Visual Attention

Intermediate article

Ronald A Rensink, University of British Columbia, Vancouver, British Columbia, Canada

CONTENTS

Visual attention is the factor controlling the selective access and integration of visual information.

INTRODUCTION

Although much of vision appears to be effortless and all-encompassing, nevertheless there are limits to what it can do. For example, consider air traffic control, where it is imperative to keep track of all moving items in a display (corresponding to the airplanes in an airspace). If only a single item is present, it can generally be tracked without problem. It is also possible to track four or five items simultaneously, although some effort is needed to do so. However, for 20 or 30 items, even a maximal effort will not suffice, and the task must be shared among several controllers. What appears to be happening in such cases is that visual perception is constrained by a consciously controlled factor within the observer – a factor that enables certain types of processing to take place, but which is limited in the extent to which it can be applied. This factor is termed *visual attention*.

Interestingly, although most observers immediately know what to do when asked to 'pay attention' to a stimulus, it has been rather difficult to give this an objective characterization. Indeed, until recently there was no general consensus on the basic function of attention, and at various times it was associated with such things as clarity of perception, intensity of perception, consciousness, and selection.

SELECTION

During the past few decades, considerable progress has been achieved by focusing on *selection* as the basic function of visual attention. Two types of selection are of particular importance. The first is *selective access* (i.e., allowing only certain parts or properties to be sent on to later processes). It was originally believed that selective access protected processors at higher levels from being overwhelmed by too much information. However, more recent research has tended to view selective access as a way to delimit control of various actions (e.g., focusing on the locations of items that are to be grasped).

The second type of selection is *selective integration* (i.e., combining selected parts or properties into structures that then form the basis of further processing). For example, three adjoining lines could be combined into a complete figure. This figure (and not the lines themselves) might then provide the basis for subsequent control of grasping. It was initially believed that such integration had to be selective in order to make good use of a limited amount of processing 'resource'. However, more recent research has tended to view selective integration in terms of the selective *coordination* of the outputs of multiple processes.

According to this more recent view, therefore, visual attention is not a unitary faculty. Instead, it is simply the selective control of information in the visual system, achieved in various ways by various processes. When considered from this perspective, several of the unresolved issues in earlier treatments of visual attention simply vanish. One such example is the issue of whether selection is 'early' or 'late' (i.e., whether it acts on simple, precategorical structures or more complex ones). Given that selective control may be carried out by a number of systems, there may not be a single site where attention acts, so this issue becomes meaningless.

Since a complete understanding of visual attention is still a long way off, this article will survey only the major behavioral techniques that have been used for its exploration and several of the more important results that have been obtained. Furthermore, it will focus entirely on the purely visual aspects of the processes involved (i.e., on how the stimuli themselves are handled, rather than how responses to them are generated). Issues such as the sequencing of multiple responses are considered to involve central control at higher levels, so will not be discussed here. (*See* **Attention**)

VISUAL ORIENTING

Consider the situation where a driver is at a red traffic light, anxiously waiting for it to turn green. Here the stop light becomes the center of the driver's perceptual world, and when the light turns green the driver will respond almost immediately. This is an example of *visual orienting*. More generally, orienting can be defined as the alignment of the perceptual apparatus to allow optimal perception of what is happening (or expected to happen). Judging by the near universality of its occurrence, orienting is highly important to our survival in the world.

Orienting is usually studied by determining how well an observer can use advance information about a target item (e.g., position) to improve performance. Typically, the observer tries to detect, identify or locate the target. Two performance measures are generally used, namely *accuracy* and *response time* (RT). The effectiveness of orienting is measured by comparing performance when the observer has advance knowledge with performance when no such knowledge is available (Figure 1).

Overt Orienting

Perhaps the simplest form of orienting is the *overt orienting* of the eyes towards a stimulus. This is not usually a matter of eye movements alone, for the torso and head also contribute to it. Thus, for example, there is a reflexive orienting of the head towards items that suddenly appear in peripheral vision; this behavior is so basic that it is found even in newborn infants.

