, we examine the conditions under which attention may be said to be directed to regions of space, as space-based theories assert, or to preattentively defined perceptual objects, as suggested by object-based theories of attention. Finally, we examine recent evidence concerning the time course of attention, both as it moves through space and as events occurring sequentially in time are selected. In this section, several different estimates of the time scale on which attentional events occur are reviewed.

A single chapter cannot completely cover the broad and active field of attention. Readers interested in pursuing these and other aspects of attention may consult three recent tutorial volumes (Dagenbach & Carr 1994, Kramer et al 1996, Pashler 1996a) or several recent monographs (LaBerge 1995, Pashler 1996b, van der Heijden 1992). In addition, there are several recent *Annual Review* chapters on attention (Desimone & Duncan 1995, Johnston & Dark 1986, Kinchla 1992).

## STIMULUS-DRIVEN AND GOAL-DIRECTED CONTROL OF ATTENTION

When William James (1890) first delineated the varieties of attention over a century ago, one major categorical boundary was the distinction between *passive* and *active* attention. The modern terms are usually *bottom-up* and *top-down* or the less metaphorical *stimulus-driven* and *goal-directed.* The idea is that the deployment of attention may sometimes depend on the properties of the image almost exclusively (e.g. sudden movement in the periphery); other times it may be under strict supervision according to the observer's goals. Mounting evidence has revealed that these two domains of attentional control almost invariably interact. With a few possible exceptions, both the properties of the image and the expectations and goals of the observer determine the attentional consequences of a given perceptual episode. (We consider the time course of attentional control in the last section of this chapter.)

The past 25 years have yielded ample evidence that the distribution of attention can be controlled by the intentions of the observer. Helmholtz (1925, p. 455) first noted this ability in the past century, but it was not until the 1950s that the perceptual consequences of the deliberate deployment of attention were first studied systematically by Mertens (1956). Much of the modern evidence for top-down control has been reviewed previously (e.g. Johnston &

Dark 1986). Significant advances in our understanding of the top-down deployment of attention began with a seminal series of studies by Eriksen and his colleagues (e.g. Eriksen & Hoffman 1972, 1973). Subjects identified a letter indicated by a bar marker and attempted to ignore other letters in the display. The amount of interference caused by the to-be-ignored letters provided a measure of the efficiency and time course of attentional deployment. Studies by Posner and colleagues (e.g. Posner 1980, Posner et al 1980) further examined top-down attentional control.

The evidence concerning the capture of attention (i.e. bottom-up control), unlike that for top-down control, is more recent and is the focus of this section. Two major categories of stimulus properties that could in principle capture attention can be distinguished: stimuli that differ substantially in one or more simple visual attributes (e.g. color, orientation, or motion) from their backgrounds—hereafter called *feature singletons* or simply *singletons*—and abrupt visual onsets. We consider the evidence about each of these categories in turn, and we return to this topic in the section entitled "The Time Course of Attention."

## *Feature Singletons and Attentional Capture*

In considering what stimulus properties might capture attention regardless of (or in spite of) the observer's state of attentional readiness, feature singletons appear to be likely candidates. Feature singletons are judged as subjectively salient, and there is ample evidence that such stimuli can be found efficiently in visual search. For example, Neisser (1967) observed that curved letters could be found easily among straight letters, and using a somewhat different paradigm, Egeth et al (1972) drew a similar conclusion. Treisman & Gelade (1980) also showed, using a visual search paradigm like that of Egeth et al (1972), that various feature singletons could be efficiently detected in visual search. Many examples have been catalogued by Treisman & Gormican (1988; see also Bravo & Nakayama 1992).

These demonstrations provide no direct evidence, however, about whether feature singletons capture attention, because in all the cited cases the stimulus in question was itself the target of search and therefore presumably elicited top-down, deliberate deployment of attention. Therefore, one must design experiments that explicitly dissociate the observer's attentional set from the properties of the stimulus array. Several such studies have been reported, but they have yielded different conclusions. We first review papers that might lead one to conclude that feature singletons do capture attention. We next consider studies that suggest otherwise and end the section with discussion of a paper that provides a possible reconciliation.

SINGLETONS CAPTURE ATTENTION Pashler (1988, Experiment 7) had subjects search for a slash (/) in an array of many *O*s or for an *O* among /s. The identity of the target was not known in advance. On some trials, two of the items in the array were uniquely colored. These were always irrelevant to the task, and subjects were told to ignore them. Pashler found that reaction time (RT) to locate the target shape was prolonged on those trials in which the color singletons appeared despite subjects' intentions to the contrary. Because of these results, one might tentatively conclude that feature singletons do capture attention.

Theeuwes (1991a, 1992) further explored the conditions under which feature singletons controlled the deployment of attention. In his tasks, subjects typically searched for an easy-to-detect target (e.g. a diamond) in an array of distractors (e.g. circles). Each stimulus had inscribed in it a line segment, either horizontal or vertical (in the target) or oblique (in the distractors). To demonstrate successful target acquisition, the subject had to indicate whether the line in the target was horizontal or vertical. Subjects were told that the critical line would always be contained within the shape singleton. On half the trials, all the stimuli had the same color (e.g. red), and on the remaining trials, one of the stimuli—never the target shape—was unique in color (e.g. green). Subjects were instructed to ignore the color variation. Additionally, the total number of elements in the display varied (5, 7, or 9). Reaction time was independent of display size, which suggests parallel processing. With respect to the issue of attentional capture, this experiment is similar in many ways to that of Pashler (1988, Experiment 7). Results were similar as well: RTs were prolonged on those trials containing a color singleton compared with the trials without a color singleton. Theeuwes (1991a) concluded that when subjects search for a target in parallel, which is possible when the target is a feature singleton, then top-down control is not possible and attention is captured even by singletons known to be irrelevant to the task.

Another example of stimulus-driven capture was reported recently by Joseph & Optican (1996). Subjects were required to search for an *L* embedded in an array of *T* stimuli; this target array was flashed briefly and then masked. Subjects were required to report the location (one of four quadrants) in which the *L* appeared. Immediately preceding the target array, a cue array was presented briefly. The cue array consisted of vertical line segments in which a single horizontal segment was embedded. Subjects were correctly informed that the location of the orientation singleton in the cue array was uncorrelated with the location of the upcoming target. (To call an uninformative stimulus a cue may seem problematic; we accede here to what appears to be common usage in the field.) Nevertheless, responses were more accurate when the cue appeared in the location subsequently occupied by the target, which suggested that attention was drawn to the cue even though it was known to be irrelevant to the task.

SINGLETONS DO NOT CAPTURE ATTENTION These preceding studies all suggest that feature singletons, even ones that are known to be task irrelevant, do capture visual attention. However, there is also evidence for the opposite conclusion. Jonides & Yantis (1988) reported that color and brightness singletons do not capture attention. Observers were required to search for a letter in an array of multiple letters. On each trial one letter differed from all the rest in color for some subjects and in brightness for others. Subjects were told that the target would occasionally be the unique element, but only on randomly selected trials. That is, the feature singleton provided no help in solving the primary task, which was to find the target letter. At issue was whether RT to find the target differed when the target did and did not happen to be the unique element. Yantis & Jonides found that it made no difference: Responses were no faster when the target was the singleton than when it was not. This result was subsequently corroborated by Theeuwes (1990).

Hillstrom & Yantis (1994) reported that not even visual motion captures attention under all circumstances. They had subjects search for a rotated T among rotated Ls. One of the stimuli on each trial exhibited one of five different types of visual motion. In all cases, the position of the target element was uncorrelated with the position of the moving element. At issue was whether RT differed according to whether the target happened to correspond to the moving element. If motion captures attention, then one would expect more rapid RTs to moving targets than to stationary ones. Hillstrom & Yantis found, however, that RT did not differ for the two conditions.

ATTENTIONAL CONTROL Thus, several studies suggest that singletons do capture attention, whereas several others suggest that they do not. Bacon & Egeth (1994) proposed a reconciliation by suggesting that these conflicting results are manifestations of two different attentional strategies adopted by subjects. Under some circumstances, subjects enter *singleton detection mode,* in which attention is directed to the location in the array exhibiting the largest local feature contrast (for further evidence, see Bravo & Nakayama 1992; Nothdurft 1992, 1993). In singleton detection mode, the location of the greatest contrast can be accessed, but not the identity of the dimension(s) on which the stimuli differ. Thus, when one is searching for, say, a shape singleton, an irrelevant color singleton may “win out” because of its greater local feature contrast. Other stimulus conditions might lead subjects to adopt *feature search mode*, in which attention is directed to locations that match some task-defined visual feature (e.g. “red” or “vertical”). Bacon & Egeth (1994) supported this proposal by showing that capture by a to-be-ignored feature singleton only occurred when the task could be carried out in singleton detection mode. When that strategy was made ineffective, then capture by irrelevant feature singletons did not occur. In one experiment, singleton detection mode was made ineffective by including several identical target shapes, thus ensuring that no one of them could serve as a singleton target. In another experiment, several unique stimuli were present that were nontargets, again ensuring that the target could not be found efficiently simply by looking for a singleton.

