Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 44 (2004) 1273–1296
Vision Research
www.elsevier.com/locate/visres

The spatial distribution of visual attention

Joetta L. Gobell [a,*], Chia-huei Tseng [a], George Sperling [a,b]

[a] Department of Cognitive Sciences, University of California, Irvine 92697, CA, USA
[b] Institute for Mathematical Behavioral Sciences and Department of Neurobiology and Behavior, University of California, Irvine, CA, USA

Received 25 November 2003; received in revised form 13 January 2004

Abstract

We use a novel search task to investigate the spatial distribution of visual attention, developing a general model from the data. Observers distribute attention to locations defined by stripes with a high penalty for attention to intervening areas. Attended areas are defined by a square-wave grating. A target is in one of the even stripes, and ten false targets (identical to the real target) are in the odd stripes; the observer must attend the even stripes and strongly ignore the odd, reporting the location of the target. As the spatial frequency of the grating increases, performance declines. Variations on this task inform a model that incorporates stimulus input, a "low pass" attentional modulation transfer function, and an acuity function to produce a strength map from which the location with the highest strength is selected. A feature-strength map that adds to the attention map enables the model to predict the results of attention-cued conjunction search experiments, and internal noise enables it to predict the outcome of double-pass experiments and of variations in the number of false targets. The model predicted performance on a trial-by-trial basis for three observers, accounting for approximately 70% of the trials. Actual trial-to-trial variation for an observer, using the double-pass method, is about 76%. For any requested distribution of spatial attention, this general model makes a prediction of the actually achieved distribution.

Keywords: Distribution of attention; Visual search; Sustained; Endogenous; Split attention

1. Introduction

Typically, attention refers to an internal state that can be rapidly modified by instructions. In practical terms, it means that observers respond differently to the same stimulus depending on their attentional state.

Spatial attention generally refers to a focus area where performance on some task is better than outside of that focus area. Here we are concerned with a general approach to the spatial distribution of attention that would apply to any requested distribution, that is, how well the observer could conform his/her actual distribution to the requested distribution. We use different spatial frequencies of requested attention as our basic tool. We first review the metaphors that previously have been used to characterize the spatial distribution of attention. Then, we consider the prior studies organized, loosely, in terms of the most important aspect of the paradigm.

1.1. Attention metaphors

Spatial attention has been likened to a spotlight (Posner, 1980), a zoom-lens (Eriksen & St. James, 1986), and a peak in an activity distribution (LaBerge, 1995); suppression of surrounding signals has been compared to a gate (Reeves & Sperling, 1986) and to troughs surrounding a peak in an activity distribution (LaBerge, 1995).

Spotlight. In the spotlight metaphor, the beam of attention is fixed in size and shape, and can be directed at a single area of the visual field (Posner, 1980). As a result, processing within that area is facilitated. The spotlight can be described as moving in an analog or discrete manner. The evidence supporting either description is mixed. For instance, Tsal (1983) argued that the distance travelled and time to travel in attentional shifts are correlated, supporting analog movement; others have provided evidence that the spotlight moves in discrete steps in which the time taken to move

* Corresponding author. Address: Center for Neural Science, New York University, New York, NY 10003, USA.
E-mail addresses: joetta@cns.nyu.edu (J.L. Gobell), sperling@uci.edu (G. Sperling).

0042-6989/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2004.01.012

the focus of attention is independent of the distance to be moved (e.g., Sperling & Weichselgartner, 1995).

Zoom-lens. Closely related to the moving-spotlight is the zoom-lens metaphor (Eriksen & St. James, 1986). The zoom-lens is also unitary. However, the area on which the ''beam'' of attention shines can be contracted or expanded as required by a task or instructions. In the zoom-lens model, the strength of the attentional beam decreases with an increase in size of the beam, but the beam cannot be split among more than one location.

Gradient. In contrast to these unitary models of attention, LaBerge and Brown (1989) proposed a gradient model of attention in which attention can be simultaneously allocated to non-contiguous locations in the visual field. In this model, the selected areas are controlled by a filter that operates on the location information in each display. Once the filter selects the location information for each selected area, the location information then allows the selection of the corresponding feature information. The selected areas and changes within these areas are described by a gradient of processing resources.

Of the more well-known metaphors, the moving-spotlight and the zoom-lens both depict attention as unitary in nature, while the gradient model of attention allows for the division of attention across more than one location separated in space.

Questions related to the understanding of spatial attention include: Can observers attend to two or more locations simultaneously? If they can, what is the nature or degree of the costs of this division? To what degree are the in between areas suppressed? These questions are certainly not exhaustive. A general theory would generate answers to these questions, and to many more. The purpose of the experiments presented here is to provide data to guide the development of a general model of the distribution of spatial attention.

1.2. Previous research

Researchers investigating the ability to attend to multiple locations simultaneously have employed a wide variety of paradigms; within these paradigms, both changes in attention instructions and in the physical layout of the stimuli have been exploited. The studies cited here can be roughly categorized as monitoring, cuing, interference, and stimulus configuration studies. In monitoring studies, observers are instructed to monitor one or more locations and perform a given task while either response time (RT) or accuracy data are collected. In cuing studies, the observer is provided a cue indicating that a target will appear in a location(s) with a certain probability. Occasionally, the target will appear in an uncued location. RT data are collected. In interference studies, observers are required to make a judgment (e.g., same–different) using information presented at two locations separated in space. In between these locations, additional information is presented and its effect on performance is studied. In stimulus configuration studies, the configuration of the stimulus to which the observer must attend is changed and the effects of that change on performance are studied.

1.2.1. Monitoring

Shaw and Shaw (1977) conducted a study in which observers identified a letter appearing at one of eight positions on an annulus whose diameter was 1 degree of visual angle. Accuracy data were well-predicted by the probability distribution of the target occurrence, consistent with simultaneously monitoring more than one location. However, the results can equally well be explained by probability matching.

Evidence of the ability to attend to certain information while suppressing other information which would fall within the same beam of a traditional zoom-lens is provided in a comparison between two conditions in the work of Melchner and Sperling (1978). Observers identified and located a letter among digits in the outer ring of the stimulus and a digit among letters in the inner ring. Under one instruction, observers were to give 90% of their attention to the outer ring, and under another they were to attend equally to both rings. Observers' performance for the inside ring gets much worse when primarily attending to the outside ring, while performance on the attended ring substantially improves. A contingency analysis showed that when observers detected one kind of target (letter or digit) they were very unlikely to detect the other kind, indicating that on a single trial observers performed only one of these two detection tasks. Observers' good performance when attending the outer ring indicates they can suppress almost completely the information lying in the center, i.e., within a single beam of attention. A long sequence of frames occurred at the rate of 10 Hz, so the suppression of the ''false targets'' on the inside had to be performed at a relatively early stage of processing prior to a limited capacity memory. The degree to which suppression of the inner items was accomplished was not addressed explicitly in this study, and it is unclear how to extend these results to other spatial arrangements without a general theory of spatial attention.

Castiello and Umilta (1992) used a task in which observers attended to two locations indicated by boxes, one in each hemifield. Observers responded when a light was briefly presented in either of the boxes. Three box sizes were used to indicate location—independently chosen for the two locations. An inverse relationship was found between box size and RT for both locations, interpreted as indirect evidence that observers can attend to two different locations at once.

