5

Visual Attention: The Active Vision Perspective

John M. Findlay
Iain D. Gilchrist

This chapter discusses the relationship between covert visual attention and overt movements of the eyes. One of the most frequently emphasised facts concerning visual attention is the ability to attend covertly to a location in the visual field without directing the eyes to that location. Based on this, the "mental spotlight" metaphor has been widely prevalent and has given rise to much experimental work. An important, but rather neglected, question concerns what role such a mental-spotlight process might play in vision. The compelling nature of the spotlight metaphor seems likely to arise because of its similarity with overt movements of the eyes. This leads to the suggestion that covert attention is involved in some form of mental scanning. We will argue that this account is not satisfactory. An alternative account, termed active vision, develops from the fact that overt eye scanning occurs several times each second in many tasks. Analysis of information intake during text reading, scene perception and visual search suggests that no additional covert scanning occurs in these cases. Evidence however suggests that attentional processes may operate to assist in pre-processing information in the visual periphery at the location to which the eyes are about to be directed. Covert attention to peripheral locations thus acts to supplement, not substitute for, actual movements of the eyes.

5.1 Introduction

There was an extraordinary proliferation of work on visual attention in the late 1990's, summarised in several recent books (Styles, 1997; Pashler, 1998; Wright, 1998) and numerous journal articles (reviews include Egeth and Yantis, 1997; Cave and Bichot, 1999). A frequent starting point in discussions is the simple demonstration that attention to visual locations can operate covertly. An observer can, while fixating one location, attend to a different location without moving the eyes to look at it. This result, known to Helmholtz, was brought into prominence by the work of Michael Posner (Posner, Nissen, and Ogden, 1978; Posner, 1980) whose experimental paradigms have proved highly productive. Here we consider

how covert attention is used in common visual tasks. We argue that covert attention does not operate as an independent scanning mechanism but instead supplements the process of overt eye scanning that forms the normal way that visual attention is redirected.

Attention to a location demonstrably affects the speed and accuracy of visual processing at that location. Visual material at the attended location is selected for processing and receives enhanced processing. Attention generally leads to benefits but an ingenious demonstration by Yeshurun and Carrasco (1998) describes a case where attention actually renders visual discriminations less effective. Two separate preconditions are recognised. Attention may be directed voluntarily in response to instructions or cues. However, attention may also be captured by environmental events. In particular, the occurrence of visual change in the periphery has the effect of summoning attention to its location. The consequences, in terms of enhanced processing of subsequent material, are similar in the two cases, but differences in time course and other properties are found (Müller and Rabbitt, 1989; Nakayama and Mackeben, 1989). These two forms of attention are distinguished in various ways (endogenous, exogenous; sustained, transient etc). At the heart of this chapter is the question: How are the attentional effects best described? Posner and followers have emphasised a spotlight metaphor whereby attention is envisaged as a mental process akin to a spotlight beam which "illuminates" a restricted region of some internal mental representation of the visual scene. Much work on visual attention has been concerned with exploring the validity of the spotlight account and offering possible alternatives. The suggested alternatives (zoom lens, hemifield inhibition, objects rather than locations), in general, retain the idea that attention operates covertly by facilitating some part of the mental representation at the expense of others.

The implicit assumption of the spotlight model is the existence of a full and rich mental representation upon which this attentional spotlight could act and across which it could move. Until recently, this assumption had been unquestionable. It appeared self-evident that the spotlight had something to illuminate. Vision was analysed as a process that started at the retinal image. The work of David Marr (1982) epitomised this approach and offered a computational account elaborating processes that might transform a raw retinal image into a three-dimensional scene representation. The 1990s saw the gradual realisation that this account of vision may be seriously misleading. As pointed out by O'Regan (1992, see also Nakayama, 1992), the traditional image-processing approach did not appear to be nearing a solution of "the real mysteries of vision". A polemic entitled "A critique of pure vision" (Churchland, Ramachandran, and Sejnowski, 1994) argued that the "picture-in-the-head" metaphor for vision was still too pervasive among vision scientists and had outlived its usefulness. The incompleteness of the mental visual representation has now become widely accepted, and more recent work on "change blindness" (Grimes, 1996; Rensink, O'Regan and Clark, 1997; Rensink, 2000) has provided a convincing immediate demonstration.

These developments necessitate a reappraisal of the role of attention in the visual process. In this chapter we will examine vision from a somewhat different

perspective – that of active vision – and discuss how attention might be incorporated into this framework.

5.2   Active Vision

Active Vision is a term that appears to have been coined by computer scientists working on visual problems. For a time, work on computer vision was dominated by the image-processing approach advocated by Marr (1982). Attempts were made to derive an elaborated representation at each location of a visual scene. These attempts were largely unsuccessful and led to an appreciation of the vast computational resources that would be entailed in order to derive a full three-dimensional representative description across the whole of an extended image. A promising alternative emerged from work that downplayed the perceptual aspect of vision and concentrated on the means whereby a vision system might lead to useful visually controlled behaviour. A key to progress appeared to be a visual sensor, of which a region, effectively a fovea, could be directed to different locations within the visual field. Such a system allows vision to deploy *deictic*, or pointing, properties that are essential for interfacing with the cognitive system (Ballard, Hayhoe, Pook and Rao, 1998). Aloimonos, Bandopadhay and Weiss (1987) termed such a system an *active vision system*. Ballard (1991) elaborated very similar ideas but preferred the term "*animate vision*[1]". Of course, this work developed in full awareness that the vision of humans, and that of other active animals, utilises the mobility of the eyes (Land, 1995). In many situations, rapid saccadic, or jump-like, movements reposition the gaze several times each second. No useful vision occurs during these movements and information intake is restricted to the successive fixations of the eye, pauses of a fraction of a second between the scanning movements. These movements are characteristic of most daily visual activity. Land, Mennie,and Rusted (1999), in an elegant study, comment that "the eye-movement [record] has the frenetic appearance of a movie that has been greatly speeded up."