## *Abrupt Visual Onsets and Attentional Capture*

A category of stimuli that behaves somewhat differently than most feature singletons is abrupt visual onset. Early studies of whether abrupt onsets capture attention were motivated by the cuing technique introduced by Eriksen and his colleagues (e.g. Eriksen & Hoffman 1972). Jonides (1981) showed that a peripheral attentional cue (i.e. a bar marker presented near the location subsequently occupied by a to-be-identified letter) draws attention "automatically," whereas a central arrowhead cue requires a deliberate shift of attention. The automaticity of the peripheral cue was demonstrated in his Experiment 2 by showing that peripheral cues drew attention whether they were informative about the location of the target or not, while central cues only controlled the deployment of attention when they were informative. Remington et al (1992) found that a peripheral cue captures attention even when it is known *never* to indicate the target location.

Yantis & Jonides (1984) proposed that peripheral cues like the ones employed by Jonides (1981) might capture attention because they have abrupt onsets. The magnocellular visual pathway is known to be quite sensitive to high temporal frequency, and one of its functions might be to signal the location to which attention should be directed (Breitmeyer & Ganz 1976 originally proposed this idea). Yantis & Jonides (1984) pursued this idea with a visual search task in which observers searched for a prespecified target letter embedded in an array of nontarget letters. The letters were formed by illuminating a subset of the segments of a figure eight as on a digital alarm clock (borrowing a technique devised by Todd & Van Gelder 1979). Each trial began with an array of complete figure-eight placeholders. These were actually letters that were "camouflaged" with irrelevant line segments. The figure eights remained on the screen for 1 s. The camouflaging line segments were then removed from the figure eights to reveal letters (these were designated "no-onset" letters), and simultaneously a single letter appeared in a previously blank location (the "onset" letter). The target letter was present on half the trials. On the target-present trials, the target was the onset letter on 1/*n* of the trials (where *n* is the total number of letters in the display). Because the target had an abrupt onset only rarely, there was no incentive to deliberately attend to it.

Even though there was no incentive to attend to the abrupt onset letter, Yantis & Jonides (1984) found that RT to find the target when it happened to be the onset letter was fast and did not vary with the number of elements to be searched, whereas RT to find the target when it was one of the no-onset letters was slower and increased linearly with the number of elements to be searched. They concluded that the onset letter captured attention on each trial. If the target happened to be the onset letter, a response was made immediately and no further searching was required, but if the onset letter was not the target, then an attentionally demanding search had to be initiated. Yantis & Jonides (1990) found that this capture, which occurs in the absence of any relevant attentional set, is prevented when subjects are induced to focus attention on a different spatial location in advance of each trial. This finding was corroborated using different approaches by Theeuwes (1991b) and by Koshino et al (1992) and Juola et al (1995).

At least two potential mechanisms could account for attentional capture by abrupt onsets (Yantis & Hillstrom 1994). One, mentioned above, is that the luminance increment activates visual pathways that respond to high temporal frequency, which in turn direct attention to the eliciting object. A second possibility is that the appearance of a perceptual object, which requires the creation of an episodic perceptual representation, elicits a shift of attention. This second possibility might be a “hard-wired” response to the need to rapidly identify new objects entering the visual field. Yantis & Hillstrom (1994) performed a series of experiments that permitted them to determine which of these accounts was correct. They used stimuli that were equiluminant with their backgrounds (e.g. random-dot stereograms in which the letters were composed of dots exhibiting binocular disparity against a zero-disparity background). These displays thus exhibited no change in mean luminance, but they did include the appearance of a new perceptual object. The experiment provided clear support for the new-object account: Attention was captured by new perceptual objects even though they did not exhibit a luminance increment. Hillstrom & Yantis (1994) corroborated this finding by showing that while motion per se does not capture attention when it is task irrelevant (as noted above), when motion segments an object from its background (as when the motion of a moth’s camouflaging wings segment it from a tree’s bark), attention is captured. For a recent debate about the new-objects account, see Gibson (1996a,b) and Yantis & Jonides (1996).

## *Interaction of Goal-Driven and Stimulus-Driven Capture*

The studies reviewed so far in this section provide evidence that under certain circumstances attention is drawn to objects (e.g. feature singletons or abrupt onsets) without a deliberate intent to direct attention there. However, in each case, top-down control plays a role. One example of top-down control is that irrelevant feature singletons capture attention only when subjects enter singleton detection mode, where feature contrast gradients control the distribution of attention. Another example is that attentional capture by abrupt onsets can be prevented or at least modulated by focused attention elsewhere in the display.

Folk et al (1992) proposed a theoretical framework for the interaction between goal-driven attentional control and stimulus-driven attentional capture. They argued that any given perceptual act entails an "attentional control setting," which is part of the explicit or implicit set of perceptual goals held by the observer at that moment. These goals might be a result of instructions provided by an experimenter (e.g. "search for the red vertical bar"), or, more often, by the individual's current plan of action in everyday life (e.g. searching for the car keys). The visual features that are of current interest (e.g. "red" or "vertical") will control the distribution of attention.

They provided evidence for this idea by showing that the deployment of attention depends critically on what the subject is set for. In one experiment, each trial consisted of a fixation display, followed in rapid succession by a cue display and a target display. Each element shown in the target display was either an **x** or an **=**. Two types of target displays were used. Color target displays consisted of three white elements and one red element, and the task was to identify the red element as quickly as possible as either an **x** or an **=**. Onset target displays consisted of only one element, and so the target was characterized as being the only element with an abrupt onset. Again, the task was to identify the target as being an **x** or an **=**. Immediately preceding the target display, a cue display appeared; this could consist of either color cues (in which one location was surrounded by red dots and the other three locations were surrounded by white dots) or onset cues (in which one location was surrounded by suddenly onset white dots and the remaining locations remained blank). Each type of target display was combined with each type of cue display. Cue validity was manipulated between blocks; in one condition the cue was 100% valid (It always indicated the location of the to-be-identified target element), and in another condition it was 100% invalid (It always indicated a nontarget element). Folk et al (1992) found that when the cue and target were of the same type, i.e. both color or both onset, cue validity had a large effect. In particular, subjects could not ignore invalid cues. However, when the cue and target were of different types (e.g. a color cue and an onset target), then the cue had little or no influence on response times. This result is consistent with the idea that the state of attentional readiness adopted by the observer determines what sorts of feature singletons will capture attention.

A similar idea motivated Wolfe's (1994) Guided Search model (see also Cave & Wolfe 1990, Wolfe et al 1989). According to Guided Search, attention is directed to objects serially in order of priority. Attentional priority is determined jointly by two things. One is top-down activation, that is, how closely an object matches the current attentional set. For example, if the subject is searching for a red vertical bar, then all red things and all vertical things will receive higher priority than things that are neither red nor vertical. Things that are both red and vertical will, of course, receive the most activation. The other determinant is bottom-up activation, that is, how much a given object differs from neighboring objects within any given perceptual dimension. For example, a red object surrounded by green objects will have greater bottom-up activation than will a red object surrounded by orange objects. These two sources of activation are combined to produce an "attention map" that determines the order in which objects are visited during visual search.

Both these theories incorporate a principle that William James recognized: The deployment of attention depends jointly on properties of the image and the goals and expectations of the observer.

## THE REPRESENTATIONAL BASIS OF VISUAL SELECTION

In an important 1981 paper, Kahneman & Henik asked, "If attention selects a stimulus, what is the stimulus that it selects?" (p. 183). Until that point—with some notable exceptions—attention was viewed (implicitly or explicitly) as similar to a spotlight directed to regions of space, "illuminating" the objects located there (e.g. Eriksen & Hoffman 1972, Hoffman & Nelson 1981, Posner et al 1980). The evidence consisted primarily of demonstrations that the spatial separation between elements significantly modulated attentional effects. For example, Hoffman & Nelson (1981) required subjects to identify a target letter that appeared in one of four locations in the visual field, and then to identify a secondary shape that was either near the target letter or elsewhere in the display. They found that identification accuracy was much better when the two stimuli were adjacent to each other, which suggested a spatial limitation in dividing attention. Downing & Pinker (1985) cued subjects to attend to one of ten boxes arranged in a horizontal row (five on either side of fixation) in anticipation of a luminance increment in one of the boxes, most often in the cued box. Detection RT was fastest when the target event occurred within the cued box, and it slowed monotonically as the distance between the target and the attended location increased. This strongly suggested a spatial gradient of selective attention.

Kahneman & Henik (1981) urged readers, however, to consider the possibility that attention might be directed not only to spatial locations but also to perceptual objects. The idea is that the raw retinal image provides only a fragmented representation of the scene because of occlusion, yet perceptual experience is coherent and “smooth.” Therefore, some early visual mechanism is required to construct representations of objects. Discovering the principles by which object representations are constructed was a major goal of the Gestalt psychologists. Kahneman & Henik suggested that often the object representations resulting from perceptual organization serve as the representational basis for visual selection.