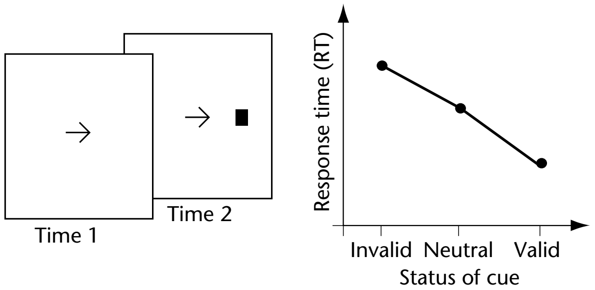

Figure 1. Covert orienting. When a cue indicates the location of the stimulus in a subsequent display, responses are faster and more accurate if the cue is valid (i.e., if the arrow points correctly to the location of the stimulus) than if the cue is uninformative (if it contains only the stem of the arrow). This improvement is considered to be due to the covert orienting of attention. Note that performance is impaired if the cue is invalid (the arrow points to the wrong location), presumably because attention needs to be reoriented from the incorrect location.

The net result of these movements is that the eye fixates on some part of the world. Since the resolution of a human retina is highest at its center, estimates of shapes or positions are most accurate for those items at the fixated location. As such, overt orienting is essentially a form of selective access. (*See* **Eye Movements**)

Covert Orienting

Observers can detect a target faster and more accurately if they are presented with a cue containing advance information about its location. Given that eye movements can be prevented, such improvement indicates the existence of a *covert orienting* performed by neural mechanisms. Enhancement begins within 50 ms of the cue and increases thereafter, reaching a peak about 200 ms after cue onset.

One explanation for this facilitation is that location is selected via a *spotlight of attention* (Posner *et al.*, 1980), which allows input only from the area that it 'lights'. Studies of this have been largely based on interference caused by irrelevant items, which can cause performance to degrade if near the target. Interference effects indicate that the spotlight covers about 1° of visual angle in central vision, although it can 'zoom out' to cover a larger area if necessary. The minimum area increases with eccentricity from the fovea, and appears to be greater in the upper visual field.

The relationship between covert and overt orienting is not a direct one. Attention does not have to be given to the fixated location – people can move their attention without moving their eyes. Moreover, although some form of attention is needed to select the target of an eye movement, this selection need not be accompanied by a withdrawal of attention from other items.

Space-based Versus Object-based Selection

Although there is considerable agreement that orienting is concerned with selection, there is less agreement about what exactly is being selected. For overt orienting, the situation is straightforward – orienting involves aligning the eyes to a particular two-dimensional location in space, and if the eyes are properly coordinated, they can select a particular depth as well. Similarly, covert orienting need not be limited to a two-dimensional location in the visual field, but can also be based on three-dimensional depth.

An important issue is whether covert orientation selects for a particular location in space (that happens to be part of some object), or for a particular object (that happens to be at some location), or for both. When two overlapping items are presented, it is easier to report two properties from a single item than to report one property from each of the pair, indicating that covert orienting can be influenced by object structure (Duncan, 1984). The extent to which selection depends on spatial location and object structure has not yet been fully elucidated.

VISUAL SEARCH

Another task that is important in everyday life is *visual search*. For example, when looking for a friend in a crowd, an observer must check each person in turn until the friend is seen. More generally, a search task involves scanning through various items (cars, keys, people, etc.) until the desired item is found.

In studies of visual search, observers typically attempt to detect, identify, or locate a given (target) item among a set of other (distractor) items in a visual display. One of two performance measures are generally used: (1) accuracy on briefly presented displays, or (2) RT on displays that are continually visible. The central issue is the way in which performance is affected by the properties of the items in the display (Figure 2).

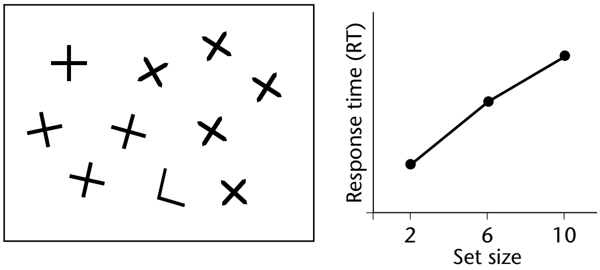

Figure 2. Visual search. The observer is asked to detect, identify, or locate a unique target placed among a set of distractor items. When the target contains a unique feature, performance is largely independent of set size (the number of items in the display). When the target is a strong conjunction of the distractor's features (when it contains the same basic features, such as the same line segments), performance is strongly dependent on the number of elements.