It is something of a mystery that textbooks of vision frequently ignore the mobility of the eye. Equally disturbing is the tendency to ignore the fundamental nonhomogeneity of the visual system that renders these movements essential. The highest visual acuity is only achieved in the region of the central fovea. The fovea is technically defined as an anatomical depression in the retina occupying about 2 degrees. In human vision, this is also, at least approximately, the area from which rods are absent and vision is mediated entirely by cones. However, while the fovea can be defined in anatomical terms, there is no corresponding functional boundary at its edge. Visual acuity declines steadily from the very centre of the

---

[1]Ballard wished to avoid possible confusion with active sensing, a term preempted in the computer-vision world. Active vision might also offer confusion with the quite different approach of J. J. Gibson (1966, etc.). Nevertheless, we believe it is the term that most appropriately captures the approach.

fovea out to the periphery (Wertheim, 1894), as do many other visual functions (Rovamo, Virsu and Näsänen, 1978).[2] The nature of this steady decline gives it a remarkable self-scaling property. As remarked by Anstis (1974), the acuity vs eccentricity tradeoff function ensures that the visibility of visual material is largely unaffected by its retinal size. Retinal size is affected by viewing distance, and the self-scaling property is surely of significance in allowing us to adapt our visual behaviour to material at a wide variety of viewing distances.

The active-vision approach emphasises the sequential nature of information intake in vision. The important questions become: What information is taken in at each fixation? How is this information integrated with that from preceding and subsequent fixations? How are the scanning movements planned and orchestrated? In one specialised area – the study of information intake during reading – much progress has been made towards answering these questions (Rayner, 1998). The reading process has some unique simplifying constraints. The eyes move across the text, in general, in a prespecified direction, and acuity limitations mean that relatively little use is made of parafoveal or peripheral vision. Other situations are less constrained and less understood, although some progress is being made in the areas of scene perception and search (Henderson and Hollingworth, 1999; Liversedge and Findlay, 2000).

How might covert attentional processes interact with overt scanning movements of the eye? Several possibilities exist. First, covert attention might be entirely absent when overt eye scanning occurs. Covert attention might be just a curious artefact of the peculiar instructions given in the psychology laboratory. Second, at the opposite extreme, overt eye movements might be incidental. Covert attention shifts might be the main way in which attention was redirected. In some accounts, it is suggested that several locations might be scanned by covert attention within each eye fixation. For this sequential covert attentional scanning to occur within a single fixation it would, of course, be necessary for covert attention to be redeployed much more rapidly than the speed of overt eye scanning.

We will argue that neither of these extreme positions is tenable, and particularly argue against the existence of "sequential intrafixational scanning". A more promising approach builds on the long recognised link between covert attention and eye movement programming. It was demonstrated some years ago (Shepherd, Findlay and Hockey, 1986; see also Hoffman and Subramaniam, 1995) that during the preparatory period of a voluntary eye movement, responses to an attentional probe were faster at the destination location of the eye movement, even when the manipulation of probe probability had rendered an alternative location more probable. Such a finding is consistent with the *pre-motor theory of attention* (Rizzolatti, Riggio, DaScola, and Umiltá, 1987; Rizzolatti, Riggio and Sheliga,

---

[2]Despite this fact, we will continue to use the convenient terms fovea, parafovea, and periphery as introduced by Rayner (1984). "Foveal vision" refers to the central 2 deg of vision and "parafoveal vision" refers to the central 10 deg with "periphery" referring to the remainder. In relation to the total visual field, the fovea occupies about 0.04% of the total area and the parafovea about 1%.

1994). This theory proposes that covert attention to a visual location involves using the mechanisms for saccade programming, but with the actual motor response withheld. We believe the pre-motor theory is essentially correct but that, rather than putting the emphasis on pre-motor theory as an account of covert attention, it should be recognised that overt attention has primacy.

Further evidence indicating this attention link comes from experiments showing that visual discriminations at the destination location of an eye movement are improved in comparison with other locations (Kowler, Anderson, Dosher and Blaser, 1995; Deubel and Schneider, 1996). These two findings demonstrated that the improved discrimination occurred before the eye movement itself commenced. Many studies have demonstrated *pre-view advantage* when saccadic eye movements are made. The visibility of material in parafoveal or peripheral vision prior to an eye movement facilitates subsequent recognition. This suggests that the improved pre-saccadic visual discrimination at the saccade destination is normally incorporated into an integrated routine for dealing with information trans-saccadically, thereby speeding up the recognition process.

This facilitation of discriminations at the destination location of a saccade is often described by saying that covert attentional processes precede the saccadic movement. It is further frequently assumed that, during the preparatory period of the saccade, covert attention moves from the fixated location to the saccade destination location in a spotlight-like manner. However, this extension assigns concrete properties to attention, such as a unique and restricted locus. As Kowler et al. (1995) point out, this assumption may be unwarranted. One tradition in attentional work has argued that spatial shifts of attention between two locations are characterised by three mental operations; disengagement of attention at the first location, movement of attention, and re-engagement of attention at the second location (Posner, Walker, Friedrich, and Rafal, 1984). Although superficially plausible, this position may be challenged. Thus one piece of supporting evidence, the "gap effect" in saccadic eye-movement programming, has proved better interpreted in an alternative way (Kingstone and Klein, 1993; Findlay and Walker, 1999).

We will now consider the role of attention in turn in the areas of reading, scene and object perception, and visual search.

5.3   Reading

Reading is a ubiquitous and very familiar activity. The study of visual information intake in reading has a long history and provides a benchmark situation for studies of active vision. An excellent review by Rayner (1998) gives a very thorough general account. We will select certain details here, which emphasise possible involvement of visual attention in the reading process.

A reader typically advances along a line of text making a sequence of fixations and saccadic movements. The average length moved is around 7–8 characters,

showing a rough correspondence to the average length of a word. Short words are frequently not fixated and long words will frequently receive more than one fixation. Occasionally a regressive movement in the reverse direction breaks the forward sequence of the scanning.

An extremely fruitful innovation for investigating the reading process was the *gaze-contingent methodology* developed by McConkie and Rayner (1975). This combined advances in computer technology with advances in eye gaze recording to present observers with visual displays where the visual information could be changed in a manner dependent on where the observer's gaze was directed. Various different detail manipulations of this generic technique have been developed, among which the *moving window* manipulation is perhaps the most widely used. In the moving window technique, as applied to reading, the subject sees normal text within a window of predefined size but outside this window the text is masked in some way, for example each text character may be replaced by an "x." Each time the eye moves, the display is changed so that the unmodified material is always presented precisely where the gaze is directed.