Because objects occupy locations, experiments designed to provide evidence for object-based accounts must demonstrate that a given finding is due to allocation of attention to a locationally invariant object representation and not to a spatial location. Several strategies have been adopted that accomplish this. Evidence for object-based theories of attention fall into two broad categories. Locations or features in an image can be probed that differ according to their relationship to object structure, but that do not differ in spatial location or separation. Attention may also be directed to moving objects, which by definition involves continuously changing spatial locations. Within each of these two categories, many specific techniques have been devised.

## *Overlap and Grouping*

One of the earliest demonstrations of object-specific attentional benefits was reported by Rock & Guttman (1981) in an experiment showing that subjects can selectively attend to one of two objects appearing in the same spatial location. Observers viewed a sequence of 10 pictures. Each picture consisted of two superimposed outline drawings of novel shapes, one drawn in red ink and one in green. The subject was asked to make an aesthetic judgment of the object drawn in red ink (or, for half the subjects, green ink) and to ignore the object drawn in the other color. The judgment task was merely a cover to induce subjects to process the items selectively. After viewing all the drawings, subjects were given a surprise recognition test. One third of the test items had been attended during the judgment task, one third had been unattended, and one third were new. Subjects were much more likely to report attended items as old than unattended or new items. Their judgments for unattended and new items did not differ. This result shows that attention need not be purely location-based, but that it is possible to selectively attend to one of two spatially coincident perceptual objects.

Duncan (1984) laid out explicitly the distinction between space-based theories and object-based theories of attention and explored the distinction empirically with a perceptual version of the memory task employed by Rock & Guttman. Subjects viewed a display consisting of a rectangle with a tilted line drawn through the middle. Each object could take on two values for each of two attributes: The line could be tilted right or left, and it could be dotted or dashed in texture. The rectangle could be tall or short, and it had a small gap in its contour on the right or left side. The display was flashed briefly and was followed by a mask. Subjects were asked to report two attributes on each trial. On some trials, the two attributes belonged to the same object (e.g. the tilt and texture of the line), while on other trials they belonged to different objects (e.g. the height of the rectangle and the tilt of the line). Responses were more accurate when the attributes belonged to the same object. This was taken as evidence that observers attend to objects as a whole: When judgments had to be made about both objects, a cost was incurred because of the need to shift attention from one object to the other.

Vecera & Farah (1994) verified that Duncan's results reflected an object-based effect (rather than what they termed a spatial "grouped array" effect). They noted that a spatial account predicts that the magnitude of the object-specific benefit should be larger when the two objects are spatially separated than when they are spatially superimposed. They found no evidence of such a pattern. However, when the task required judgments that did not involve accessing the shapes of the objects but required only the detection of a small dot at various locations on the object contours, then only space-based effects were observed. They concluded that simple detection tasks may access a strictly spatial level of representation, while shape discrimination tasks require object-based representations and therefore yield object-based attentional effects.

The studies just reviewed employing two overlapping objects revealed that one can selectively attend to an object occupying the same spatial location as another object, as object-based theories predict. A related technique is to show that when attention is directed to one part of an object, other parts of the object enjoy an attentional benefit, whereas equally distant locations in other objects do not. Several examples of this approach have been reported. Baylis & Driver (1993) showed that judging the relative locations of two "corners" of a complex stimulus was more difficult when they belonged to two objects rather than one. This was the case even when the one-object and two-object displays were physically identical, with instructions determining how many objects were seen in the display. Baylis & Driver found, like Duncan, that judgments about two parts of a single object were made faster than those about parts of two different objects (for further discussion of this procedure, see Baylis 1994, Gibson 1994).

M Behrmann, RS Zemel & MC Mozer (unpublished manuscript) documented a similar object-specific benefit using a perceptual matching task. Subjects were shown a display in which two rectangles, one oriented at +45° and the other at –45°, overlapped such that one rectangle was seen as being in front of the other. Two of the rectangle ends had either two or three "bumps" on them, and subjects were required to report whether the number of bumps was the same or different on the two ends (the other two ends were straight). Of greatest interest was whether the ends to be judged were part of the same perceptual object. For example, the rectangle ends to be judged could be at either end of the partly occluded rectangle, or one could be at one end of the occluding rectangle, and the other could be at one end of the partly occluded rectangle. The main result was that judgments made about two parts of the same object were faster than judgments made about parts of two different objects, even when the object in question was partly occluded. This finding is related in many ways to the result reported by Duncan (1984). In this case, however, the partly occluded objects were fragmented in the image. Perceptual organization mechanisms were required to put the object fragments together into coherent object representations.

Egly et al (1994) had subjects view a display containing two vertically (or, on other trials, horizontally) oriented rectangles presented side by side. One end of one of the rectangles was cued (its local contour was briefly brightened), and after a short delay one end of one of the rectangles was filled in (this was the "target"). Subjects were to press a button when the target appeared (a simple detection task). They were told that the target would appear at the cued location on 80% of the trials (the valid condition), at the other end of the cued object on 10% of the trials (the same-object condition), and at the same end of the uncued object on 10% of the trials (the different-object condition). The latter two locations were equally distant from the cued location, but they differed in their relation to the cued object. The authors found that mean RT in the valid condition was faster than the other two conditions. More revealing was the presence of an object-specific benefit: RTs in the same-object condition were faster than in the different-object condition. This outcome is consistent with an object-based account.

A related study was carried out by Yantis & Moore (1995). They used rectangle pairs like those used by Egly et al, but in some conditions they placed an occluding surface in front of the rectangles. At issue was whether the object-specific benefits documented by Egly et al persisted when the objects containing the target events were partly occluded and required perceptual organization to be completed [as in the Behrmann and colleagues (unpublished manuscript) study]. The results revealed a robust object-specific benefit. Yantis & Moore (1995) went on to show that when the perceptual task required of subjects was a temporal-order judgment, no object-specific benefit was observed; instead only location-based effects were observed. This latter result parallels the similar finding by Vecera & Farah (1994) noted earlier.

Several studies have shown that the attentional effects of image features can vary as a function of how they are perceptually grouped, rather than where they are located in the image. Such results support object-based theories of attention, on the assumption that the function of perceptual grouping is to create object representations. For example, Driver & Baylis (1989; Baylis & Driver 1992) asked whether interference from to-be-ignored stimuli during target identification depended only on relative spatial location or on more complex grouping principles. They employed the flankers task developed by Eriksen & Eriksen (1974) in which subjects are required to report the identity of a centrally located target letter and ignore adjacent noise letters. Eriksen & Eriksen found that when the noise letters were assigned to a response that conflicted with the response associated with the target letter, responses were significantly slowed, which suggests that attention could not be completely focused on the target letters.

Baylis & Driver (1992) constructed displays in which the stimuli were colored letters. For example, in their Experiment 2, five letters were arranged in a row. The first, third, and fifth letters were one color (e.g. red), and the second and fourth letters were another color (e.g. green). The letters X and Y were assigned to one response (e.g. press the right button), and C and S were assigned to another response (e.g. press the left button). H and T were neutral letters not assigned to a response; they never served as target letters. Subjects were supposed to press the button assigned to the identity of the middle (third) letter in the string. Baylis & Driver (1992) found that the identity of the letters that matched the target in color, and not the letters that were spatially closest to the target, had the greatest influence (facilitation and inhibition of RT). For example, the string $X^r\ H^g\ S^r\ H^g\ X^r$ (where the superscript indicates that letter's color) produced longer RTs than did the string $H^r\ X^g\ S^r\ X^g\ H^r$, even though the response-incompatible Xs are closer to the target in the second string than in the first string. Driver & Baylis (1989) obtained qualitatively identical results using grouping via common motion.

These results have been corroborated by Kramer & Jacobson (1991), who found that the extent to which flanking elements interfered with the identification of a target depended on whether the flankers were joined via connecting line segments to the target (producing large interference effects) or to other objects (producing smaller effects).

## *Motion*

A second approach to exploring the representational basis of visual selection is to separate objects from their location via motion. Kahneman et al (1992) introduced a priming technique that produces what they term an object-specific "re-viewing" effect. We here describe a simplified version of their Experiment 4. Each trial began with the appearance of a square and a triangle on opposite sides of the display (e.g. above and below fixation) for 500 ms. A capital letter then appeared within each shape for 1 s and then disappeared; this initial display constituted the “preview field.” The empty shapes smoothly moved to new positions to the left and right of fixation over a period of 590 ms, at which time they stopped and a target letter appeared within one of the shapes. The subjects were required to name the target letter vocally as quickly as possible after it appeared. The target letter could either be one of the two in the preview field, or it could be new. When it was one of the preview letters, it either appeared within the same object as it did in the preview field, or in the other object. This led to three possible trial types. For example, if the preview field consisted of an S in the square and a P in the triangle, then a *same-object trial* would consist of an S in the square during the target display, a *different-object trial* would consist of an S in the triangle during the target display, and a *no-match* trial would consist of a V in either shape.

Kahneman et al (1992) found that naming latencies were much slower for no-match trials than for the other conditions. More importantly, however, RTs were significantly faster for the same object condition than for the different object condition. They interpreted this finding as follows. When a visual object appears in the visual field, an object file is created for it. An object file is a temporary episodic representation of a visual object, containing a record of its location, its various attributes (including, perhaps, its name), and its recent history (Kahneman & Treisman 1984). In the case of the experiment described above, object files for the square (and the letter appearing with it in the preview display) and for the triangle (and its preview letter) are created at the start of the trial. When the target letter appears after the shapes move to their new locations, then the object file is reaccessed, and if the object file corresponding to the shape in which the target appears contains a trace of the target letter (as it would in the same-object condition) then naming latency is speeded relative to the no-match condition. This is a form of object-specific perceptual priming.