In response to Castiello and Umilta (1992), McCormick, Klein, and Johnston (1998) used the same task with the addition of a probe procedure to provide a more direct measure of the distribution of attention. The lack of any difference in RT data for probes presented in between the two locations and for probes presented within either of the two locations supported the presence of unified attention across both locations.

Using electrophysiological measures, Heinze et al. (1994) used P1 amplitude modulation as evidence that attention cannot be divided but rather operates as a zoom-lens. In their study, observers were required to monitor 2 of 4 locations. One symbol was then presented in each of the four locations, and observers responded when these symbols matched. After this primary task, a probe appeared at 1 of the 4 locations. When the observer was attending to two separated locations, the intervening location showed P1 modulation; while the observer was attending to two adjacent locations, probes at the other locations showed no such P1 modulation. This study demonstrates that, when the information in the intervening location is not too disruptive to task performance, observers may use a unitary region of attention.

Using illusory line motion (ILM) as an index of attention, Schmidt, Fisher, and Pylyshyn (1998) presented observers with four dots to monitor plus a central fixation point. The dots were removed and a straight line appeared connecting the fixation point with either a dot location or a location in between two dots. Data showed a significantly higher occurrence of ILM when the line pointed to an attended location than when it pointed between two locations, consistent with attention at multiple locations and reduced attention to the locations in between.

Bichot, Cave, and Pashler (1999) used several variations on a monitoring task to provide evidence of separate attention windows that open and close independently. In all of the tasks, they compared performance when target items were presented simultaneously (in the same frame) with successive presentation (alternating frames). No difference in accuracy was found between the simultaneous and successive conditions, indicating observers were not switching back and forth between the locations but selecting both regions. Extending this investigation to require attention to more than two locations would provide more insight into the existence and degree of any costs.

An electrophysiological study by Eimer (2000), in which observers were instructed to monitor concentric rings in a display, found effects on ERPs consistent with the splitting of attention. Instructions required observers to either attend to an area that a single zoom-lens would cover without incorporating other areas or two rings that a zoom-lens could not selectively cover. Differences in the ERPs suggested some flexibility in distribution that is inconsistent with a zoom-lens of attention.

1.2.2. Cuing

In a classic cuing study, Posner, Snyder, and Davidson (1980) asked observers to press a key at the onset of a light in one of four locations. Prior to each block, observers were told which two of the four locations to prepare to attend. At the beginning of each trial a number at fixation indicated which location was 65% likely to occur. The other three location probabilities were either all equally likely or 25%, 5%, and 5% likely. RTs were faster at the 25% likely location than at 5% likely locations only when the 25% location was adjacent to the 65% likely location. This was interpreted as evidence that observers cannot monitor two locations unless they fall within the same unitary beam of attention. However, this task had the problematic aspect of providing little incentive to monitor the second location since the target stimulus usually occurred in the cued location. Another point of concern is that information that occurs at unattended locations does not interfere with attended information. Rather, it requires a response as well. There is little incentive to ignore uncued locations unless using information from the uncued locations would be detrimental to performance.

Using a similar paradigm, Eriksen and Yeh (1985) required observers to determine which of two target letters was present in an 8 letter annulus display. A precue indicated the primary location while the secondary location was opposite the primary. The validity of the cue varied from trial to trial (primary/secondary cue probabilities were 70%/10%, 40%/40%, or 100%/0%). Benefits of cuing were found only for the primary location, and not the secondary. This paradigm, however, does not address how observers can distribute attention if two target locations are equally (or nearly equally) likely across trials. Instead, in this study the primary location is far more likely to be the location of the target across trials (70% versus 16%). With such probabilities, observers might do well to mostly attend the primary location.

1.2.3. Interference

Using a partial report procedure, Awh and Pashler (2000) tested the ability of observers to split attention over two non-contiguous locations. Observers had to identify two target digits that appeared in an array of distractor letters. Spatial cues indicated the two most likely target locations. A strong accuracy advantage at cued locations compared with the intervening ones suggests that the primary mechanism supporting the flexible deployment of spatial attention is the suppression of interference from stimuli at unattended locations, though this suppression is not complete.

In a priming study, Pan and Eriksen (1993) used the interference caused by information falling in between two attended locations to argue that the focus of attention is unitary. In this study, two target letters appeared left and right of fixation at three possible eccentricities, and observers indicated whether they were the same. An incompatible or compatible distractor appeared in between these two target letters, priming a response. RTs were impacted by the irrelevant noise letter appearing between comparators, indicating that observers did not ignore the information. The authors interpret this as a failure to distribute attention in a non-unitary fashion. However, observers did not know where the distractor letter would appear on each trial, so they were unable to predict which areas needed to be suppressed. Good performance in this paradigm requires suppressing unattended areas.

Using this same/different task, Kramer and Hahn (1995) found an interaction between response type (same or different) and distractor type (prime same or different) only when the distractors had an onset. While this provides some evidence that observers can split attention, it does not attempt to investigate systematic changes in performance. Additionally, the interaction of response time and distractor type in the onset condition may be indicative of only partial division of attention rather than none.

In an extension of their 1995 study, Hahn and Kramer (1998) added probes at target, in between, and outside locations. Both the RT data from these probes and the accuracy data indicate that observers were indeed able to selectively attend to the separated locations and ignore the middle in the non-onset condition.

1.2.4. Configuration

Podgorny and Shepard (1983) asked observers to distribute attention across a subset of squares on a 3×3 grid and then indicate whether a probe dot appeared on or off of the attended area (equally likely). RTs were shorter for the attended area than the unattended, and correlated with the "compactness" of the attended area and not with the number of attended squares. These results suggest that the distribution of attention depends upon the stimulus configuration in a systematic manner.

1.2.5. Summary

There is ample evidence that under a variety of conditions observers can, to some extent, distribute attention to multiple locations. The evidence suggests that this disjoint distribution does come at a cost relative to maintaining a unitary focus of attention. The results from several of the studies suggest the importance of suppression; when observers were required to suppress information and knew where that information would be located, they were better able to disjointly attend.

The problem with these studies, and with many more similar ones, is that they do not lead naturally to a formulation of exactly what an observer's spatial distribution of attention actually is and how well that distribution conforms to the requested distribution. Additionally, achieving spatial distributions of attention that conform closely to requested distributions requires not only explicitly defining regions where stimulus input is to be enhanced but regions where it is to be suppressed. The importance of suppression regions has not been fully appreciated nor has the use of suppression been fully exploited.

1.3. Outline

In the task described below, the information within the unattended areas is extremely detrimental to task performance, thus providing a maximum incentive to make attention conform to the requested spatial distribution. Splitting attention into different disjoint regions will be essential to successfully perform the task. Without this suppression requirement, even if observers were explicitly instructed to divide attention, they might find it is easier or sufficient to use a unitary window of attention. Without a suppression requirement, demonstrating that observers are not distributing attention disjointly fails to distinguish between the inability to split attention and merely a preference for unitary distribution.

In our paradigm, observers are required to report the location of a target disk appearing in one of several attended locations while ignoring identical disks (false targets) appearing in unattended locations. The presence of many false targets makes it critical for observers to suppress (or ignore) the unattended regions.