Pop-out

It is easy to see a single yellow dot among an array of blue dots. More generally, a target with a unique property can often be detected rapidly and with little dependence on the number of items in the display. Such *pop-out* is believed to indicate the presence of a distinctive property (or *feature*) in the target item. According to this view, various features (e.g., color or orientation) are computed rapidly and in parallel across the visual field. A unique feature will be *salient*, and so attract attention to its location, causing it (and its properties) to be selectively accessed. Salience can also arise via *differences* between adjacent features. An item will pop out if it is the only one with an orientation that differs greatly from those of its neighbors, even if its orientation occurs elsewhere in the display.

Support for this view comes from the existence of *search asymmetries*, where a switch in the role of target and distractor items strongly affects performance. For example, a 'Q' among a set of 'O's pops out, whereas an 'O' among a set of 'Q's does not. This can be explained by the 'Q' having a unique feature (the tail) that is not present in the 'O's, and thereby becoming salient. Since 'O' is distinguished from 'Q' by the absence of a feature, it can never be salient. Search asymmetries yield a powerful method for identifying features – pop-out if some feature in the target does not exist in the distractors, and a slower search otherwise.

Emergent Features

Many features have been found that are simple properties of the image (e.g., orientation, size, and color). However, other features are *emergent*, being derived from the image in a relatively complex way. For example, when targets are triangles and distractors are forked-shaped items constructed of the same line segments, the triangles pop out, indicating that some property of the line arrangement (presumably closure) acts as a feature. The existence of such features is due to processes operating *preattentively* (i.e., prior to the application of selective integration).

Several types of preattentive process are known to exist. These include grouping between items, grouping within items, and completion of occluded items. Features have also been found that are based on the (recovered) three-dimensional scene rather than the two-dimensional retinal image (e.g., surface convexity and three-dimensional orientation).

Although preattentive processes can make search easy under some conditions (i.e., if the target contains an emergent feature), they can also sometimes make it more difficult. For example, if target and distractors are both line configurations with distinctive line segments, search can still be difficult if the overall length of the configurations is the same. This indicates that selective access is easiest not for simple attributes of the image (e.g., individual line segments), but rather for the more complex structures formed by preattentive processes.

Conjunctions

An important type of search is the *conjunction task*, where the target and distractor sets contain the same features but differ in how they are assembled. Conjunctions can be either *weak* (at the level of sets) or *strong* (at the level of items). An example of a weak conjunction is search for a red vertical line among green vertical lines and red horizontal lines. Here the set of distractors contains all of the features in the set of targets. However, at the level of individual items, the target remains unique. In contrast, in a strong conjunction each target item would contain the same features as a distractor item (e.g., an 'L' among 'T's).

If highly discriminable properties are involved, search for weak conjunctions can be relatively easy. This has been explained in terms of *guided search*, where a guidance mechanism either inhibits items with non-target features, or excites items with target features. Once these mechanisms have reduced the number of possible items, the remainder can be searched more easily – if a target has a unique feature, it will pop out. Thus, for example, if searching for a red vertical line, all green items can be inhibited, leaving the single remaining (red) vertical line to pop out. According to this view, therefore, search for weak conjunctions may simply involve selective access.

In a strong conjunction, items differ only in the way in which their features are arranged, and so guidance cannot be used. Search for strong conjunctions is often difficult, requiring at least 30–50 ms for each item. This has been taken to indicate the serial application of a spotlight of attention that binds the features of each item together, integrating them into an *object file* (i.e., a coherent collection of features) (Kahneman *et al*., 1992). As such, search for strong conjunctions is primarily an issue of selective integration.

Several issues with regard to strong conjunctions are currently unresolved. Although it has been proposed that selective integration acts on one item at a time, at a rate of 50 ms per item, it has also been proposed that it might act on clusters of four or five items, at a rate of 200–300 ms per cluster. Indeed, a serial mechanism may not even be involved – search could be carried out via a parallel process with a processing speed dependent on the number of items in the display. It is also not known what happens to object files after attention has been withdrawn. Although it has been suggested that object files exist for some time, it may be that they dissolve almost immediately (Wolfe *et al*., 2000).

INDUCED FAILURES OF PERCEPTION

Suppose that we are looking for a pencil, and we believe that it is yellow. If the pencil is indeed yellow, the attentional control based on our belief can facilitate its perception. However, if the pencil is blue, this same control can be detrimental – indeed, we may completely fail to see a blue pencil that is directly in front of us. More generally, failures of selective access and selective integration can provide important insights into the operation of the mechanisms involved. Such studies can also provide information about what aspects of perception continue when selective access and selective integration fail.