Another study in which motion was used to separate objects from their locations was reported by Yantis (1992), who used a multi-element visual tracking procedure devised by Pylyshyn & Storm (1988). On each trial, ten elements (small plus signs) appeared on the screen, usually in random locations. A subset of these ($n$ = 1 to 5) was flashed several times. This constituted the *target set.* The targets stopped flashing, and all 10 elements began to move about the screen independently, changing direction at random times and bouncing off one another and the edges of the screen. After 7 s, the elements stopped moving, and one was flashed. Subjects were to report whether this probe element was a member of the target set or not. Pylyshyn & Storm (1988) had shown that this task could be carried out with reasonably good accuracy, and to explain this they invoked a theory of visual indexing (Pylyshyn 1989) according to which each target element is independently indexed at the start of the trial. When the probe appears, it is queried to determine whether it is among the indexed set.

Yantis (1992) suggested instead that this task was carried out by grouping the target elements into a "virtual polygon," an object representation that is analogous to an object file (Kahneman & Treisman 1984, Kahneman et al 1992). Evidence for selective attention to the target elements supports an object-based representation for selection, because the target and nontarget elements could not be distinguished on the basis of spatial location (They were spatially intermixed as they moved so that no convex spatial region contained all and only targets). The experiments were designed to show that task performance was modulated by factors that influenced how easily perceptual groups could be created and maintained. In Experiment 4, for example, the configuration of target elements was either unconstrained or was constrained to remain convex during motion. The convexity constraint ensured that the virtual polygon would remain coherent throughout motion (i.e. the ordering of vertices along the perimeter remained constant), and this permitted observers to use a single object representation throughout motion. Performance in the constrained condition was significantly better than in the unconstrained condition.

Several studies have attempted to dissociate objects and locations in an *inhibition of return* (IOR) paradigm. Summoning covert attention to a spatial location with a task-irrelevant peripheral cue can either speed or slow detection of a subsequent target. When the target follows the onset of the cue by 150 ms or less, RTs are usually faster when targets are displayed in the cued location than in an uncued location (e.g. Maylor 1985, Posner & Cohen 1984). This facilitation has been thought to reflect one consequence of attending to a location. When the target follows the cue by more than 300 ms, however, time to detect a target is often faster for targets presented at previously uncued locations than at previously cued locations. It is this effect that Posner et al (1985) called inhibition of return. We consider here just one aspect of inhibition of return, namely whether it is associated with a spatial location or a perceptual object. (For discussions of other aspects of this phenomenon, see e.g. Abrams & Dobkins 1994, Klein & Taylor 1994, Kwak & Egeth 1992, Rafal et al 1989.)

Early experiments suggested that IOR was associated with spatial locations, specifically spatial locations defined in fixed environmental coordinates (Maylor & Hockey 1985, Posner & Cohen 1984). However, these locations often happened to coincide with objects. For example, in Posner & Cohen's (1984) study, displays consisted of squares. When brightened or dimmed, these squares served as cues. A target could then appear in a cued or an uncued square. However, the squares appeared in fixed locations on a screen, and thus it was not possible to say whether the resulting IOR was associated with a particular location, a particular square, or both.

In an effort to distinguish among these possibilities, Tipper et al (1991) set the squares into motion. Subjects fixated a central location around which two diametrically opposed squares revolved in a clockwise direction along the circumference of an imaginary circle. At a certain time during this circular motion one square was briefly cued (brightened). Both squares continued to revolve for a variable time until the target (a dot) was shown inside either the cued or the uncued square. Consider the case in which a square was cued as it reached the leftmost point of the circle, and the target dot was presented after the pair of squares had completed an additional half-turn (i.e. 180°) around fixation. If RT were slower when the target appeared at the left location, which is the same location in environmental coordinates as the original cue, than when it appeared at the right location, a location-based account would be supported. The pattern of results observed by Tipper et al (1991) clearly favored the opposite outcome. RT was slower when the target appeared within the previously cued object, which suggested that IOR is object-based under these conditions. In a subsequent study Tipper et al (1994) found evidence for both location-based and object-based IOR.

We have so far distinguished between object-based and location-based representations. This simple distinction might lead us to think that when attention is paid to an object, then the entire object benefits (or, in the case of IOR, suffers) equally. We conclude with a brief description of one additional study that suggests representations may be more complex than that. Gibson & Egeth (1994) argued that the conception of an object as *independent* of location should not be understood to imply that an object is *devoid* of location. (See also Baylis & Driver 1993, Farah et al 1990.)

An outcome of visual object processing appears to be a structural description that includes an explicit specification of relative locations of parts or surfaces within an object (e.g. Hummel & Biederman 1992). Thus, although objects are distinct from the spatial locations that they occupy, there exist other intraobject locations that may be fixed with respect to the overall object. That is, an object can be construed as a “microenvironment” within which specific locations may be tagged by the mechanisms that produce IOR or attentional facilitation. To test these notions, Gibson & Egeth (1994) employed a computer-generated depiction of a “brick” that rotated in depth in the time between the presentation of a cue and a subsequent target. The results of a series of four experiments showed that IOR was associated both with locations on the brick that remained fixed with respect to the brick as well as with locations that were fixed in reference to the unmoving environment.

## THE TIME COURSE OF ATTENTION

The deployment of attention from one stimulus to another is by no means instantaneous. A substantial body of research has explored the temporal characteristics of attentional deployment. In this section we examine (*a*) how quickly attention can be directed to a particular stimulus, (*b*) how long attention remains directed at a particular stimulus (the dwell time of attention), and (*c*) how attention moves from location to location.

### *Directing Attention*

There is a substantial literature concerning how attention may be covertly directed to a particular stimulus or to a location in the visual field, which was covered in a recent *Annual Review of Psychology* article (Kinchla 1992). Only basic findings are recounted here. In studies by Eriksen and his colleagues (e.g. Eriksen & Collins 1969, Eriksen & Rohrbaugh 1970) stimuli were briefly displayed letters on the circumference of an imaginary circle. A cue indicating the location of the to-be-reported letter could be shown in advance of the letter display. Accuracy of report increased with increasing stimulus-onset asynchrony (SOA) between the cue and the target letter. There was substantial improvement with just a 50-ms SOA, and the effect of the precue was asymptotic by about 200 ms. However, the story is not quite as simple as that description might suggest. As discussed in the section on attentional control, apparently two different mechanisms can direct attention to a stimulus or stimulus location—one that is stimulus-driven and another that is goal directed. We review here studies that reveal the time course of these mechanisms.

In a study by Müller & Rabbitt (1989), subjects fixated the center of a display while four boxes were present in the periphery of the display at the corners of a larger imaginary square. Subjects had to discriminate the orientation of a T presented in one of the boxes; the remaining three boxes contained plus signs. Before the presentation of these characters, subjects received a cue that was either the brief brightening of one of the four boxes (a peripheral cue) or the presentation of an arrow at the center of the display that pointed at one of the boxes (a central cue). These cues were partially valid. Half of the time they indicated the box that contained the critical T-shape, and half of the time they indicated one of the boxes that contained a plus sign. Performance was examined as a function of the SOA between the cue and the characters. The results showed that the peripheral cue had a fast-acting effect on performance. For example, with a valid peripheral cue, performance was quite good even at the shortest SOA (100 ms). It improved as SOA increased to 175 ms and then declined somewhat to a stable level for SOAs beyond 400 ms. In contrast, a valid central cue was virtually ineffective at 100 ms; performance increased steadily until, at 400 ms, it reached approximately the same stable level as that achieved by the peripheral cue. Thus, the peripheral cue was characterized as having a fast, transient response, and the central cue was characterized as having a slow, sustained response. More specifically, central cues elicit a deliberate shift of attention that is characterized by a monotonic rise to an asymptote, while peripheral cues produce a quick rise and then fall to a lower asymptotic level (and, perhaps, inhibition of return at still longer intervals). Similar findings have been reported by Kröse & Julesz (1989), Nakayama & Mackeben (1989), and Cheal & Lyon (1991).

## *The Dwell Time of Attention*

VISUAL SEARCH In much of the recent research on attention the visual search paradigm has been used to probe the mechanisms of attention. Among other things, this paradigm has been used to estimate the amount of time spent per item in the visual display. Let us take as a starting point search for a T in any orientation in a background of Ls in any orientation (see e.g. Bergen & Julesz 1983, Egeth & Dagenbach 1991, Wolfe et al 1989). This task is demanding and may well require serial processing (a requirement for any straightforward estimate of time per item). If one plots mean RT against display size, the resulting target-absent and target-present functions are nearly linear and have substantial slopes that stand in roughly a 2:1 ratio. For example, in the study by Wolfe et al (1989, Experiment 4) the present and absent slopes for one set of conditions were 19.2 and 41.6 ms per item, respectively, and for another set of conditions were 24.9 and 60.9 ms per item, respectively. Taken together, these two data sets suggest a serial search that inspects nontargets at the rate of approximately 50 ms per item until the target is found, with the shallower slope of the target-present function due to the subject terminating the search upon finding the target after half of the stimuli (on average) have been inspected. The slope of the target-absent function can be construed as the time that attention dwells on an item before moving to the next item. Obviously this dwell time will depend on many factors, such as the difficulty of the discrimination between targets and nontargets (see e.g. Cheal & Lyon 1992, Palmer 1994).