Successful performance of the task is evidence of the ability of observers to divide attention among disjoint locations, i.e., to utilize non-unitary attention distribution. Additionally, the systematic changes in performance with changes in stimulus configuration will provide important insights into the characteristics of the attentional distribution. From these data, we construct a model of an attentional modulation transfer function. This model, based on Fourier principles, allows for the prediction of the distribution of attention across novel patterns—i.e., it is a general model.

2. General methods

Observers perform a search task which requires the distribution of visual attention across multiple disjoint locations. On each trial, a red-green square-wave grating indicates the areas to attend and those to ignore. After the square-wave fades into the background, a 12×12 search array of white disks appears. Observers must report the location (row and column) of the large white disk in the attended area (e.g., red). Observers tend to ignore the fine structure required by the map unless false targets (identical to targets) are placed in unattended areas to force confinement of attention to the to-be-attended stripes. Our procedure requires observers to suppress the information in the unattended regions because the presence of false targets there interferes with their ability to detect and localize the real target. In preliminary experiments, where no false targets were present, response accuracy was greater than 98%, demonstrating, first, that detecting the target in the absence of false targets is trivial and, second, that observers are able to accurately report the location of a detected target.

The criterion for a response to be judged as correct required that the observer reported a location equal or adjacent to the target location. This relaxed criterion was used because in preliminary experiments with no false target and a strict criterion (report exact target location) observer accuracy fell to 77%, i.e., on 21% of trials the observers misreported the location of the target to an adjacent square of the 12×12 array.

In order to investigate attention across a range of configurations, the red-green square-wave grating is changed from trial to trial. Orientation (horizontal or vertical), spatial frequency (four values, where ATT_1 is shown with a phase shift of 0 and π; the number of locations covaries with this measure) and color phase (two possibilities) of the grating are varied. See Fig. 1(a) for an illustration of these gratings.

2.1. Observers

A total of nine students—graduate and undergraduate—from the University of California, Irvine, participated in the experiments. All had normal vision.

Participants not associated with the laboratory received compensation of $8 per session. Each person provided written consent and was treated in accordance with the ''Ethical Principles of Psychologists and Code of Conduct'' (American Psychological Association, 1992). Consent forms and procedures were approved by

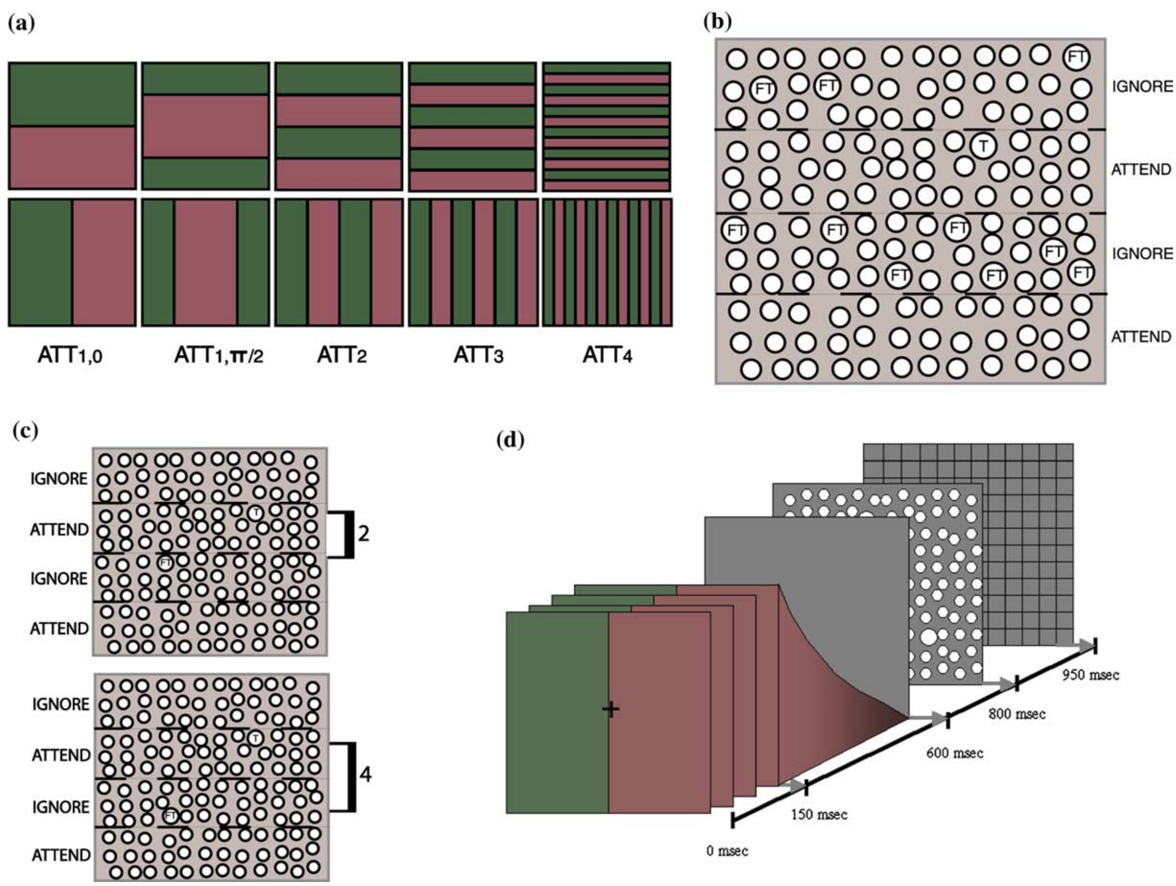

Fig. 1. (a) Attention conditions as defined by attention-instruction images. For each observer, one color is defined to be the attended color. A single target occurs on that color, either one or ten false-targets occur on the other color. In addition to the 10 horizontal and vertical attention conditions, there are 10 conditions identical to those shown except that the red and green colors are interchanged. (b) A sample ten-false-target search array in which the target (T) appears in an attended region, in row 4, column 9. The ten false targets (FT) are located in the unattended regions. Dashed lines and all other labels are only for illustrative purposes and were not present in the actual search array. (c) Two examples of one-false-target control trials of the same attention condition showing two different row-separation values (two and four). Target (T) and false target (FT) labels and other labels and dashed lines were not present in the actual displays. (d) Schematic representation of a single trial of one attention condition. All examples here appeared in all possible reflections and rotations.

the Institutional Review Board of the University of California, Irvine.

2.2. Stimuli

The basic stimulus consisted of three separate images: the indication of which regions to attend (i.e., the attention instructions), the search array, and the response screen. Each image subtended 12.5 degrees of visual angle horizontally and vertically.

Attention-instruction image (attention instruction). The attention-instruction image consisted of a red-green square-wave grating of one of four spatial frequencies: 1, 2, 3, and 6 cycles per image (cpi). The 1 cpi attention-instruction image was presented in two different phases—in one phase half of the display was red (green) and the other half green (red), while in the other phase the outer two quarters of the display were red (green) while the center half of the display was green (red). In all cases, the stripes were oriented either horizontally or vertically, and could either begin with red or begin with green (see Fig. 1(a)). With two stripe orientations and five gratings there are 10 possible attention instructions and $12 \times 12 = 144$ possible target locations in the search array for a total of 1440 experimental conditions.