This technique has been used to measure the *perceptual span* in reading. If the manipulation does not affect the speed of reading, it is reasonable to assume that the material outside the window is playing no part in the normal process. Conversely, when the window size is reduced so that reading speed is affected adversely, then it can be deduced that material from regions outside the boundary must normally be processed. This allows measurement of the perceptual span, the region from which visual information is extracted. Summarising many results, the perceptual span is found to extend no more than 15 characters to the right of fixation and only 3–4 characters to the left. Information about the visual characteristics of letters is only available within a smaller window, extending 8–12 characters to the right; the contribution to the span of material more peripheral than this comes exclusively from word boundaries.

The asymmetry of the span is an immediate indication that attentional effects are involved. Visual acuity declines systematically away from the fovea with no left-right differences. The asymmetry has been found to reverse when readers read text in languages such as Hebrew, where the direction of reading goes from right to left. This finding alone allows rejection of one of the suggestions made earlier, that attentional processes are not operative when free eye scanning is allowed.

A study by Blanchard, McConkie, Zola, and Wolverton (1984) addressed the possibility that, during each fixation, a covert attentional scanning process operates within the material in the perceptual span. For example, suggestions had been made that reading operates as a sequential letter-by-letter serial encoding process (Gough, 1972). Blanchard et al. programmed a gaze-contingent display where the display changed three times during each fixation. An example is shown in Fig. 5.1.

This sequence of display changes occurred on every fixation. The critical change occurred to one carefully chosen five letter word of the text, where one letter was changed from its first exposure (time T_1) to its second (time T_2). Thus, during the early part of the fixation (T_1), the first word ("tombs") is present, whereas

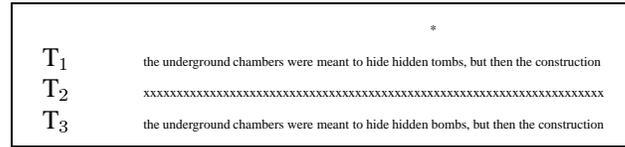|	
	*
T_1	the underground chambers were meant to hide hidden tombs, but then the construction
T_2	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
T_3	the underground chambers were meant to hide hidden bombs, but then the construction

FIGURE 5.1. Text changes in the experiment of Blanchard et al. (1984). The experimental subject read a line of text. During every fixation, the text was changed at the critical point marked by the asterisk. Up to time T_1, the word "tombs" was displayed. After a further time T_2, the word "bombs" replaced it. The brief masking stimulus presented (for 30 msec) at time T_2 prevented the attention-grabbing effect that would otherwise occur when the letter was changed.

during the second part (T_3), the second word ("bombs") has replaced it. Equivalent changes were made where letters at different positions in the word sequence (e.g., changing the word "tombs" to "tomes") were made. A critical prediction of the left-to-right serialisation hypothesis is an effect of the relative locations of fixation and critical letter on the probabilities of reporting the first or the second word. When the changed letter occurs to the left of the fixation location, then the first word should be more likely to be reported, whereas when the critical letter lies to the right, an increase in the probability of reporting the second letter is expected. The results did not provide any evidence to support the attentional-scanning prediction. Subjects would sometimes report the first word and sometimes the second[3]. The likelihood of reporting each word depended on the timing of the changes (times T_1 and T_3) but was not affected by the relative fixation position.

The results of the Blanchard et al. (1984) study suggest that a sequential left-to-right intake of letters does not occur. Such a possibility is an extreme form of the sequential intrafixational scanning position. It could only occur if covert attention could scan the letters at a very rapid rate. Rates of attentional scanning will be considered further in a subsequent section, and it will be argued that there is no good evidence that scanning of attentional focus can occur at a rate that is considerably in excess of the rate of overt scanning. We consider next a proposal that has widespread, although not universal, support. This is the idea that within a fixation, the attentional focus is initially on the fixated word and moves at some point during the fixation to the following word in the text.

The length of the perceptual span is such that, frequently, two words are present within the span. If the moving window is reduced so that only one word is available, reading is possible but with some cost. Inhoff and Rayner (1986) found that gaze durations on a word were about 30 msec shorter when the word had been previously visible in the periphery of vision than when the word had not been previewed. The usual interpretation of this finding is that during a fixation, in-

---

[3]On a small proportion of occasions (35%), both words could be reported. These trials were excluded from subsequent analyses.

formation is taken in, both from the word fixated, and also from the following word (on the right of fixation for English readers). This information can be used to speed up some process, possibly lexical access, at the time when the following word is, in its turn, fixated. It may also be used in skipping the next word. The term *peripheral preview advantage* succinctly describes the result. Peripheral preview in reading-like tasks has been considerably investigated (Rayner, 1998). A principal concern has been what type of information is extracted and integrated across the saccade that brings the word into foveal vision. Rayner (1998) reviews the evidence and suggests that abstract letter codes and phonological codes are used for this purpose.

The peripheral-preview effect is well established and shows the operation of attentional processes in reading. However there are two alternative positions concerning how these attentional operations are effected. One position argues that covert attention is located at the fovea during the initial part of the fixation and, at some point during the fixation, shifts to the destination of the subsequent saccade. The second position argues that during the fixation, material from the fovea and the saccade destination location is processed in parallel. The first position has been advocated in a series of models of increasing sophistication produced by the Amherst group (Morrison, 1984; Henderson and Ferreira, 1990; Reichle, Pollatsek, Fisher, and Rayner, 1998). The second position is linked to the work of McClelland and Mozer (1986).

The idea that attention shifts at a specific point in time prior to the actual eye movement appears intuitively plausible but it is difficult to support with direct evidence. It has been shown that foveal information can be extracted if only visible for the first 50 msec of the fixation (Slowiaczek and Rayner, 1987). However the study of Blanchard et al. (1984) discussed earlier shows that information presented at a later point during the fixation is also processed. Other tests involve further assumptions. The models that involve a discrete attention shift identify a process that triggers this shift of attention. This is generally related to the lexical processing of the foveated word. A testable prediction is thus generated. Because the processing of the foveal word determines when attention shifts, then the content of the parafoveal word cannot affect the timing of this attention shift. If the further assumption is made that the attention shift is followed after a fixed time by the actual eye movement, neither will it affect the fixation time. Henderson and Ferreira (1993) tested this prediction and reported no effect of the processing difficulty of the parafoveal word on fixation times. However, Kennedy (1998; see also Inhoff, in press) presents a contrary finding.[4]

In summary, work on reading has very clearly demonstrated attentional effects through the asymmetry of the perceptual span. It is clear that reading is charac-

---

[4]It may be noted that there is a further lively debate about how the extent of parafoveal preview depends on the difficulty of the fixated word (Henderson and Ferreira, 1990; Kennison and Clifton, 1995; Schroyens, Vitu, Brysbaert and d'Ydewalle, 1999). Although this has relevance for theories involving attentional movement, it brings in the "limited-capacity" view of attention, and a full discussion would go beyond the scope of this chapter.

terised by active sampling of visual information with the eyes and that a small but significant contribution to the process comes from a preview of the material to which the eyes are to be directed next. There is still debate about whether this preview occurs because the attentional focus makes a specific movement during the fixation, or whether the foveal and the attended parafoveal material are processed in parallel. Importantly for the current chapter, there is no support for a more elaborated sequence of covert attentional movements (for example a letter-by-letter scan) within a fixation.