Use of the slope of a search function as an estimate of how long attention dwells on a stimulus has substantial face-validity, but it is not without interpretive problems (e.g. Palmer & McLean 1995; Townsend 1971, 1990). Perhaps the biggest problem is that the underlying serial model may be inappropriate. Suppose, for example, that nontargets were rejected in parallel by a limited-capacity process (such that it takes longer to work the more elements there are in the display). One could still compute a slope, but it would not accurately reflect the time course of attention across discrete items in the display. For examples of models with varying degrees and kinds of parallelism, see Duncan

& Humphreys (1989), Grossberg et al (1994), Hoffman (1979), Palmer & McLean (1995), and Wolfe (1994).

RAPID SERIAL VISUAL PRESENTATION (RSVP): WHOLE REPORT If we move beyond the search task, we find other paradigms that cast light on the time course of attention. Several involve the sequential display of stimuli as opposed to the simultaneous display of the visual search task. (For some early applications of such displays to the issue of whether processing is parallel or serial, see e.g. Eriksen & Spencer 1969, Shiffrin & Gardner 1972, Travers 1973.) Some of the sequential display procedures have yielded estimates of dwell time that are shorter than the roughly 50 ms estimates we get from visual search studies, while others have provided substantially longer estimates.

Saarinen & Julesz (1991) presented two, three, or four numerals in random positions on a ring surrounding fixation. Each numeral was followed by a mask in the same location. Each numeral and each mask was presented for 33 ms, and there was a blank interval of variable duration (0, 33, or 67 ms) between each numeral and its following mask. Each mask appeared simultaneously with the appearance on screen of the next numeral. Thus, SOAs in this experiment were 33, 67, or 100 ms. At the end of the stimulus presentation the subject was to type in all the numerals in the sequence in the correct order. Not surprisingly, as the number of numerals in the sequence increased, the proportion of trials on which the subject could correctly identify all of them decreased. The authors emphasized, however, that performance was above chance even when there were four numerals in the sequence and the SOA was 33 ms. They concluded that the speed of focal visual attention can be quite fast (at least 50 ms per item), with performance still respectable at 33 ms per item.

In a subsequent experiment, exposure durations as short as 16.7 ms were used, in addition to a condition in which stimuli were presented simultaneously (Hung et al 1995). Again performance was better than chance even with four-numeral sequences presented at the shortest SOA (i.e. 16.7 ms). Accuracy of report in the correct order in that condition was approximately 0.2%; their estimate of chance in that condition was 0.02%.

There is a problem, however, with basing this argument on the fact that performance was above chance. Suppose subjects always saw the first numeral clearly but, because of capacity limitations, saw none of the following items, which they would have to guess randomly. Even such minimal information would lead to above-chance performance, but such performance could not then be converted into a meaningful estimate of dwell time per item. By our own calculations, performance in the Saarinen & Julesz study appears to be too good to be accounted for in terms of subjects seeing one item and guessing three. If anything, it is more like seeing three clearly and guessing one. Thus, the work of Julesz and his colleagues strongly suggests a high speed for focal attention, but it is not clear whether a precise estimate of that speed is possible based on this technique.

The research of Saarinen & Julesz (1991) and Hung et al (1995) attempted to estimate dwell time by presenting stimuli sequentially and determining how quickly they can be presented while still maintaining above-chance performance. A different approach to estimating dwell time asks instead how slowly stimuli need to be presented to keep report accuracy at a high level. Some early research using RSVP was concerned with reading and so used letter sequences that formed words. Kolers & Katzman (1966) presented six letters one after the other in the same spatial location; they found that it took an SOA of 375 ms for accurate report (over 90% correct) of the letters in a sequence. Haber & Nathanson (1969) used a similar display format and presented words that varied in length from four to eight letters. They found that the SOA required for asymptotic performance increased with word length. For four-letter words they estimated the critical SOA to be 65 ms, and for eight-letter words 110 ms. Haber & Nathanson gave several reasons for believing that the relationship they found between word length and SOA may be artifactual. For example, there was no mask before the first letter or after the last letter. Thus two of the four letters in a four-item list are particularly easy, but only two of the eight items in an eight-item list were particularly easy. Thus, the 65 ms estimate of required processing time is probably too short. It is possible that the 110 ms estimate is also too short, at least if we consider the results of Travers (1973). In a condition in which the letters of a word were presented sequentially in the same spatial location (with the string preceded and followed by a mask) an exposure duration of 375 ms yielded between 80 and 85% of words (not letters) correctly identified.

RAPID SERIAL VISUAL PRESENTATION: PARTIAL REPORT One problem with interpreting the aforementioned studies is that the use of words creates opportunities for all sorts of guessing strategies to occur. To avoid these problems, one might present random letter strings (e.g. Kolers & Katzman 1966, Travers 1973). However, this creates problems of its own. In particular, memory requirements come to dominate task performance. One solution to this problem is to eliminate the need for whole report (Sperling 1960). A variety of interesting designs have adopted this approach. They have in common the requirement that subjects report the status of just one or two items, called target items, that are differentiated from the other items in the stream in some way.

In a series of four experiments, Broadbent & Broadbent (1987) distinguished targets from nontargets in several different ways. In one task, subjects had to report two uppercase target words presented in an RSVP stream of otherwise all lowercase words. Subjects were unable to report both targets if they were presented in temporally adjacent positions. Moreover, this deficit persisted even when words were separated by one, two, or three intervening nontargets. (At the exposure duration of 80 ms, three intervening items translates into an SOA of 320 ms between successive targets.) In another experiment using a somewhat more difficult discrimination (targets were designated by the presence of a hyphen on either side), the deficit in reporting both words was present for temporal separations of up to 480 ms. The difficulty of reporting both targets was not limited to situations in which target and foil were distinguished by a simple physical feature. Similar results were obtained in another task in which all items were lowercase and subjects had to report animal names. This lengthy refractory period is consistent with Duncan's (1980) claim that it is difficult to process two targets at the same time.

The RSVP studies supply an appreciation of what the phrase "at the same time" means. Roughly speaking, poor performance on the second target may be viewed as reflecting the duration of processing of the first target. However, this is a simplification. When items follow one another rapidly, subjects often process them in the "wrong" order. For example, in the Broadbent & Broadbent (1987) experiment using uppercase targets, when the two targets were temporally adjacent the probabilities of reporting the first and second targets were 0.46 and 0.35 respectively, but the probability of reporting both correctly was only 0.075. Thus, apparently on many trials subjects were able to report the second but not the first target. In contrast, when the targets were separated by three intervening items, the corresponding probabilities were 0.45, 0.14, and 0.075. Here the deficit would appear to be mostly, but not entirely, due to prolonged processing of the first target. Reeves & Sperling (1986) and Weichselgartner & Sperling (1987) provided detailed temporal analyses of responses in multitarget tasks.

The RSVP studies we have reviewed suggest that the allocation of attention to a target in a stimulus stream produces a fairly protracted deficit. It is not clear what the nature of this deficit is. In studies that required word identification the deficit may be in word identification, or it may be in some lower-level visual process. Similarly, in the Weichselgartner & Sperling (1987) study the deficit may be in memory mechanisms, or it may be in perceptual or attentional processes. Raymond et al (1992) suggested that what we are seeing in these studies is a suppression of visual processing. They wrote that "these data suggest that the mechanisms involved in target identification are temporarily shut down after use. It is as if the perceptual and attentional mechanisms blink" (p. 851).

Raymond et al (1992) designed a dual-task RSVP experiment in which response requirements were somewhat simpler than in preceding studies. Letters were presented one at a time at a rate of 11 per second. One letter was white; all the rest were black. Subjects had to identify the white letter. On half of the trials there was an "*X*" somewhere in the stream (but never prior to the white letter). On the other half of the trials there was no *X*. After reporting the identity of the white letter, the subject was to indicate whether the stream had contained an *X*. (Note that on some trials the white letter was an *X*.) Both responses were unspeeded, and the memory load was minimal. The focus of this study was on the consequences associated with paying attention to a target; for this reason, the white letter was referred to as the *target*, while the *X* was referred to as the *probe*. In the control condition, the subject was instructed to ignore the white letter and just indicate whether the probe was present or not. Note that the stimuli were identical in the experimental and control conditions. This allows one to determine whether posttarget performance deficits are due to sensory factors such as masking of the probe by the target or to the attentional demands of identifying the target letter.