Change Blindness

A number of studies have examined the ability of an observer to detect, identify, or localize the occurrence of a change in a display. Performance is measured in one of two ways, either by accuracy on a pair of displays containing a single change, or by RT on displays that continually alternate between an original and a modified image. In both cases, observers often experience great difficulty in reporting the presence of a change that is made simultaneously with another event, such as an eye movement, blink, or flash (Figure 3).

This *change blindness* has been regarded as evidence that attention is needed in order to see change (Rensink, 2000). According to this view, when the local signals due to the change are swamped (or otherwise neutralized) so that they no longer capture attention, a time-consuming attentional scan of the display is needed. The observer will be blind to the change until the appropriate item is attended.

Although a selective process of some type is involved in perceiving change, the nature of this selectivity has not yet been established. One possibility is that a complete representation of the original display is formed, and that change blindness results from a limited ability to compare this representation with the current image. According to this view, the key factor is selective access to a

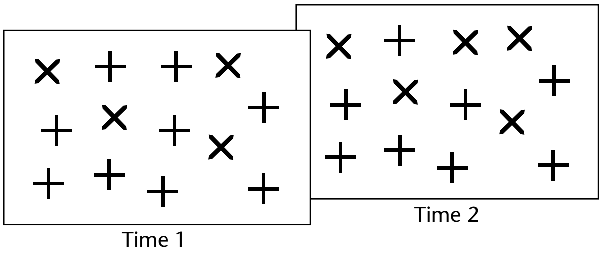

Figure 3. Change blindness. The observer is asked to detect, identify, or locate a unique change among a set of distractor items. This change can either be shown just once, or it can be continually presented by alternating between the two displays. Performance is extremely poor, provided that successive displays are separated by an interval of at least 80 ms.

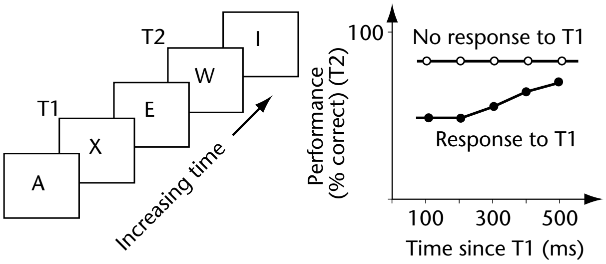

Figure 4. Attentional blink. When the observer is asked to report on two items (e.g., the colors of the consonants in a stream of letters), performance on the second of these (T2) is impaired if it is made within several hundred milliseconds of the occurrence of the first (T1). However, this degradation does not occur if no response is required of the first letter (e.g., if the observer reports only the color of the W).

comparison mechanism. Another possibility is that a complete representation of the display is never formed – representations of unattended items are volatile, and are simply replaced by representations of subsequent stimuli. Here attention is believed to act much as it does in visual search for strong conjunctions, integrating selected items over time and space so that they have a degree of spatiotemporal continuity, which then allows them to be seen to change.

Attentional Blink

Another phenomenon involving the failure to perceive stimuli is the *attentional blink*. Here a stream of successive items (usually letters) is presented at a location, and the observer attempts to report the presence of an item that has been designated in some way (e.g., by its color or its identity). Performance is measured by the accuracy of response. Whereas it is easy to detect a target when items appear at a rate of less than about 10 items per second, it becomes much more difficult to do this if the observer also has to respond to a target (T1) appearing earlier in the stream. Somehow, responding to T1 induces a 'blink' that makes it difficult for the observer to see – or at least respond to – a second target (T2) for the next few hundred milliseconds (Figure 4).

Control experiments have shown that this blink is not due to perceptual, memory or output limitations. Instead, it appears to result from attention being given to T1, leaving subsequent stimuli unattended and thus vulnerable to replacement by the items that follow them. Evidence in favor of this explanation is provided by the fact that T2 is difficult to report only if it is followed by another item. (Note that this replacement mechanism is similar to that proposed for change blindness.) Unseen T2 items can also facilitate the processing of semantically related items (Shapiro *et al.*, 1997), providing further evidence that a considerable degree of processing can take place in the absence of awareness (and thus presumably in the absence of attention).