5.4   Scenes and Objects

Although much less intensively investigated, some of the principles emerging from studies of active vision in reading are also at work during the more general situation of viewing visual material. An important difference is that scenes allow "gist" information to be extracted very rapidly. The nature of this gist information has been a topic of considerable recent interest, in view of the demonstrations (change-blindness) that much less specific information is available than once thought. One suggestion is that gist information relates to statistical properties of colour and contour distribution rather than more fully analysed information (Henderson and Hollingworth, 1999).

Attempts have been made to describe the useful field of view for picture perception using a gaze-contingent window technique in a manner similar to that used in studies of reading. Saida and Ikeda (1979) created a viewing situation with an electronic masking technique in which subjects saw only a central square region of an everyday picture. Viewing time and recognition scores were impaired unless the window was large enough for about half the display to be visible. This fraction applied both to 14 deg by 18 deg displays and to 10 deg by 10 deg ones. A follow up study by Shiori and Ikeda (1989) looked at the situation where the picture was seen normally inside the window but in reduced detail outside. Even very low-resolution detail (very well below the acuity limit at the peripheral location) aided performance considerably. Shiori and Ikeda used a pixel-shifting technique to reduce resolution. Van Diepen, Wampers, and d'Ydewalle (1998) have developed a technique whereby a moving window can be combined with Fourier filtering. Somewhat surprisingly, performance was better when the degraded peripheral material was low-pass filtered than when it was high-pass filtered.

When scenes are viewed containing a number of different objects, fixations occur on these objects and are important. Nelson and Loftus (1980) allowed subjects a limited amount of time to scan such a scene, and they then performed a recognition test, viewing a version of the scene in which one object had been changed. Their task was to detect the changed object. The detection rate for small objects (1 deg) was above 80% for objects that had been directly fixated and higher if they received two fixations. However, it fell to 70% for objects where the closest original fixation had been at a distance from the object in the range 0.5 – 2 deg

FIGURE 5.2. The experiment of Pollatsek et al. (1984). On fixation 1, visual material was presented in peripheral vision, which the subject was required to fixate and name. During the saccade prior to the second fixation, the material could be changed in various ways or left unchanged.

and to a chance level (or in one case slightly above chance) for objects viewed more peripherally.

As with words, parafoveal preview aids object identification. A series of studies at Amherst University has used the boundary technique as shown in Fig. 5.2 to explore the preview benefit of presaccadic material in peripheral vision. Pollatsek, Rayner and Collins (1984) measured the speed of object naming when various forms of preview occurred, using as a baseline a square containing a cross. They found that full peripheral preview produced a substantial benefit (100–130 msec). Benefits, albeit of a reduced magnitude (90 msec), also occurred if the previewed picture came from the same category as the target picture. A small benefit (10 msec) occurred when the previewed picture had the same overall visual shape as the picture (e.g., carrot > baseball bat). This preview information only occurs for previewed material at the saccade destination: material elsewhere in the visual field does not generate any priming (Henderson, Pollatsek and Rayner, 1989).

The emerging view from this discussion is that scene perception, apart from the initial gist, is built up from the serial eye scan. A result apparently in contradiction to this position was reported by Loftus and Mackworth (1978). They asked people to look at scenes, some of which contained an object that was highly unlikely in the context (e.g., an octopus in a farmyard). Control scenes had a visually similar but contextually probable object (a tractor in a farmyard), and subjects' eye scans were recorded. Loftus and Mackworth reported that the anomalous object was detected rapidly and often when the eyes were located in a remote part of the scene immediately prior to the movement to the anomaly. This implied that detailed analysis of scene content in peripheral vision was occurring and influencing subsequent behaviour. However, as reviewed by Henderson and Hollingworth (1998, 1999), several attempts to replicate Loftus and Mackworth's findings have now been attempted and all have failed to replicate the crucial findings.

Work on object and scene perception shows that peripheral vision is much more important in this area than in reading. However, the identification of objects within scenes appears to proceed in a way quite similar to the identification of words

during reading. Peripheral preview processes are likely to be operative, but none of the current work on scene perception makes use of the concept of spotlight-based shifts of covert attention.

5.5   Search

In visual search, the task is to locate a target on the basis of some visual properties. Everyday examples of visual-search tasks abound (finding a cup in the kitchen, finding a book on a bookshelf). The target book may be specified precisely (Harris and Jenkin's *Vision and Action*) or loosely (a red object). It is also easy to set up controlled laboratory tasks to investigate visual search. Typically a display is presented containing a target together with a number of additional items, termed nontargets or distractors.

A dominant tradition in visual search was initiated with a seminal paper by Treisman and Gelade (1980). They argued that some primary visual properties allow a search in parallel across large displays. In such cases the target appears to 'pop out' of the display. For example there is no problem in searching for a red item amongst distractor items coloured green, blue or yellow. In other cases, the paradigm example being a 'feature conjunction' search for a target that is both green and a cross when distractors include red crosses and green circles, the task is much more difficult, suggesting the use of a different search strategy. Treisman and Gelade argued that in the pop-out tasks preattentional mechanisms permit rapid target detection, in contrast to the conjunction task, which was held to require a serial deployment of attention over each item in turn.

Treisman and Gelade introduced an experimental paradigm that differentiated the different types of searches. They measured the time taken for an observer to make a speeded two choice decision concerning the presence or absence of a target in a visual display. Half the displays contained a target and in the remaining the target was absent. The critical independent variable was the number of display items. The *search function* shows how the response time depends on this variable. Figure 5.3 shows the classic search functions from Experiment 1 of Treisman and Gelade's paper.