The attention-instruction image was presented for 150 ms, at which point it began fading continuously into the gray background over the next 450 ms. The motivation for incorporating this fading was to eliminate the negative afterimage that otherwise would have followed the presentation of a brief attention-instruction image.

Search array. The search array was a 12×12 array of white disks on a gray background. All disks were circular; 11 of the disks subtended 0.34 degrees of visual angle and the remaining 133 disks subtended 0.23 degrees of visual angle. Each disk's location was jittered both horizontally and vertically, so the distance between the disks varied from trial to trial. The amount of jitter along each dimension for each dot was chosen from a uniform distribution with minimum $-0.3°$ and maximum $0.3°$. The average distance from the center of one disk to the center of a neighboring disk was approximately $1°$.

Of the 11 larger disks, 10 were randomly placed in regions of one color and one was placed in a region of the other color. The 10 disks on one color will be referred to as false targets (a.k.a. foils), while the single disk is the target of the search task. Smaller disks, distractors, were then placed in the remaining locations. See Fig. 1(b) for an example of a disk array.

Response screen. The response screen consisted of a 12×12 grid outlined with white lines on a gray background to guide the observer's response. This response grid overlapped the location of the search array perfectly, demarcating each of the possible disk locations. Each square in the grid subtended approximately $1°$ vertically and horizontally. Digits indicating the row and column numbers were present down the left side and along the top of this grid.

2.3. Apparatus

Stimuli were generated in MATLAB, using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Stimuli were displayed on an Apple Multiple Scan 720 Display monitor, powered by a Power Macintosh 7500/100 computer. The display resolution was set at 640×480 pixels, 60 Hz, which at 120 cm viewing distance subtended 18×13.5 degrees of visual angle. The gray background had a luminance of 20 cd/m^2, the white discs were 73 cd/m^2, the red stripes 13 cd/m^2, and the green stripes 12 cd/m^2.

2.4. Procedure

Instructions and task. Each observer was randomly assigned to attend to areas marked by either red or green. This color designation remained the same for that individual for the duration of his or her participation. The observer's task was to report the location of the larger white disk that occurred on the attended color.

Training. Prior to the experimental sessions, all observers completed one training session. Training consisted of viewing several sample trials at a very low speed with the experimenter explaining each step. The observer then completed one session in which the duration of the search arrays began at 1500 ms and incrementally shortened to 150 ms as the observer's performance improved. The practice session was terminated when the observer satisfactorily performed with a 150 ms display duration. Satisfactory performance was defined as correctly locating the target disk in 5 out of 10 successive trials. A training session typically lasted 45 min, similar in duration to the experimental sessions.

Trial sequence. On any given trial, unless specified otherwise, the sequence of events was as follows: the observer sat in a windowless unlit room 120 cm from the display and fixated a cross shown in the center of the display. The attention-instruction image appeared for 150 ms. During the following 450 ms the red-green grating continuously faded into a completely gray background which then remained unchanged for 200 ms. Then a 12×12 array of disks appeared for 150 ms. 150 ms was chosen to eliminate the possibility of an observer making a useful eye movements during the time the display was visible. See Fig. 1(d) for a schematic representation of a single trial of one attention condition.

The observer's task was to search the attended locations for the larger white disk. After the 150 ms presentation of the disk array, the response grid appeared and remained on until the observer responded. Responses were made using a keyboard where target location was indicated. Row and column responses were entered using the following key presses: 's', 'd', 'f', 'j', 'k', and 'l' indicated rows/columns 1 through 6, respectively; 'w', 'e', 'r', 'u', 'i', and 'o' indicated rows/columns 7 through 12, respectively.

3. Experiment 1: Attention modulation transfer function, excluding false-target crowding

3.1. Main goals

In Experiment 1, observers performed the task described in Section 2. A priori, we expect, and propose to measure, a decline in performance with an increase in the spatial frequency of the attention cue. This is an inevitable consequence of a decreased amplitude of attentional modulation as the spatial frequency of the requested distribution of attention increases. Additionally, we expect and propose to measure the performance decline with an increase in target eccentricity, due to reduced acuity in the periphery as compared to central vision. A performance advantage along the horizontal dimension of the display is also anticipated, consistent with many studies showing that performance in many tasks declines more slowly with eccentricity along the horizontal as compared with the vertical dimension of the visual field (e.g., Awh & Pashler, 2000; Carrasco, Talgar, & Cameron, 2001).

3.2. False-target crowding

The more false targets there are, and the closer they are to the real target, the worse we expect detection performance to be. Our goal is to study spatial attention per se, and we use false targets only to force spatial attention to mold itself to the requested attention distribution. Therefore, it is necessary to determine to what extent the presence of the false targets themselves (false-target crowding) perturbs our estimate of the distribution of attention.

3.3. Identical stimuli, different attention instructions

To estimate the influence of false-target crowding, occasional one-false-target trials (stimuli with only one false target instead of 10) are—unbeknownst to the observers—embedded in a series of stimuli with ten false targets. In a ten-false-target trial, at high spatial frequencies false targets are closer on average to targets than at lower spatial frequencies. By presenting identical one-false-target stimulus matrices with different attention instructions, the effect of the spatial frequency of the attention instructions can be determined absolutely. Additionally, we can measure the effect of false-target crowding. Indeed, by systematically varying the row-separation of the false target from the target in one-false-target trials, the effect on performance of the distance between a false and a real target can be specifically investigated. It is worth iterating that the one-false-target trials allow the presentation of the identical stimulus while varying only the attention conditions—the classical approach to the measurement of attention (Sperling & Dosher, 1986).

For the one-false-target trials, it is hypothesized that within a particular target/false-target configuration, performance will decline with an increase in spatial frequency of the attention instructions. This would indicate that the decline in performance with an increase in spatial frequency cannot be explained by the effect of false-target crowding. It also is expected that ten-false-target trials will result in a lower accuracy than similar one-false-target trials due to the increased interference of nine additional false targets resulting from incomplete suppression of the unattended regions. Additionally, the difference between the ten-false-target and one-false-target trials should increase with increasing spatial frequency of the requested attention distribution. We expect this interaction because, on the average, high spatial frequencies of the attend/not-attend grid permit false targets to approach closer to real targets than do low spatial frequencies.

3.4. Method

3.4.1. Subjects

Three observers ran the experiment with embedded false-targets and three observers ran the experiment without embedded false-targets. All subjects had normal vision and were graduate and undergraduate students in the Cognitive Sciences/Psychology Departments at UCI. All but CT and JG were naive to the purposes of the study.

3.4.2. Stimuli

One part of Experiment 1 consists of ten-false-target trials as described in Section 2. In another phase of Experiment 1, trials containing only one-false-target were added. The one-false-target trials varied along two dimensions: first, the attention condition, and second, the stimulus configuration—the number of rows or columns in between the target and the false target (called "row separation"). In Fig. 1(c), two examples of one-false-target stimuli are depicted—a target to false target separation of four rows, and a separation of two rows.