The relevant data are the percentages of correct detections of the probe as a function of the position of the probe in the series. When the subject did not have to identify the white letter, probe detection was very good, averaging about 90% correct, and did not vary as a function of probe position. However, when subjects did have to identify the white letter, probe detection probability was similar to control performance at position 0, i.e. when the probe and target coincided, and declined steadily until, at position 3, probe detection was less than 50%. Performance then recovered gradually until, by position 6, it once again did not differ from control performance. This substantial and extended dip in performance of the experimental condition was referred to by Raymond et al (1992) as the *attentional blink*. In that study, it was statistically significant in the posttarget interval from 180 to 450 ms. We are concerned here chiefly with documenting the existence and extent of the attentional blink. Several attempts to provide theoretical accounts of the phenomenon have been proposed recently (e.g. Chun & Potter 1995, Grandison et al 1996, Raymond et al 1995, Seiffert & DiLollo 1996).

RAPID SERIAL VISUAL PRESENTATION: MINIMAL SEQUENCES The partial report procedure is clearly a simplification of the whole report method. However, RSVP with a multi-item sequence is a daunting task nevertheless, requiring selection of stimuli presented at high speed. One further simplification has been introduced. Duncan et al (1994; see also Ward et al 1996) presented just two stimuli sequentially and had subjects identify both. The stimuli were presented in two different locations and were both postmasked. When the stimuli were presented close together in time, the first stimulus interfered with the second. The investigators measured the interval over which the interference persisted. They took this to be an index of the time course of the first object's attentional demand (or, as they put it, of the dwell time of attention). The result is consistent with other estimates from the RSVP literature. Dwell times were several hundred milliseconds. A summary estimate of 500 ms would not be far off the mark.

The simplicity of the experimental paradigm makes this perhaps the strongest piece of evidence that search cannot shift between objects fast enough to account for the reaction times that are obtained in typical visual search tasks.

WHY THE DISCREPANT RESULTS? We take the high speed (17–33 ms) estimates of dwell time derived from the studies of Saarinen & Julesz (1991) and Hung et al (1995) to be provocative but, for the reasons stated above, not yet sufficiently secure to serve as the basis for further theoretical speculation. What remains, then, is to consider why the results from search tasks (50 ms per item) and a wide variety of RSVP tasks (500 ms per item) yield such different results. This section must be considered speculative because little work has addressed this question.

Moore et al (1996) noted that the experiments by Duncan et al (1994) and Ward et al (1996) used masked stimuli. Typically, masked stimuli are not used in visual search tasks. If dwell time depends on the specific stimuli and tasks used, then perhaps the discrepancy is more apparent than real. More specifically, attention may remain focused longer for a difficult discrimination than for an easy discrimination, and it is reasonable to think that masking may have made the discrimination difficult and thus led to an unusually long dwell time.

The Moore et al (1996) experiment was very similar in design to the experiments by Duncan and his colleagues. The one crucial difference is that in one condition both the first and second stimulus were postmasked immediately after their exposure (as in Duncan et al), while in another condition the first and second stimuli were masked at the same time (immediately after the exposure of the second stimulus). The critical question was whether the change in masking status of the first stimulus would affect the dwell time of attention on that stimulus. It did. Dwell time was reduced to about 200 ms. Although still longer than the times derived from visual search tasks, this time is less than half of the estimate in Duncan et al (1994). It seems possible that other differences between the methods might reduce the time still further.

Bennett & Wolfe (1996) have made a further effort to bridge the methodological gap between visual search and RSVP procedures. Subjects searched for a rotated T among rotated Ls, a task that would seem likely to elicit serial search. Stimuli were presented one at a time at random locations in a large field. SOAs varied across trials: 26, 52, 78, or 104 ms. Once presented, a stimulus remained in view until the end of the trial. Reaction time was measured from the onset of the trial until the subject pressed a key. What was of interest was how well subjects could “keep up” with the sequential presentation of items. If subjects could move attention from stimulus to stimulus at the same rate that they were being shown, then the slope of the function relating mean RT to the time in the sequence when the target was found should be 1.0. The seven subjects kept up with stimulus presentation at SOAs of 104, 78, and 52 ms, and fell behind only at 26 ms. This suggests that subjects can discriminate rotated Ts and Ls at a rate of about 50 ms per item.

With this interesting paradigm and result we have come full circle. Although presentation was sequential, the estimate of dwell time was as short as those from simultaneous presentations in classical studies of visual search. It is by no means clear just what it is about these various paradigms that leads to such differing results; this remains a problem for further research. Nevertheless, it seems reasonable to conclude, following MM Chun, JM Wolfe & MC Potter (unpublished manuscript) that "while it is *possible* to tie up attention for several hundred ms after a target has been detected, such commitment by no means represents a *mandatory* minimum dwell time for successful processing."

## *The Movement of Attention*

It is widely accepted that attention can be shifted from one location to another in the visual field without any concomitant movement of the eyes. However, the nature of the shift is less clear. Does attention move in an analog, continuous fashion, or is the shift of attention accomplished abruptly, without any actual movement?

Several investigators have obtained results that they took to support the idea that, like a spotlight, attention moves continuously through space (e.g. Shulman et al 1979) and thus requires more time to move a greater distance (e.g. Tsal 1983). For example, Tsal (1983) had subjects rapidly discriminate an "X" from an "O". Stimulus presentations were 4°, 8°, or 12° to the left or right of fixation. At a variable time before the letter was presented, a cue was briefly flashed at the location in which the letter was to appear. Tsal reasoned that the cue should be beneficial. On the basis of the assumption that attention takes time to move, he reasoned further that the maximum benefit of the cue should occur progressively later in time the further away from fixation the stimulus appeared. This was the pattern he observed.

This research suggesting that attention moves in an analog fashion has been criticized in some detail by Eriksen & Murphy (1987) and Yantis (1988). For example, one problem with the study by Tsal is that it did not include a control for general arousal or alertness. Yantis (1988, p. 205) concluded that both the Tsal (1983) and Shulman et al (1979) experiments "are simply inconclusive about whether attention shifts have continuous or discrete dynamics." However, there are some other approaches that avoid the problems of the aforementioned studies.

Sagi & Julesz (1985) showed evidence for an abrupt relocation of attention. Subjects had to decide whether two simultaneously presented stimuli were the same or different. The stimuli—rotated Ts and Ls—were presented at varying separations and were followed by a mask to limit processing time. The finding of chief interest was that at any given stimulus-mask onset asynchrony, discrimination accuracy was independent of distance, which they took as evidence for "fast, noninertial shifts of attention" (p. 141).

Converging evidence of distance-independent relocation of attention was provided by Kwak et al (1991). Subjects made same-different judgments about pairs of Ts and Ls that varied in separation. In their experiments, stimuli appeared at varying separations on an imaginary circle, to control acuity. The dependent variable of chief interest was reaction time. As in the Sagi & Julesz study, performance was independent of separation. This was true for both upright Ts and Ls (Experiment 1) and rotated Ts and Ls (Experiments 2, 3). Remington & Pierce (1984) reported a similar result.

In both the Sagi & Julesz (1985) and Kwak et al (1991) experiments, it is important to establish that the tasks actually required attention. If the tasks were accomplished preattentively, there would be little reason to speak of reallocation of attention. We focus here on the analysis provided by Kwak et al (1991, Experiment 3). To establish that the two stimuli were examined one after the other, they used a diagnostic based on the additivity of a within-display visual quality manipulation (Egeth & Dagenbach 1991). This diagnostic applies to situations in which processing of stimuli must be exhaustive. This is the case in a same-different matching task because both stimuli must be examined to make a correct response.

The two letters in the display were, independently, either high or low in contrast. If the letters are processed sequentially, then the slowing caused by presenting low-contrast characters should be additive. That is, if making one character low in contrast increases mean RT by 20 ms, then making both characters low in contrast should increase mean RT by 40 ms. However, if processing is parallel, then the 20 ms slowing produced by one low-contrast letter should not be exacerbated by making the other letter low in contrast. That is, the effects of reducing the contrasts of the two letters should be subadditive. The results showed a clear pattern of additivity for the rotated Ts and Ls. Thus, attention presumably had to move serially from one item to the other. Sagi & Julesz (1985) used some very different experimental diagnostics that also pointed to serial processing. Thus, it seems reasonable to suggest that it is appropriate to speak of the (null) results of the distance manipulation in these studies as indicating that it does not take longer for attention to move greater distances.

Sperling & Weichselgartner (1995) have independently reported evidence that longer movements of attention do not require more time. They further showed that attention can skip over an intervening obstacle without any time cost. Together these results suggest that the movement of attention is “quantal” rather than analog.

## Concluding Remarks

Our review reveals significant recent advances in the understanding of attention. Of course, many of the issues and some of the mechanisms that occupy journal pages today were anticipated to some degree as long as a century ago by William James and others. Nevertheless, many empirical details have been clarified, richer theoretical frameworks have evolved, and important new ideas, such as the distinction between object-based and location-based selection, have been advanced and developed. There is no reason to believe that this recent progress will subside any time soon; behavioral studies of attention, augmented by neuroimaging studies of the functioning brain and neuropsychological studies of brain-damaged patients, promise new insights into the mechanisms of visual attention.