The traditional interpretation of the search function is that the display-size-dependent increases shown in the search functions for conjunction searches come about through an item by item serial scan of covert attention through the display. If the display does not contain a target, it is assumed that every item in the display is scanned before a target-absent response is given. If the search is self-terminating in displays that do contain a target, then on average half the display items must be scanned before the target is found. This accounts for the result, seen above, and often repeated, that the slope of the search function in the target-absent cases is approximately twice that of target-present cases. Using this logic, the speed of search scanning may be estimated from the slope of the target-absent data and is about 60 ms/item.
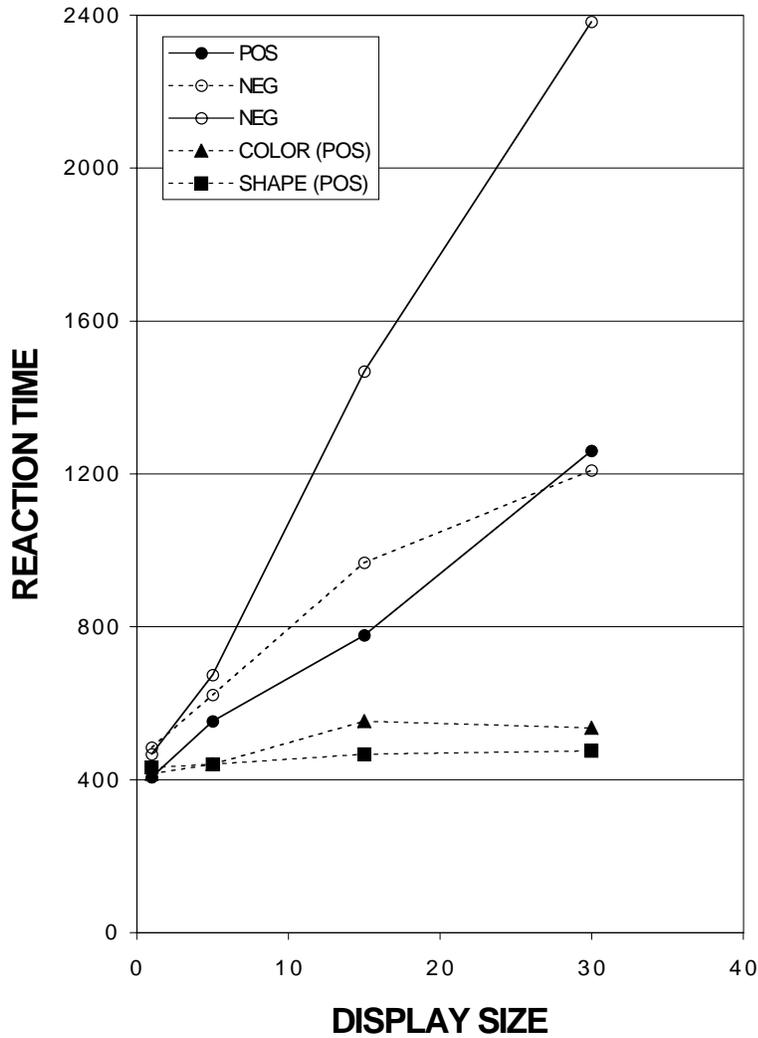
FIGURE 5.3. Results from Experiment 1 of Treisman and Gelade (1980). The plots show a series of search functions, in which the time taken to respond whether a target is present or absent in the display is plotted against the number of elements in the display. The dashed lines (disjunction) show the results for cases where the target can be identified through a single feature (colour or shape), and the solid lines (conjunction) show the results from a feature-conjunction search.
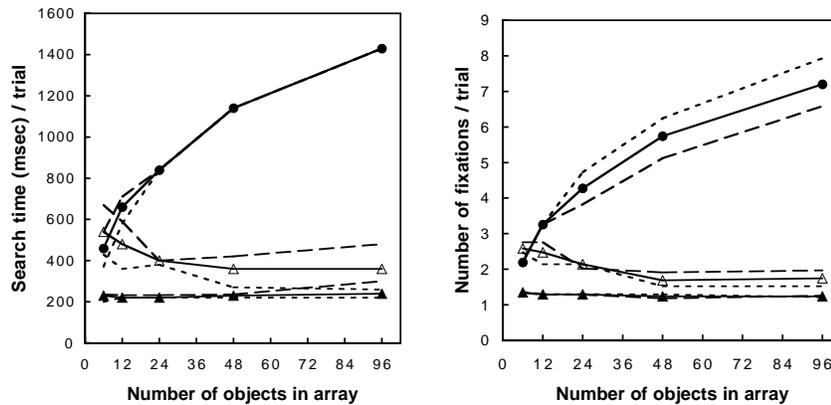
FIGURE 5.4. Results from the experiment of Motter and Belky (1998a) in which monkeys carried out a simple search task (triangle data points) or a conjunction search task looking for a target with a particular colour and orientation (circle data points). The left-hand plot shows the search function, plotting the average time taken to move the eyes to the target against the size of the display. The right-hand plot shows the average number of fixations before the target was located.

Plots similar to those of Figure 5.3 have been produced many times. Figure 5.4 shows a recent one from Motter and Belky (1998a). The left-hand plot shows a standard search function, with the triangles representing feature searches and the circles conjunction searches. The participants in this study were trained rhesus monkeys. The displays contained a target on every trial and the search time represents the time for the animals to 'find' the target with their eyes. The right hand plot shows the mean number of fixations before the target was located in this way. The search function for the conjunction task is very comparable with those found when humans carry out a similar task, although the search rate is somewhat faster.

It is clear that the variation in the number of eye fixations is highly correlated with the variation in search times. Similar correlations have been shown in human search (Binello, Mannan and Ruddock, 1995; Zelinsky and Sheinberg, 1997). Conjunction search thus involves serial overt scanning with the eyes. Invoking the assumption discussed earlier about self-terminating search, the data imply that a number of items (approx. 7-8) are dealt with on each fixation to give a search rate of 20 msec/item.