Six row separation values were used, 1–6. The average Euclidean distance separation corresponding to these in degrees of visual angle are 5.4, 5.8, 6.4, 7.1, 7.8, and 8.6. Because the expected spacing between disks is approximately 1.0°, these distances also represent the target to false target separations in terms of the disk-to-disk distances. Because of geometric constraints, it was not possible to combine all attention conditions with all separation values.

A sample of each of the 1440 possible conditions of the ten-false-target trials and each of the 296 one-false-target trial configurations were presented once in a randomized order. 17% of the trials were one-false-target trials.

3.4.3. Procedure

The procedure followed that described in Section 2. Observers were not told about the presence of one-false-target trials.

3.5. Results

A total of three observers completed Experiment 1 with one-false-target trials. Fig. 2(a) plots accuracy against attention condition for the ten-false-target trials, in both the horizontal and vertical conditions. The graphing conventions established here hold for the rest of the experiments when applicable.

Chance is defined as the probability of being scored as correct when the target location is chosen at random from among the attended locations. Because locations adjacent to the target are counted as correct when they fall in the attended area, chance performance differs slightly for each condition. These values are indicated by individual dots on the bottom of the figure.

The results show that as the spatial frequency of the attention-instruction image increases, accuracy decreases.

Fig. 2(b) plots accuracy at each target location for the ten-false-target trials, collapsed across attention condition. For example, the upper left corner indicates performance on all trials in which the target appeared in row 1, column 1 of the display. Performance is best in the center of the display (around fixation) and falls off as the distance from fixation increases. Accuracy generally falls off slightly more quickly in the vertical direction than in the horizontal.

Results for three other observers who ran the experiment without the one-false-target trials follow a very similar pattern. See Fig. 10 for these data. However, with respect to the 144-location plot in Fig. 2(b), one of these three extra observers showed a performance pattern consistent with a shifted point of fixation. Her performance by target location was similar to all other observers, but the peak was shifted upward and to the left by about two rows and columns.

Fig. 3(a) shows data for the one-false-target trials with data for the ten-false-target trials (from Fig. 2)

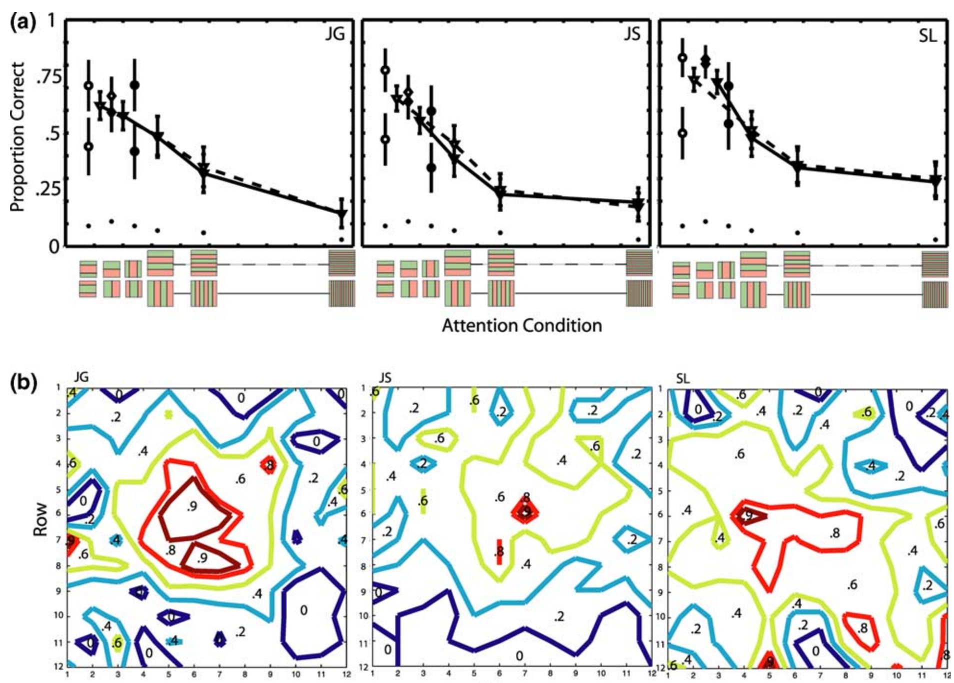

Fig. 2. (a) Accuracy by attention condition for three observers. Dashed line connects horizontal attention conditions, solid line connects vertical attention conditions. Open circles to the far left represent the horizontally oriented $ATT_{1,\pi/2}$ condition for both color phases, while filled circles represent the vertically oriented $ATT_{1,\pi/2}$ condition. Diamonds represent the horizontal and vertical conditions for the $ATT_{1,0}$ condition. Two leftmost triangles represent the averages, for each color phase, of the three data points for ATT_1. Error bars represent the 95% confidence intervals. Dots indicate chance performance. (b) Accuracy by target location for three observers, averaged over all attention conditions. Digits inside the regions defined by contours indicate the lower bound of proportion correct within that region. Contours are drawn at the boundaries of 0.9, 0.8, 0.6, 0.4, and 0.2 proportion correct.

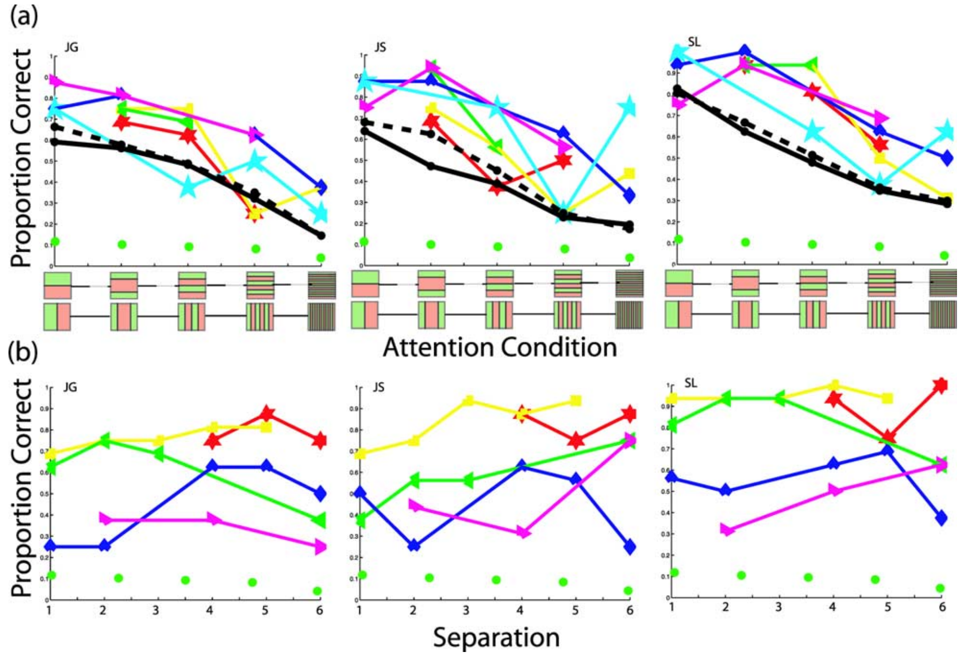

Fig. 3. (a) Performance across attention conditions as a function of target to false target separation. For reference, solid and dashed lines with closed circles indicate performance on ten-false-target trials (taken from Fig. 2a). All other lines indicate performance on one-false-target trials with a separate line for each separation value: 6-point stars = 1, squares = 2, leftward triangles = 3, diamonds = 4, rightward triangles = 5, 5-point stars = 6. It was not possible to measure all separation values in all attention conditions. Dots indicate chance accuracy. (b) Accuracy by target to false target separation value for each attention condition: stars = $ATT_{1,0}$, squares = $ATT_{1,\pi/2}$, leftward triangles = ATT_2, diamonds = ATT_3, and rightward triangles = ATT_4. Dots indicate chance accuracy.

included for reference. For the one-false-target trials, each data point represents performance at a particular target to false target separation value for a particular attention condition averaged over both orientations. Each line plots the performance averaged over the attention conditions for one of the six row-separation values. For all six separation values, there is a clear trend of declining performance with an increase in spatial frequency. For all but a few conditions, performance on the ten-false-target trials is worse than on the one-false-target trials.