## Literature Cited

Abrams RA, Dobkins RS. 1994. Inhibition of return: effects of attentional cueing on eye movement latencies. *J. Exp. Psychol.: Hum. Percept. Perform.* 20:467–77

Bacon WF, Egeth HE. 1994. Overriding stimulus-driven attentional capture. *Percept. Psychophys.* 55:485–96

Baylis GC. 1994. Visual attention and objects: two-object cost with equal convexity. *J. Exp. Psychol.: Hum. Percept. Perform.* 20: 208–12

Baylis GC, Driver JS. 1992. Visual parsing and response competition: the effect of grouping factors. *Percept. Psychophys.* 51:145–62

Baylis GC, Driver JS. 1993. Visual attention and objects: evidence for hierarchical coding of locations. *J. Exp. Psychol.: Hum. Percept. Perform.* 19:451–70

Bennett SC, Wolfe JM. 1996. Serial visual search can proceed at 50 msec per item. *Invest. Ophthalmol. Vis. Sci.* 37:298 (Abstr.)

Bergen JR, Julesz B. 1983. Parallel versus serial processing rapid pattern discrimination. *Nature* 303:696–98

Bravo MJ, Nakayama K. 1992. The role of attention in different visual-search tasks. *Percept. Psychophys.* 51:465–72

Breitmeyer BG, Ganz L. 1976. Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychol. Rev.* 83:1–36

Broadbent DE, Broadbent MHP. 1987. From detection to identification: response to multiple targets in rapid serial visual presentation. *Percept. Psychophys.* 42:105–13

Cave KR, Wolfe JM. 1990. Modeling the role of parallel processing in visual search. *Cogn. Psychol.* 22:225–71

Cheal ML, Lyon DR. 1991. Central and peripheral precuing of forced-choice discrimination. *Q. J. Exp. Psychol.* 43A: 859–80

Cheal ML, Lyon DR. 1992. Attention in visual search: multiple search classes. *Percept. Psychophys.* 52:113–38

Chun MM, Potter MC. 1995. A two-stage model for multiple target detection in rapid serial visual presentation. *J. Exp. Psychol.: Hum. Percept. Perform.* 21:109–27

Dagenbach D, Carr TH, eds. 1994. *Inhibitory Processes in Attention, Memory, and Language.* San Diego: Academic. 461 pp.

Desimone R, Duncan J. 1995. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18:193–222

Downing CJ, Pinker S. 1985. The spatial structure of visual attention. See Posner & Marin 1985, pp. 171–87

Driver JS, Baylis GC. 1989. Movement and visual attention: the spotlight metaphor breaks down. *J. Exp. Psychol.: Hum. Percept. Perform.* 15:448–56

Duncan J. 1980. The locus of interference in the perception of simultaneous stimuli. *Psychol. Rev.* 87:272–300

Duncan J. 1984. Selective attention and the organization of visual information. *J. Exp. Psychol.: Gen.* 113:501–17

Duncan J, Humphreys G. 1989. Visual search and stimulus similarity. *Psychol. Rev.* 96: 433–58

Duncan J, Ward R, Shapiro K. 1994. Direct measurement of attentional dwell time in human vision. *Nature* 369:313–15

Egeth H, Jonides J, Wall S. 1972. Parallel processing of multidimensional displays. *Cogn. Psychol.* 3:674–98

Egeth HE, Dagenbach D. 1991. Parallel versus serial processing in visual search: further evidence from subadditive effects of visual quality. *J. Exp. Psychol.: Hum. Percept. Perform.* 17:551–60

Egly R, Driver J, Rafal RD. 1994. Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *J. Exp. Psychol.: Gen.* 123:161–77

Eriksen BA, Eriksen CW. 1974. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* 16:143–49

Eriksen CW, Collins JF. 1969. Temporal course of selective attention. *J. Exp. Psychol.* 80:264–61

Eriksen CW, Hoffman JE. 1972. Temporal and spatial characteristics of selective encoding from visual displays. *Percept. Psychophys.* 12:201–4

Eriksen CW, Hoffman JE. 1973. The extent of processing noise elements during selective encoding from visual displays. *Percept. Psychophys.* 14:155–60

Eriksen CW, Murphy TD. 1987. Movement of attentional focus across the visual field: a critical look at the evidence. *Percept. Psychophys.* 42:299–305

Eriksen CW, Rohrbaugh JW. 1970. Some factors determining efficiency of selective attention. *Am. J. Psychol.* 83:330–42

Eriksen CW, Spencer T. 1969. Rate of information processing in visual perception: some results and methodological considerations. *J. Exp. Psychol.* 79:1–16

Farah MJ, Brunn JL, Wong AB, Wallace MA, Carpenter PA. 1990. Frames of reference for allocating attention to space: evidence from the neglect syndrome. *Neuropsychologia* 28:335–47

Folk CL, Remington RW, Johnston JC. 1992. Involuntary covert orienting is contingent on attentional control settings. *J. Exp. Psychol.: Hum. Percept. Perform.* 18:1030–44

Gibson BS. 1994. Visual attention and objects: one versus two or convex versus concave? *J. Exp. Psychol: Hum. Percept. Perform.* 20:203–7

Gibson BS. 1996a. Visual quality and attentional capture: a challenge to the special role of abrupt onsets. *J. Exp. Psychol.: Hum. Percept. Perform.* In press

Gibson BS. 1996b. The masking account of attentional capture: a reply to Yantis and Jonides (1996). *J. Exp. Psychol.: Hum. Percept. Perform.* In press

Gibson BS, Egeth H. 1994. Inhibition of return to object-based and environment-based locations. *Percept. Psychophys.* 55:323–39

Grandison TD, Ghirardelli TG, Egeth HE. 1996. Masking of the target is sufficient to cause the attentional blink. *Percept. Psychophys.* In press

Grossberg S, Mingolla E, Ross WD. 1994. A neural theory of attentive visual search: interactions of boundary, surface, spatial, and object representations. *Psychol. Rev.* 101: 470–89

Haber RN, Nathanson LS. 1969. Processing of sequentially presented letters. *Percept. Psychophys.* 5:359–61

Helmholtz H von. 1925. (1866). *Treatise on Physiological Optics,* Vol. 3, ed./transl. JPC Southhall. Washington, DC: Opt. Soc. Am. 3rd ed.

Hillstrom AP, Yantis S. 1994. Visual motion and attentional capture. *Percept. Psychophys.* 55:399–411

Hoffman JE. 1979. A two-stage model of visual search. *Percept. Psychophys.* 25:319–27

Hoffman JE, Nelson B. 1981. Spatial selectivity in visual search. *Percept. Psychophys.* 30:283–90

Hummel JE, Biederman I. 1992. Dynamic binding in a neural network for shape recognition. *Psychol. Rev.* 99:480–517

Hung GK, Wilder J, Curry R, Julesz B. 1995. Simultaneous better than sequential for brief presentations. *J. Opt. Soc. Am. A* 12: 441–49

James W. 1890. *The Principles of Psychology.* New York: Holt

Johnston WA, Dark VJ. 1986. Selective attention. *Annu. Rev. Psychol.* 37:43–75

Jonides J. 1981. Voluntary versus automatic control over the mind's eye's movement. In *Attention and Performance,* ed. JB Long, AD Baddeley, pp. 187–203. Hillsdale, NJ: Erlbaum. 9th ed.

Jonides J, Yantis S. 1988. Uniqueness of abrupt visual onset in capturing attention. *Percept. Psychophys.* 43:346–54

Joseph JS, Optican LM. 1996. Involuntary attentional shifts due to orientation differences. *Percept. Psychophys.* 58:651–65

Juola JF, Koshino H, Warner CB. 1995. Tradeoffs between attentional effects of spatial cues and abrupt onsets. *Percept. Psychophys.* 57:333–42

Kahneman D, Henik A. 1981. Perceptual organization and attention. In *Perceptual Organization,* ed. M Kubovy, JR Pomerantz, pp. 181–211. Hillsdale, NJ: Erlbaum

Kahneman D, Treisman A. 1984. Changing views of attention and automaticity. In *Va-*

*rieties of Attention*, ed. R Parasuraman, DA Davies, pp. 29–61. New York: Academic

Kahneman D, Treisman A, Gibbs BJ. 1992. The reviewing of object files: object-specific integration of information. *Cogn. Psychol.* 24:175–219

Kinchla RA. 1992. Attention. *Annu. Rev. Psychol.* 43:711–43

Klein RM, Taylor TL. 1994. Categories of cognitive inhibition with reference to attention. See Dagenbach & Carr 1994, pp. 113–50

Kolers PA, Katzman MT. 1966. Naming sequentially presented letters and words. *Lang. Speech* 9:84–95

Koshino H, Warner CB, Juola JF. 1992. Relative effectiveness of central, peripheral, and abrupt-onset cues in visual search. *Q. J. Exp. Psychol.* 45A:609–31

Kramer AF, Coles MGH, Logan GD. 1996. *Converging Operations in the Study of Visual Selective Attention.* Washington, DC: Am. Psychol. Assoc.