Are the items processed within each fixation dealt with in parallel, or is there a process of covert attentional scanning that scans each item serially? Several pieces of evidence suggest the former. First, both theoretical and empirical work in the Treisman tradition have supported parallel processing of a small number of items (Pashler, 1987; Eckstein, 1998) and Treisman herself has been ready to support this (Treisman and Sato, 1990). Second, more detailed studies of eye movements in search provide support for the same conclusion.
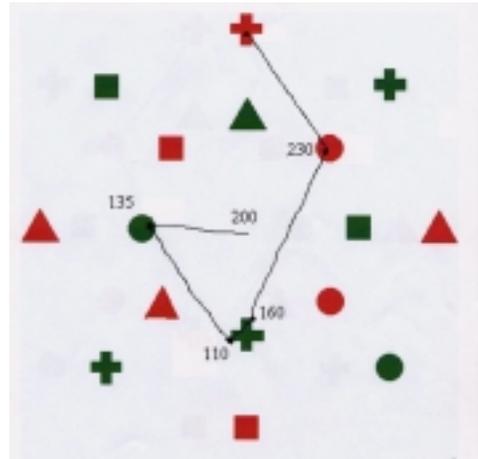
FIGURE 5.5. Feature-conjunction search task of the type studied by Findlay (1997), together with a record of a subject's eye-scan data. The target in this display was a cross located in the centre-top of the display, and subjects were instructed to move their eyes to the target as rapidly as possible. The figures show the fixation times in milliseconds for fixations before the target was reached.

In our laboratory, we have studied a task of feature-conjunction search using displays of the type shown in Fig. 5.5. Subjects were required to move their eyes from the central position to locate a prespecified target. In displays with the target in the inner ring, three out of four subjects tested showed high accuracy in making a single saccade to the target (Findlay, 1997). The latency of these saccades was in the range 200-250 msec, not significantly longer than situations that involved a search for a simple pop-out feature. If the target was in the outer ring, only on a small percentage of occasions was the first saccade directed to it. On the remainder, the saccade was nearly always directed to a distractor in the inner ring. The latency of these erroneous saccades was slightly shorter (218 msec) than those of correct saccades (228 msec). Erroneous saccades were more likely to be directed to distractors sharing a feature with the target than onto distractors sharing neither target feature.

All of these findings are readily explained on the assumption that a number of display elements are processed in parallel during the first fixation but are less easy to explain with a model in which covert attention scans around the display prior to any overt eye movement. In particular, such a serial scanning model would appear to predict that the latencies for incorrect saccades would be longer than those for correct saccades, on the basis of any reasonable hypothesis linking covert to overt scanning. Motter and Belky (1998a,b) also found that the latency of saccades that led to immediate target capture was no different from that in the remainder of cases and used this fact to argue against serial covert scanning. This parallel-processing model also accords with current work on the neurophysiology of visual search (Schall, 1995; Schall and Hanes, 1998; Schall and Thompson, 1999).

The tasks we have described previously were characterised by peripheral-preview effects. There is some indication in visual search that peripheral preview plays a role. This is shown by the occurrence of extremely short fixations (<100 msec) found when eye-scanning records during search are analysed. We have found tat these short fixations occur regularly (Findlay, Gilchrist, and Brown, 1998, 2000) and they have also been reported by McPeek, Keller, and Nakayama (1999). Indications that peripheral preview is involved come from the finding that the short fixations only occur when the eyes land on (or near) a distractor, never when the eye lands on a target. Occasions when the eye lands on or near a target are often followed by a further saccade, usually a small corrective movement. However, these saccades are preceded by a fixation of normal length (100-180 msec). Because the very brief fixations are too short for visual information to be taken in during the fixation, the decision to make the second saccade must have been already made before the first movement occurred; that is, a pair of "pipelined" saccades is programmed. Because the decision is dependent on the nature of the material at the saccade destination, peripheral preview is involved[5].

Work in the area of visual search thus produces convergent findings with those previously discussed. During a fixation, visual information is processed in parallel with enhanced processing at the area of the following saccade location. There is no support for sequential intra-fixational scanning by covert attention.

5.6   Rethinking Covert Attention

Whenever an eye movement directs the gaze to a location in the visual field, the enhanced visual resolution found in foveal vision becomes available to analyse material at this new location. If attention is moved covertly, this resolution benefit does not operate. Can any complementary benefits be found for covert shifts of attention?

One suggestion might be that the locus of covert attention could be moved around the visual scene more rapidly than the eyes. We have noted cases where researchers have looked for evidence of a fast rate of covert scanning and found no evidence to support such a phenomenon. Further estimates of the rate at which covert attention can be moved between items on a display have been reviewed by Egeth and Yantis (1997). A wide variety of estimates are described, ranging from a few tens of milliseconds per item to several hundred milliseconds per item. Some of the faster estimates come from the slope of search functions and in the previous section we have already argued that alternative interpretations of these data may be made. Other fast estimates (Saarinen and Jules, 1991) come

---

[5]Peripheral preview might be involved in two ways. First, it is possible that information from the saccade destination is acquired at a point in time during the programming of the first saccade too late to cancel it but in time to produce a saccade pair. The second possibility is that the peripheral preview of the *target* prevents a second saccade from occurring that would otherwise have been made. Further work will be necessary to distinguish between these possibilities.

from rapid serial visual presentation (RSVP). When a set of digits are exposed sequentially, but extremely rapidly, the order of the digits can often be reported. This may be used to argue that exogenously driven attention can be moved rapidly (although Egeth and Yantis note certain caveats). Theeuwes, Kramer, Hahn and Irwin (1998) have also produced data showing that exogenous factors may operate during the preparatory stage of an eye movement, but at a location different from that of the saccade destination. During the latency period whilst an observer was preparing a saccade to one target destination, a new visual stimulus was presented at a different location to the saccade target. Frequently this distractor 'captured' the saccade, which was directed to the distractor rather than to the intended target. However, on those occasions in which the eyes were not so captured, improved visual discrimination was found at the location of the distractor.

In the case of endogenous movements of covert attention, estimates of the speed of redeployment are much slower. Duncan, Ward, and Shapiro (1994) and Sperling and Weichselgartner (1995) used variants of the RSVP technique to assess the speed with which endogenous (centrally driven) attention can be relocated. Their estimates (300 - 500 msec) are very much slower and comparable with the speed at which overt eye movements can be made.

We thus argue that there is no convincing evidence that covert attention can move rapidly enough to produce sequential intrafixational scanning unless fixations are artificially prolonged. Covert attention appears to operate in close conjunction with actual saccadic eye movements by preprocessing the material at the location of the forthcoming movement in a way that subsequently allows faster processing of the material when fixated. In the case of reading, this speeds up the rate of information intake by some 10-15%. In this task at least, the absence of covert attention would only impair task performance to a small extent. The loss of the ability to make eye-scanning movements, however, would be devastating.[6]

5.7   Conclusion

We have examined three areas in which visual tasks are carried out using eye scanning. In reading, scene viewing and free visual search, the eyes make saccadic movements several times each second. We have argued that, in these tasks, there is no evidence for additional spatial scanning by movements of a covert attention spotlight. However, covert attention is involved in these tasks, as shown by the asymmetry of the reading perceptual span and the phenomenon of peripheral preview. Nevertheless, the overt movements of the eyes are primary and the

---

[6]The eye scanning need not necessarily be mediated by eye movements of the conventional type however. We have studied an individual with congenital ophthalmoplegia (Gilchrist, Brown and Findlay, 1997; Gilchrist, Brown and Findlay, & Clarke 1998). She has no functional eye muscles but can nevertheless use her vision to read and carry out other tasks. Saccade-like movements of the head substitute for a normal individual's eye movements.

covert processes only operate in association with these movements.