Fig. 3(b) depicts data for the one-false-target trials only. In this figure, each line plots the performance within one attention condition across the six separation values. There is no consistent pattern across the separation values. Within some attention conditions (e.g., JS ATT_2 or SL ATT_4) proportion correct increases with an increase in separation value. Within other attention conditions (e.g., JG ATT_2) accuracy tends to decrease with an increase in separation value. In still others (e.g., JG ATT_4 or JS $ATT_{1,0}$) there is no clear pattern to the variations in accuracy with an increase in separation value. This may be due to the fact that average Euclidean distance from target to false-target varies only by a factor of 1.6 as the row separation varies from 1 to 6. However, there is a clear and consistent decrease in proportion correct with an increase in spatial frequency of the attention condition, and this occurs even in the one-false-target trials in which identically configured displays are viewed in all attention conditions.

3.6. Discussion

Spatial low-pass filter. Performance declines with an increase in the spatial frequency of the attention cue. This pattern of performance is consistent with a "low-pass filter" of attention; as the spatial frequency of the attention distribution required increases, the modulation of attention decreases. This means that at the lower spatial frequencies, the attention modulation between attended and unattended locations is relatively high. Due to this high modulation, information in attended regions is strongly enhanced while the information in unattended regions is strongly suppressed or ignored, and detection performance is very good. As the spatial frequency of the requested attention distribution increases, the enhancement of attended areas and the suppression of unattended areas are both reduced, resulting in little difference between the attended and unattended regions, and therefore a lower detection accuracy.

Number of locations versus Fourier theory. This experiment and those that follow do not specifically deal with number of locations, a traditional variable of interest in the study of spatial attention. The problem with number of locations as an explanatory variable is that it does not generalize well. When two separated locations, A and B, are connected by a thin channel C, do they then become one location or three locations, A, B, and C? Are all one-location or all three-location spatial configurations equally difficult to attend? How can one characterize these differences?

The problem with locations as an explanatory variable is that there is no formal theory of locations that would enable one to predict the attention response to the vast universe of different configurations of attention that might be requested. The advantage of Fourier theory is that once the attentional response to every frequency of spatial sine wave (of requested spatial attention) is known, the attentional response to any arbitrary requested configuration can be computed. Of course, there is no a priori guarantee that this linear systems approach to a comprehensive theory of spatial attention theory will be successful. But, the experiments described here form the foundation for such an approach and for subsequent tests of its adequacy.

Even if a theoretical approach to spatial attention based on number of locations were ultimately necessary, the Fourier theory would still be needed to describe the ability of attention to conform itself to the requested distribution at each location. We will show below that an attention theory based on a single attentional process that can mold itself (subject to spatial frequency constraints) to the requested distributions of attention (which involve from 1 to 6 attended locations) gives an excellent account of the data from the experiments described herein.

Eccentricity. The observed fall-off in accuracy with an increase in target eccentricity is consistent with the well-known fall-off in acuity with increasing eccentricity. Targets that fall further from fixation are more difficult to distinguish from distractors than are those falling near fixation. As expected, the fall-off is somewhat faster along the vertical dimension than along the horizontal. A general model of spatial attention must take this particular limitation of the visual system into account.

One-false-target trials. Within the one-false-target trials, observations on stimuli with identical target to false target separations can be compared under different attention conditions. Within trials with identically configured stimuli, the data show a consistent decline in accuracy with the increasing spatial frequency of the attention condition. This provides direct evidence that the patterns observed in the data from ten-false-target trials can be attributed at least in part to the increase in spatial frequency, independent of the target to false target separation.

The number of false targets present does affect performance, as can be seen in Fig. 3(a). The data from ten-false-target trials fall consistently below those from the one-false-target trials, except perhaps for subject JG in the case of a separation value of six. At the separation value of six, the average eccentricities of the target and false target are necessarily greater in the one-false-target than in the ten-false-target stimuli. This increase in average eccentricity with decrease in number of targets at high spatial frequencies may account for the apparent paradox (of improved performance with more false targets).

In Fig. 3(b) there is no consistent relationship between the separation value and performance. If, for instance, increased separation between the target and false target resulted in increased proportion correct, then each line in this Fig. 3(b) (representing a particular attention condition) would have a positive slope. If increased separation resulted in decreased proportion correct, each line would have a negative slope. As can clearly be seen, there are instances of positive, negative, and close-to-zero slopes. What is evident in Fig. 3(a) and (b) is that proportion correct decreases with an increase in the spatial frequency of the attention condition. This is evidenced in Fig. 3(b) by the order of the separate lines on the graph. The line depicting performance for $ATT_{1,0}$ has generally the highest proportion correct, while that for ATT_4 has the lowest.

Consistent with the many previous findings described above, there are clear costs of distributing attention in such a way that suppression of intervening regions is required. The higher the spatial frequency of the requested attention distribution, the poorer the performance.

4. Experiment 2: Foreperiod

The previous experiment varied the precise form of the requested distribution of spatial attention. However, attention experiments typically compare results from an active state of attention with those from a neutral state of attention in order to illustrate the presence of attentional factors. In Experiment 1, there was only one type of attentional state, namely an active state (attend to a particular spatial configuration). The neutral state of attention was not utilized. Comparison to a neutral state is needed to estimate a component of attention that might be common to all the requested attentional distributions of Experiment 1.

Another issue that cannot be directly addressed by data of Experiment 1 is that of the level of processing at which attentional selection occurs. For example, although it seems obvious to observers that attentional selection occurs at an early level—they are unaware of most distractors—it nevertheless might be argued that attentional selection occurs at a high level of processing. A priori, it might be possible for observers to record all the relevant information from the display in an iconic and/or short term memory and then to find the target—

the location in which the correct color and correct size coexist—by a memory search.

In Experiment 2 foreperiod (the time from the onset of the attention-instruction image to the search array onset) is varied. This will enable comparison of (1) neutral versus specific attention states and (2) an evaluation of early versus late attentional selection. The foreperiod variation includes a post-cue condition in which the attention-instruction image appears after the search array. In the case of early attentional selection, the performance with a post-cue will be vastly inferior to performance with a pre-cue. In the case of late selection, a post-cue immediately after the search array should still enable good performance.