Kramer AF, Jacobson A. 1991. Perceptual organization and focused attention: the role of objects and proximity in visual processing. *Percept. Psychophys.* 50:267–84

Kröse J, Julesz B. 1989. The control and speed of shifts of attention. *Vis. Res.* 23:1607–19

Kwak HW, Dagenbach D, Egeth H. 1991. Further evidence for a time-independent shift of the focus of attention. *Percept. Psychophys.* 49:473–80

Kwak HW, Egeth H. 1992. Consequences of allocating attention to locations and to other attributes. *Percept. Psychophys.* 51: 455–64

LaBerge D. 1995. *Attentional Processing: The Brain's Art of Mindfulness.* Cambridge, MA: Harvard Univ. 262 pp.

Maylor E. 1985. Facilitory and inhibitory components of orienting in visual space. See Posner & Marin 1985, pp. 189–204

Maylor E, Hockey R. 1985. Inhibitory components of externally controlled covert orienting in visual space. *J. Exp. Psychol.: Hum. Percept. Perform.* 11:777–87

Mertens JJ. 1956. Influences of knowledge of target location upon the probability of observation of peripherally observable test flashes. *J. Opt. Soc. Am.* 46:1069–70

Moore CM, Egeth H, Berglan L, Luck S. 1996. Are attentional dwell items inconsistent with serial visual search? *Psychon. Bull. Rev.* In press

Müller HJ, Rabbitt PMA. 1989. Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *J. Exp. Psychol.: Hum. Percept. Perform.* 15:315–30

Nakayama K, Mackeben M. 1989. Sustained and transient components of focal visual attention. *Vis. Res.* 29:1631–47

Neisser U. 1967. *Cognitive Psychology.* New York: Appleton-Century-Crofts. 351 pp.

Nothdurft HC. 1992. Feature analysis and the role of similarity in preattentive vision. *Percept. Psychophys.* 52:355–75

Nothdurft HC. 1993. Saliency effects across dimensions in visual search. *Vis. Res.* 33: 839–44

Palmer J. 1994. Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vis. Res.* 34:1703–21

Palmer J, McLean J. 1995. *Imperfect, unlimited-capacity parallel search yields large set-size effects.* Presented at Soc. Math. Psychol., 28th, Irvine, CA

Pashler H. 1988. Cross-dimensional interaction and texture segregation. *Percept. Psychophys.* 43:307–18

Pashler H, ed. 1996a. *Attention.* London: Univ. Coll. London Press

Pashler H. 1996b. *The Psychology of Attention.* Camgridge, MA: MIT Press

Posner MI. 1980. Orienting of attention. *Q. J. Exp. Psychol.* 32:3–25

Posner MI, Cohen Y. 1984. Components of visual orienting. In *Attention and Performance,* ed. H Bouma, DG Bouwhuis, pp. 531–55. Hillsdale, NJ: Erlbaum. 10th ed.

Posner MI, Marin O, eds. 1985. *Attention and Performance.* Hillsdale, NJ: Erlbaum. 11th ed.

Posner MI, Rafal RD, Choate L, Vaughan J. 1985. Inhibition of return: neural basis and function. *Cogn. Neuropsychol.* 2:211–28

Posner MI, Snyder CRR, Davidson BJ. 1980. Attention and the detection of signals. *J. Exp. Psychol.: Gen.* 109:160–74

Pylyshyn ZW. 1989. The role of location indexes in spatial perception: a sketch of the FINST spatial-index model. *Cognition* 32: 65–97

Pylyshyn ZW, Storm RW. 1988. Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spat. Vis.* 3:179–97

Rafal RD, Calabresi PA, Brennan CW, Sciolto TK. 1989. Saccade preparation inhibits reorienting to recently attended locations. *J. Exp. Psychol.: Hum. Percept. Perform.* 15: 673–85

Raymond JE, Shapiro KL, Arnell KM. 1992. Temporary suppression of visual processing in an RSVP task: an attentional blink? *J. Exp. Psychol.: Hum. Percept. Perform.* 18:849–60

Raymond JE, Shapiro KL, Arnell KM. 1995. Similarity determines the attentional blink. *J. Exp. Psychol.: Hum. Percept. Perform.* 21:653–62

Reeves A, Sperling G. 1986. Attention gating in short-term visual memory. *Psychol. Rev.* 93:180–206

Remington RW, Johnston JC, Yantis S. 1992. Involuntary attentional capture by abrupt onsets. *Percept. Psychophys.* 51:279–90

Remington RW, Pierce L. 1984. Moving attention: evidence for time-invariant shifts of visual selective attention. *Percept. Psychophys.* 35:393–99

Rock I, Guttman D. 1981. The effect of inattention on form perception. *J. Exp. Psychol.: Hum. Percept. Perform.* 7:275–85

Saarinen J, Julesz B. 1991. The speed of attentional shifts in the visual field. *Proc. Natl. Acad. Sci. USA* 88:1812–14

Sagi D, Julesz B. 1985. Fast noninertial shifts of attention. *Spat. Vis.* 2:141–49

Seiffert AE, DiLollo V. 1996. Low-level masking in the attentional blink. *J. Exp. Psychol.: Hum. Percept. Perform.* In press

Shiffrin RM, Gardner GT. 1972. Visual processing capacity and attentional control. *J. Exp. Psychol.* 93:72–82

Shulman GL, Remington RW, McLean JP. 1979. Moving attention through visual space. *J. Exp. Psychol.: Hum. Percept. Perform.* 5:522–26

Sperling G. 1960. *The information available in brief visual presentations.* Psychol. Monogr. 74 (Whole No. 498)

Sperling G, Weichselgartner E. 1995. Episodic theory of the dynamics of spatial attention. *Psychol. Rev.* 102:503–32

Theeuwes J. 1990. Perceptual selectivity is task-dependent: evidence from selective search. *Acta Psychol.* 74:81–99

Theeuwes J. 1991a. Cross-dimensional perceptual selectivity. *Percept. Psychophys.* 50: 184–93

Theeuwes J. 1991b. Exogenous and endogenous control of attention: the effect of visual onsets and offsets. *Percept. Psychophys.* 49:83–90

Theeuwes J. 1992. Perceptual selectivity for color and form. *Percept. Psychophys.* 51: 599–606

Tipper SP, Driver JS, Weaver B. 1991. Object-centered inhibition of return of visual attention. *Q. J. Exp. Psychol.* 43A:289–98

Tipper SP, Weaver B, Jerreat LM, Burak AL. 1994. Object- and environment-based inhibition of return of visual attention. *J. Exp. Psychol.: Hum. Percept. Perform.* 20: 478–99

Todd JT, Van Gelder P. 1979. Implications of a transient-sustained dichotomy for the measurement of human performance. *J. Exp. Psychol.: Hum. Percept. Perform.* 5: 625–38

Townsend JT. 1971. A note on the identifiability of parallel and serial processes. *Percept. Perform.* 10:161–63

Townsend JT. 1990. Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychol. Sci.* 1:46–54

Travers JR. 1973. The effects of forced serial processing on identification of words and random letter strings. *Cogn. Psychol.* 5: 109–37

Treisman AM, Gelade G. 1980. A feature-integration theory of attention. *Cogn. Psychol.* 12:97–136

Treisman AM, Gormican S. 1988. Feature analysis in early vision: evidence from search asymmetries. *Psychol. Rev.* 95: 15–48

Tsal Y. 1983. Movements of attention across the visual field. *J. Exp. Psychol.: Hum. Percept. Perform.* 9:523–30

van der Heijden AHC. 1992. *Selective Attention in Vision.* New York: Routledge, Chapman & Hall. 310 pp.

Vecera SP, Farah MJ. 1994. Does visual attention selection objects or locations? *J. Exp. Psychol.: Gen.* 123:146–60

Ward R, Duncan J, Shapiro K. 1996. The slow time-course of visual attention. *Cogn. Psychol.* 10:79–109

Weichselgartner E, Sperling G. 1987. Dynamics of controlled visual attention. *Science* 238:778–80

Wolfe JM. 1994. Guided search 2.0: a revised model of visual search. *Psychon. Bull. Rev.* 1:202–38

Wolfe JM, Cave KR, Franzel SL. 1989. Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol.: Hum. Percept. Perform.* 15:419–33

Yantis S. 1988. On analog movements of visual attention. *Percept. Psychophys.* 43: 203–6

Yantis S. 1992. Multi-element visual tracking: attention and perceptual organization. *Cogn. Psychol.* 24:295–340

Yantis S, Hillstrom AP. 1994. Stimulus-driven attentional capture: evidence from equiluminant visual objects. *J. Exp. Psychol.: Hum. Percept. Perform.* 20:95–107

Yantis S, Jonides J. 1984. Abrupt visual onsets and selective attention: evidence from visual search. *J. Exp. Psychol.: Hum. Percept. Perform.* 10:601–21

Yantis S, Jonides J. 1990. Abrupt visual onsets and selective attention: voluntary versus automatic allocation. *J. Exp. Psychol.: Hum. Percept. Perform.* 16:121–34

Yantis S, Jonides J. 1996. Attentional capture by abrupt visual onsets: new perceptual objects or visual masking? *J. Exp. Psychol.: Hum. Percept. Perform.* In press

Yantis S, Moore C. 1995. *Spread of visual attention behind an occluding surface.* Presented at Annu. Meet. Psychon. Soc., 36th, Los Angeles