References

Aloimonos, J., Bandopadhay, A. and Weiss, I. (1988). Active vision. *Int. J. Comp. Vis.*, 1:333-356.

Anstis, S. M. (1974). A chart demonstrating variations in acuity with retinal position. *Vis. Res.*, 14:589-592.

Ballard, D. H. (1991). Animate vision. *Artif. Intel.*, 48:57-86.

Ballard, D. H., Hayhoe, M. M., Pook, P. K. and Rao, R. P. N. (1998). Deictic codes for the embodiment of cognition. *Behav. Brain Sci.*. 20:66-80.

Binello, A., Mannan, S. and Ruddock, K. H. (1995). The characteristics of eye movements made during visual search with multi-element stimuli. *Spatial Vis.*, 9:343-362.

Blanchard, H. E., McConkie, G. W., Zola, D. and Wolverton, G. S. (1984). Time course of visual information utilization during fixations in reading. *J. Exp. Psych. Human Percept. Perf.*, 10:75-89.

Cave, K. R. and Bichot, N. P. (1999). Visuospatial attention: beyond the spotlight model. *Psychonomic Bull. and Rev.*, 6:204-223.

Churchland, P. S., Ramachandran, V. S. and Sejnowski, T. J. (1994). A critique of pure vision. In C. Koch and J. L. Davis (Eds.) *Large Scale Neuronal Theories of the Brain*, pp. 23-60, Cambridge, MA: MIT Press.

Deubel, H. and Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vis. Res.*, 36:1827-1837.

Duncan, J., Ward, R. and Shapiro, K. (1994). Direct measurement of attention dwell time in human vision, *Nature*, 369:313-315.

Eckstein, M. P. (1998). The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychol. Sci.*, 9:111-118.

Egeth, H. E. and Yantis, S. (1997). Visual attention: control, representation, and time course. *Annu. Rev. Psych.*, 48:269-297.

Findlay, J. M. (1997). Saccade target selection in visual search. *Vis. Res.*, 37:617-631.

Findlay, J. M., Gilchrist, I. D. and Brown, V. (1998). Saccades in search: the role of trans-saccadic memory. *Invest. Ophth. and Vis. Sci.*, 39:S165.

Findlay, J. M. and Walker, R. (1999). A model of saccadic eye movement generation based on parallel processing and competitive inhibition. *Behav. and Brain Sci.*, 22:661-674.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

Gilchrist, I. D., Brown, V., Findlay, J. M. and Clarke, M. P. (1998). Using the eye movement system to control the head. *Proc. Roy. Soc. B*, 265:1831-1836.

Gough, P. (1972). One second of reading. In J. F. Kavanagh and I. G. Mattingley (Eds.) *Language by Ear and by Eye. The Relationship Between Speech and Reading*. pp. 335-358, Cambridge, MA: MIT Press.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.) *Perception*. Volume 5 of the Vancouver Studies in Cognitive Science. New York, NY: Oxford University Press.

Henderson, J. M. and Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: implications for attention and eye movement control. *J. Exp. Psych.: Learning, Memory and Cognition*, 16:417-429.

Henderson, J. M. and Ferreira, F. (1993). Eye movement control during reading: fixational measures reflect foveal but not parafoveal processing difficulty. *Canadian J. Exp. Psych. - Revue Canadienne de Psychologie Expérimentale*, 47:201-221.

Henderson, J. M. and Hollingworth, A. (1998). Eye movements during scene viewing: an overview. In G Underwood (Ed.) *Eye Guidance in Reading and Scene Perception*. pp. 269-293, Amsterdam: North-Holland/Elsevier.

Henderson, J. M. and Hollingworth, A. (1999). High-level scene perception. *Ann. Rev. Psych.*, 50:243-271.

Henderson, J. M., Pollatsek, A. and Rayner, K. (1989). Covert visual attention and extrafoveal information use during object identification. *Percept. and Psychophys.*, 45:196-208.

Hoffman, J. E. and Subramaniam, B. (1995). The role of attention in saccadic eye movements. *Percept. and Psychophys.*, 57:787-795.

Inhoff, A. W. and Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: effect of word frequency. *Percept. and Psychophys.*, 40:431-439.

Kennedy, A. (1998). The influence of parafoveal words on foveal inspection time: evidence for a processing trade-off. In G. Underwood (Ed.) *Eye Guidance in Reading and Scene Perception*. pp. 149-179. Elsevier: Amsterdam.

Kennison, S. M. and Clifton, C. (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: implications for eye movement control. *J. Exp. Psych.: Learning, Memory and Cognition*, 21:68-81.

Kingstone, A. and Klein, R. M. (1993). Visual offset facilitates saccade latency: does pre-disengagement of attention mediate this gap effect? *J. Exp. Psych., Human Percept. and Perf.*, 19:251-265.

Kowler, E., Anderson, E., Dosher, B. and Blaser, E. (1995). The role of attention in the programming of saccades. *Vis. Res.*, 35:1897-1916.

Land, M. F. (1995). The functions of eye movements in animals remote from man. In J. M. Findlay, R. Walker and R. W. Kentridge (Eds.) *Eye Movement Research: Mechanisms, Processes and Applications*. pp. 63-76. Elsevier: Amsterdam.

Land, M. F., Mennie, N. and Rusted, J. (1999). The roles of vision and eye movements in the control of activities of everyday living. *Percept.*, 28: 1311-1328.

Liversedge, S. P. and Findlay, J. M. (2000). Saccadic eye movements and cognitive science. *Trends in Cog. Sci.*, 4:6-14.

Loftus, G. R. and Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *J. Exp. Psych., Human Percept. and Perf.*, 4:565-572.

McClelland, J. L. and Mozer, M. C. (1986). Perceptual interactions in two-word displays: familiarity and similarity effects. *J. Exp. Psych., Human Percept. and Perf.*, 12:18-35.