The time course of the blink itself has also been taken to correspond to an attentional *dwell time* (the time needed to integrate the properties of an attended object into a coherent form). According to this view, once attentional processing has started on T1, it cannot be simply halted when T2 is encountered, but must run its course. The duration of this blink appears to be about 300–400 ms.

Inattentional Blindness

A different approach to exploring attention involves asking observers to attend to a given event, and then introducing an unexpected item at some point. Performance is measured by the accuracy of responses to questions about the intrusive stimuli. Interestingly, observers often have great difficulty in reporting such stimuli, even though the items are easily seen if they are expected. This failure to report unexpected – and therefore unattended – stimuli is called *inattentional blindness* (Figure 5).

Early studies were *selective*, requiring the observer to attend to a subset of the stimuli (e.g., to attend only to the white-shirted or black-shirted players in a basketball game). Observers often had difficulty noticing the appearance of the unexpected stimulus under these conditions. Although superimposed images were used initially, later studies showed that these failures occurred also even when the stimuli were elements of a single scene and the unexpected stimulus was a person

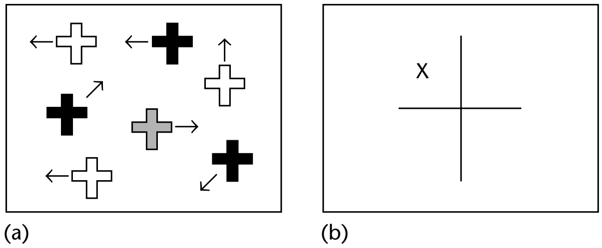

Figure 5. Inattentional blindness. (a) In a *selective* task, the observer must select a subset of the items in the display (i.e., track a set of moving black crosses among a set of white ones). While engaged in such selection, the observer will generally be blind to the occurrence of an unexpected item, even if it is unique (e.g., a gray cross moving in a unique direction). (b) In a *nonselective* task, the observer is presented with a pair of lines and asked to judge which of the two lines is longer. After a number of such tests, a display is presented that contains an extra element (e.g., a letter). Again, detection of these unexpected elements is generally quite poor.

dressed as a gorilla (Simons and Chabris, 1999). For selective tasks, inattentional blindness appears to be at least partly due to observers inhibiting the features of the irrelevant stimuli. If this were so, it would illustrate a failure of selective access.

Mack and Rock (1998) introduced a *nonselective* variant, where attention could be given to all stimuli that were present before the appearance of the unexpected item. Here observers attended to an overlapping pair of lines (one horizontal and one vertical) and had to judge which line was longer. Again they often failed to see the unexpected item under these conditions, even when it was at the center of fixation. It is still unknown whether inhibition is also an important factor in nonselective tasks.

Although observers may not report seeing an unexpected item, such items can still influence conscious perception. For example, surrounding lines can induce a length illusion in the test lines that are perceived, even if the surrounding lines themselves are not reported. Again this indicates that representations of considerable sophistication are constructed in the absence of awareness (and thus presumably in the absence of attention).

ATTENTIONAL CONTROL

Whereas much of the research on visual attention has centered on the nature of the mechanisms involved (e.g., their speed, or what they select), an equally important issue concerns the way in which these mechanisms are controlled. Two types of control appear to exist, namely goal-driven *direction* and stimulus-driven *capture*. These behave quite differently, with direction corresponding to a slow, sustained process, and capture being faster and more transient. It has been argued that these mechanisms control different types of attentional process, direction being involved with selective access, and capture being involved with selective integration (Briand and Klein, 1987).

Attentional direction occurs, for example, when observers in an orienting task engage their attention on the stimulus to which an arrow is pointing. It also occurs in guided search, where observers select the features to be enhanced or suppressed. In both cases, observers voluntarily select the appropriate locations or features in order to facilitate performance, and they are able to refrain from this if performance is adversely affected (e.g., if cues are misleading).

In contrast, attentional capture is largely involuntary, and interference can result from any features consistent with an *attentional control setting*. For example, visual search for a unique orientation is impaired if one of the distractors has a unique color, even though color is irrelevant to the task. In this case, capture occurs because the control setting can be set only for a *difference* in some feature; the relevant type of difference (e.g., orientation) cannot be represented. Indeed, a unique feature of any kind will capture attention when this control setting has been selected.