Insofar as there is early attentional selection, we expect that, as the duration of the foreperiod increases, performance will improve up to some limiting time. Improvement in performance with the same stimulus configuration but with a longer foreperiod results from the increased time to prepare to process information at the to-be-attended locations (post-cue time variations enable estimates of forgetting). In either case, the later the attentional-instruction image relative to the search array, the less time there is to prepare, and the lower is the expected accuracy of target detection. A systematic increase in accuracy with increased foreperiod would provide the traditional evidence of two distinct attentional states with exactly the same stimulus, as well as insight into the time course of the switch from a neutral to a spatially configured attention state. The time course of spatial-attention preparation, which is inferred from the foreperiod manipulation, can then be compared with the time course of attention observed in other attention paradigms.

4.1. Methods

4.1.1. Observers

Three observers participated, two of whom are common to Experiment 1.

4.1.2. Stimuli and procedure

The stimulus configuration and procedure were as described in Section 2, with the exception of the foreperiod between onset of the attention-instruction image and the search array.

On any given pre-cue trial, the sequence of events was as follows: The attention instruction appeared for a randomly selected foreperiod of either 17, 167, 333, or 1000 ms. The attention-instruction image remained on during most of the foreperiod to mitigate possible forgetting of the instruction. The pre-cue was followed by a uniformly gray display for 17 ms. The fading interval of Experiment 1 was not included because it would have prevented the use of short foreperiods. After the 17 ms blank interval, a 12×12 array of disks appeared for 150 ms followed immediately by the response grid which remained on until the observer typed on a keyboard the row and column location of the target and his or her confidence in the response.

The post-cue trials differed only in order of presentation. The search array appeared first, for 150 ms, followed by a gray interval for 17 ms, followed by the attention instructions which remained on for 333 ms. 333 ms was chosen as the post-cue foreperiod based on pilot studies which indicated that at 333 ms performance had essentially reached asymptote.

4.2. Results

Three observers completed this task. The first three panels in Fig. 4 plot accuracy versus attention condition for each of the attention instruction foreperiods, collapsed across the vertical and horizontal conditions for each observer. There are 1250 observations for each observer, for a total of 50 observations per data point. The dashed lines indicate the two most extreme conditions: the post-cue condition and the condition in which the pre-cue was presented for 1000 ms. As the spatial frequency of the attention condition increases, accuracy decreases. Additionally, the proportion of correct

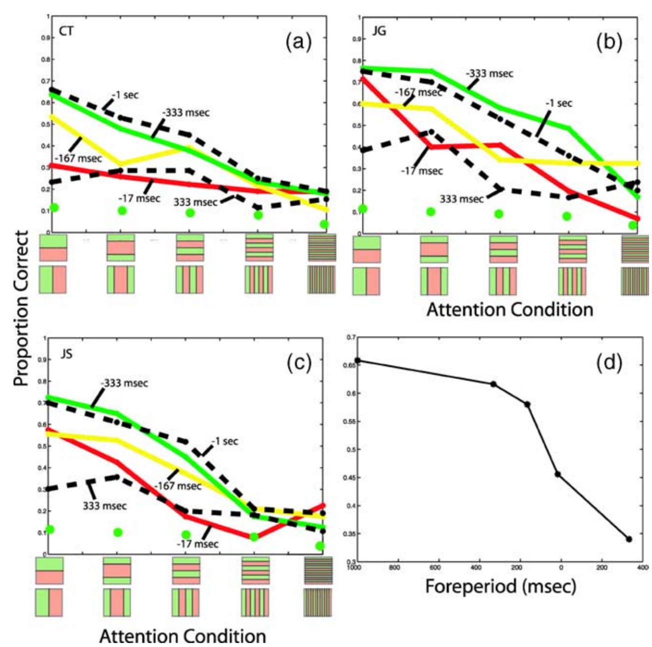

Fig. 4. Accuracy as a function of foreperiod. Panels (a)–(c) show different observers. The ordinate is the proportion of correct target detections. The abscissa represents the attention condition, averaged over vertical and horizontal and over all phases. The curve parameter is the cue foreperiod. Dashed lines indicate the extreme foreperiods: earliest precue (1 s) and the post-cue. Dots indicate chance performance. Error bars indicate 95% confidence intervals. Panel (d) depicts performance, collapsed across all observers, for ATT_1 and ATT_2. The horizontal line indicates chance.

detections increases as the duration of the foreperiod increases, up to 350 ms.

As expected, performance varies enormously with foreperiod. Overall, performance is worst for the post-cue condition and almost as bad for the 17 ms cue (30 ms foreperiod). Performance systematically improves with increased foreperiod, until the attention instructions are presented for 333 ms (350 ms foreperiod). A further increase of the foreperiod to 1 s provides insignificantly increased accuracy. Nearly all of the improvement in performance occurs between foreperiods 30 and 350 ms.

The last panel of Fig. 4 shows this change in performance with change in foreperiod averaged across all observers, for ATT_1 and ATT_2. The graph shows a fall-off in performance as the foreperiod moves from 1 s before the search array to 333 ms after the search array.

4.3. Discussion

The results provide clear evidence of two distinct attentional states—a state in which attention is spatially configured to conform to the attention-instruction image, and a neutral attentional state in which the observer perceptually processes the stimulus without knowing in which locations the target may and may not occur. That is, in the post-cue condition, observers do not have information that would allow them to prepare spatial attention, and therefore they are in a neutral state when the search array is presented. With a post-cue, performance is enormously worse than with a pre-cue. Any correct post-cue responses cannot be based on early attentional selection; therefore the post-cue performance represents a baseline of performance against which foreperiods of pre-cues can be compared to determine the benefits of attention.

The large improvement in performance between foreperiods of 350 and 30 ms is similar to the preparation times for attention found in the literature (for reviews see Sperling & Weichselgartner, 1995; Shih & Sperling, 2002). With respect to the issue raised in the introduction of comparing two kinds of attentional states (neutral and spatially configured) with the same stimuli, the data provide indisputable evidence of multiple attention states (neutral, versus prepared for one of the 10 possible spatial configurations).

In addition, the data from the post-cue condition help to distinguish between an early, perceptual attention process and a late attentional process involving short-term memory. That is, one may be concerned that task performance depends not on attention to the to-be-attended stripes, but rather on a memory of the stimulus. While the question of "levels of processing" (early versus late) cannot be completely clarified here, the poor post-cue performance indicates that observers are not searching their memory of the search array after the attention instructions have been received. Specifically, in the post-cue condition, the search array appears before the attention instructions although with the same duration as in Experiment 1. This allows the same processing time for each display but does not allow for any conformation of attention to particular locations prior to the presentation of the search arrays. Not only was performance in this post-cue condition extremely poor, the attention cue must preceed the target stimulus by 167 ms to acheive the major benefit of attention. As in many other attention tasks that involve spatial attention, perceptual preparation well in advance of stimulus presentation is essential for good performance.

5. Experiment 3: Retinal, cortical, or object coordinates for attention?

In all previous experiments, the viewing distance was fixed at 120 cm, and the attention instructions were described in terms of their spatial frequency. However, it has not yet been demonstrated whether it is the absolute or the relative "object-based" spatial frequency which is important. This is an important issue because it concerns the level of processing at which spatial attention control signals are generated (as opposed to the level of processing at which they take effect).