McConkie, G. W. and Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Percept. and Psychophys.*, 17:578-586.

McPeek, R. M., Keller, E. L. and Nakayama, K. (1999). Concurrent processing of saccades. *Behav. and Brain Sci.*, 22:691-692.

Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: evidence for parallel programming of saccades. *J. Exp. Psych., Human Percept. and Perf.*, 5:667-682.

Motter, B. C. and Belky, E. J. (1998a). The zone of focal attention during active visual search. *Vis. Res.*, 38:1007-1022.

Motter, B. C. and Belky, E. J. (1998b). The guidance of eye movements during active visual search. *Vis. Res.*, 38:1805-1818.

Müller, H. J. and Rabbitt, P. M. A. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *J. Exp. Psych., Human Percept. and Perf.*, 15:315-330.

Nakayama, K. (1992). The iconic bottleneck and the tenuous link between early visual processing and perception. In C. Blakemore (Ed) *Vision: Coding and Efficiency*, Cambridge University Press.

Nakayama, K. and Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vis. Res.*, 29: 1631-1647.

Nelson, W. W., and Loftus, G. R. (1980). The functional visual field during picture viewing. *J. Exp. Psych., Hum. Learn. Mem.*, 6:391-399.

O'Regan, J. K. (1992). Solving the 'real' mysteries of visual perception: the world as outside memory. *Can. J. Psych.*, 46:461-488.

Pashler, H. (1987). Detecting conjunction of color and form: re-assessing the serial search hypothesis. *Percept. and Psychophys.*, 41:191-201.

Pashler, H. (Ed.) (1998) *Attention*. Hove: Psychology Press.

Pollatsek, A., Rayner, K. and Collins, W. E. (1984). Integrating pictorial information across saccadic eye movements. *J. Exp. Psych., Gen.*, 113:426-442.

Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psych.*, 32:3-25.

Posner, M. I., Nissen, M. J. and Ogden, M. C. (1978). Attended and unattended processing modes: the role of set for spatial location. In H. L. Pick and I. J. Saltzman (Eds), *Modes of Perceiving and Processing Information*. pp. 137-157. Hillsdale: Lawrence Erlbaum.

Posner, M. I., Walker, J. A., Friedrich, F. J. and Rafal, R. D. (1984). Effects of parietal lobe injury on covert orienting of visual attention. *J. Neurosci.*, 4:1863-1874.

Rayner, K. (1984). Visual search in reading, picture perception and visual search. A tutorial review. In H. Bouma and D. Bouwhuis (Eds.) *Attention and Performance X*. pp. 67-96. Hillsdale, NJ: Lawrence Erlbaum.

Rayner, K. (1998). Eye movements in reading and information processing. 20 years of research. *Psychol. Bull.*, 124:372-422.

Reichle, E. D., Pollatsek, A., Fisher, D. F. and Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychol. Rev.*, 105:125-147.

Rensink, R., O'Regan, J. K. and Clark, J. J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychol. Sci.*, 8:368-373.

Rensink, R. A. (2000). Seeing, sensing, and scrutinizing. *Vis. Res.*, in press.

Rizzolatti, G., Riggio, L., DaScola, I. and Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia*, 25:31-40.

Rizzolatti, G., Riggio, L. and Sheliga, B. M. (1994). Space and selective attention. In C. Umiltá and M. Moscovitch (Eds.) *Attention and Performance XV*. pp 231-265. Cambridge, MA: MIT Press.

Rovamo, J., Virsu, V. and Näsänen, R. (1978). Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. *Nature*, 271:54-55.

Saarinen, J. and Julesz, B. (1991). The speed of attentional shifts in the visual field. *Proc. Nat. Acad. Sci.*, 88:1812-1814.

Saida, S. and Ikeda, M. (1979). Useful visual field size for pattern perception. *Percept. and Psychophys.*, 25:119-125.

Schall, J. D. (1995). Neural basis of saccade target selection. *Rev. Neurosci.*, 6:63-85.

Schall, J. D. and Hanes, D. P. (1998). Neural mechanisms of selection and control of visually guided eye movements. *Neural Networks*, 11:1241-1251.

Schall, J. D. and Thompson, K. G. (1999) Neural selection and control of visually guided eye movements. *Ann. Rev. Neurosci.*, 22:241-259.

Shiori, S. and Ikeda, M. (1989). Useful resolution for picture perception as a function of eccentricity. *Percept.*, 18:347-361.

Schroyens, W., Vitu, F., Brysbaert, M. and d'Ydewalle, G. (1999). Eye movement control during reading: foveal load and parafoveal processing. *Q. J. Exp. Psych.*, 52A:1021-1046.

Shepherd, M., Findlay, J. M. and Hockey, G. R. J. (1986). The relationship between eye movements and spatial attention. *Q. J. Exp. Psych.*, 38A:475-491.

Slowiaczek, M. L. and Rayner, K. (1987). Sequential masking during eye fixations in reading. *Bul. Psychonom. Soc.*, 25:175-178.

Sperling, G. S. and Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. *Psych. Rev.*, 102:503-532.

Styles, E. (1997). *The psychology of attention.* Hove: Psychology Press.

Theeuwes, J., Kramer, A. F., Hahn, S. and Irwin, D. E. (1998). Our eyes do not always go where we want them to go. *Psychol. Sci.*, 9:379-385.

Treisman, A. and Gelade, G. (1980). A feature integration theory of attention. *Cog. Psych.*, 12:97-136.

Treisman, A. and Sato, S. (1990). Conjunction search revisited. *J. Exp. Psych., Human Percept. and Perf.*, 16:459-478.

Van Diepen, P. M. J., Wampers, M. and d'Ydewalle, G. (1998). Functional division of the visual field: moving masks and moving windows. In G. Underwood (Ed.) *Eye Guidance in Reading and Scene Perception*. Chapter 15, pp. 337-355. Amsterdam: Elsevier.

Wertheim, T. (1894). Über die indirekte Sehschärfe. Z. Psych. 01 Phys. Sinnesorgans, 7:121-187.

Wright, R. D. (Ed) (1998). *Visual Attention*. New York, NY: Oxford University Press.

Yeshurun, Y. and Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, 395:72-75.

Zelinsky, G. J. and Sheinberg, D. L. (1997). Eye movements during parallel-serial visual search. *J. Exp. Psych. Human Percept. and Perf.*, 23:244-262.