The sudden appearance of an item has a privileged status, in that capture can occur regardless of the setting. Interestingly, the relevant factor is not the motion signal that accompanies the appearance, but rather the appearance of the object itself (Yantis and Hillstrom, 1994). However, this type of capture has its limits, in that it appears to be effective only if attention is not already engaged on some task.

RELATIONSHIP TO CONSCIOUSNESS

Although recent research views attention primarily in terms of selection, an older tradition (stemming from William James) viewed it primarily in terms of conscious perception. Traces of this older tradition persist in two functions that are currently ascribed to attention, namely selective access or integration for a process that eventually affects conscious perception, and selective entry into consciousness itself. With regard to the first of these, the relationship between attention and consciousness is unproblematic for a process that is under conscious control. However, there is increasing evidence that many actions are performed without any involvement of consciousness. Given that such actions require selection for effective operation (e.g., manual grasping may need to select items with a horizontal orientation), there may be selective mechanisms acting on processes that never involve consciousness, at least in terms of immediate control. Whether these mechanisms should be described as 'attentional' would seem to be a matter of convention. (*See* **Motor Control and Learning**)

With regard to the entry of stimuli into consciousness, there may well be selective processes with exactly this function. For example, change blindness, the attentional blink, and inattentional blindness are all phenomena in which the failure to report an otherwise highly visible stimulus could be attributed to a failure of consciousness to access the appropriate representation. However, it is not yet clear whether this is actually the case – although a failure to report could be due to a failure to consciously see a stimulus, it could also be due to a failure to remember it. It is also not yet clear whether conscious experience necessarily arises via attention – non-attentional processes may exist that can provide conscious experience of at least some aspects of the visual field. As with many of the issues relating to attention, a clearer understanding of these matters must await future developments.

References

Briand KA and Klein R (1987) Is Posner's bean the same as Treisman's glue? On the relation between visual orienting and feature integration theory. *Journal of Experimental Psychology: Human Perception and Performance* **13**: 228–241.

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

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

Mack A and Rock I (1998) *Inattentional Blindness*. Cambridge, MA: MIT Press.

Posner MI, Snyder CR and Davidson BJ (1980) Attention and the detection of signals. *Journal of Experimental Psychology: General* **109**: 160–174.

Rensink RA (2000) Seeing, sensing and scrutinizing. *Vision Research* **40**: 1469–1487.

Shapiro K, Driver J, Ward R and Sorensen RE (1997) Priming from the attentional blink: a failure to extract visual tokens but not visual types. *Psychological Science* **8**: 95–100.

Simons DJ and Chabris CF (1999) Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception* **28**: 1059–1074.

Wolfe JM, Klempen N and Dahlen K (2000) Postattentive vision. *Journal of Experimental Psychology: Human Perception and Performance* **26**: 693–716.

Yantis S and Hillstrom AP (1994) Stimulus-driven attentional capture: evidence from equiluminant visual objects. *Journal of Experimental Psychology: Human Perception and Performance* **20**: 95–107.

Further Reading

Allport A (1992) Attention and control: have we been asking the wrong questions? A critical review of twenty-five years. In: Meyer DE and Kornblum S (eds) *Attention and Performance XIV*, pp. 183–218. Cambridge, MA: MIT Press.

Braun J and Sagi D (1990) Vision outside the focus of attention. *Perception and Psychophysics* **48**: 45–58.

Cave KR and Bichot NP (1999) Visuospatial attention: beyond a spotlight model. *Psychonomic Bulletin and Review* **6**: 204–223.

He S, Cavanagh P and Intriligator J (1997) Attentional resolution. *Trends in Cognitive Sciences* **1**: 115–121.

Laberge D (1995) *Attentional Processing: the Brain's Art of Mindfulness*. Cambridge, MA: Harvard University Press.

Pashler HE (ed.) (1998) *Attention*. Philadelphia, PA: Taylor & Francis.

Simons DJ (ed.) (2000) *Change Blindness and Visual Memory*. Hove, UK: Psychology Press.

Treisman A (1985) Preattentive processing in vision. *Computer Vision, Graphics and Image Processing* **31**: 156–177.

van der Heijden AHC and Bem S (1997) Successive approximations to an adequate model of attention. *Consciousness and Cognition* **6**: 413–428.

Wright RD (ed.) (1998) *Visual Attention*. New York, NY: Oxford University Press.