A priori, we expect that the decision about where to attend is a high-level process. Even when, in our experiments, the to-be-attended locations are represented as colored areas in a spatial array that conforms exactly to the search stimulus, attentional control is image based. For example, Parish and Sperling (1991) found that letter detection in noise was object dependent. That is, letter identification accuracy was completely independent of retinal spatial frequency over a 13:1 range of frequencies (generated by varying the viewing distance 13:1). In this experiment, spatial frequency is varied by halving the viewing distance of the search task to 60 cm. Halving the viewing distance halves all the spatial frequencies in the experiment. If the spatial attention were frequency dependent, on a graph of performance versus log frequency (somewhat similar to Fig. 2(a)), halving the spatial frequencies would be expected to shift all the data to the left by log(2).

Retinal inhomogeneity and possible attentional inhomogeneity as a function of eccentricity represent complications to a theory of attention. However, the model to be presented below enables us to separate acuity and attentional components of performance. The data from Experiment 3, in which spatial frequency is varied, provide a foundation for approaching the issue of object versus retina-based scaling of attention.

5.1. Methods

5.1.1. Stimuli and procedure

The stimuli were identical to those described in Section 2, except for the change in retinal image size when the viewing distance was halved to 60 cm. The new stimulus measurements are provided below.

At 60 cm viewing distance, the display was 25×25 degrees of visual angle squared. The attention instruction image was of four spatial frequencies: 0.16, 0.24, 0.48 and 0.96 cpd. The target and false targets subtended 0.68 degrees of visual angle and the distractors subtended 0.45 degrees of visual angle. The distance from the center of one disk to the center of an adjacent disk was approximately 2°. The amount of jitter along each dimension for each dot was chosen from a uniform distribution with minimum −0.6° and maximum 0.6°. At the 120 cm viewing distance, stimulus measurements were as in Experiments 1 and 2.

For observer JG, the data for a viewing distance of 120 cm were data from Experiment 1 that were obtained without one-false-target trials. For AH and CST, observers completed 864 trials at both 120 cm and 60 cm viewing distance. CST trained and completed the first 864 trials at 120 cm, while AH trained and completed the first 864 trials at 60 cm. Observers then completed 864 trials at the other viewing distance. Other procedural aspects of the experiment were conducted as described in Section 2.

5.2. Results

Three observers completed this task. Fig. 5 plots accuracy against attention condition for the horizontal and vertical conditions. To prevent clutter in Fig. 5, only the performance averaged over all phases of each attention spatial frequency is plotted. The performance on each of the conditions (and subconditions not shown separately) was similar to that of the earlier experiments. Basically, as the spatial frequency of the attention condition increases, accuracy decreases.

The change in viewing distance did not change performance in any substantial or systematic way. Performance is nearly identical even with a halving of the viewing distance, indicating scaling across a wide range of frequencies. This is similar to the scaling in the letter identification task of Parish and Sperling (1991) cited above. Additionally, there is no indication that training at 120 cm versus 60 cm viewing distance had a significant effect on performance.

5.3. Discussion

As seen in Fig. 5, performance did not substantially change with a change in viewing distance. This pattern of results suggests that the distribution of spatial attention may be better understood in terms of "object based" spatial frequency. rather than retinal spatial frequency. However, alternative interpretations are possible, so a resolution of the object versus retinal frequency issue must await a formal statement and testing of the alternatives.

6. Experiment 4: Attentionally-cued conjunction search

In a conjunction search task, an observer must find a target defined jointly by two attributes, e.g., large and red, in the presence of distractors that are large and green, small and red, and small and green. There is a close conceptual similarity between the attention-defined search studied in Experiments 1–3 and conjunction search. In our attention experiments, observers searched for a target that was defined by large and attended, in the presence of false targets that were large and unattended, and distractors that were small and attended and small and unattended. In terms of neural circuitry, very similar anatomical structures could serve attention-defined search and conjunction search. Experiment 4 examines a novel attentionally-cued variant of conjunction search.

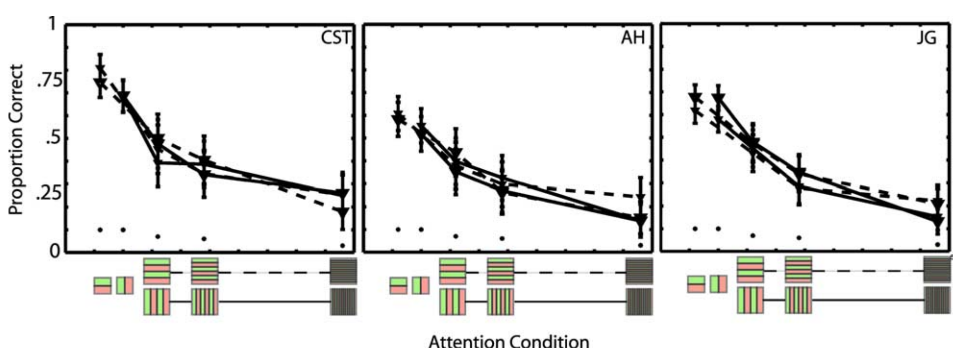

Fig. 5. Accuracy for each of the attention conditions at viewing distances of 60 and 120 cm. Dashed line is horizontal condition, solid is vertical. Thin line is 60 cm viewing distance, thick line is 120 cm. Dots indicate chance. Error bars indicate 95% confidence intervals.

To study the joint operation of attention-defined search and conjunction search, the search task was altered so that the attention-instruction image remained on during the exposure of the search array. When the red-green attention-instruction image remains on (instead of turning off as in Experiments 1–3) and is superimposed on the search array, the task becomes a conjunction search during the period in which the red-green and search arrays are visible simultaneously. The attributes of disk size and of the color surrounding the disk now define the target. The task becomes an attentionally-cued conjunction search when the attention-instruction image precedes the search image. We expect that performance in Experiment 4 will be better than in the previous experiments. The addition of the color to the search array itself can strengthen the attentional discrimination between to-be-attended and unattended areas that has been established by the prior attention-instruction image.

6.1. Methods

6.1.1. Stimuli and procedure

The stimuli and procedure are identical to those described in 2, with one exception. The attention-instruction image does not fade out but remains on continuously throughout the exposure of the search array and the response grid. The colors of the attention-instruction image were chosen to be approximately isoluminous with the gray background of Experiments 1–3. In this way, the luminance contrast between the disks and the background remained unchanged with the addition of color, thereby leaving the intrinsic difficulty of target discrimination in Experiment 4 equal to that in Experiments 1–3.

6.2. Results

Four observers completed this experiment. Results are qualitatively very similar to those of Experiment 1. Conjunctions search accuracy versus attention condition is displayed in Fig. 6(a). As the spatial frequency of the attention condition increases, accuracy decreases. As in Experiment 1, accuracy is slightly higher in the horizontal condition, especially for the lower spatial frequency conditions. Additionally, accuracy falls off more quickly in the vertical direction than in the horizontal. Not shown here, but as in Experiment 1, performance is best in the center of the display (at fixation) and falls off as the distance from fixation increases. Unlike Experiment 1, performance was better in the lower visual field than in the upper visual field.

Fig. 6(b) displays the relative improvement in search accuracy in attention-cued conjunction search compared to pure attention-cued search in Experiment 1. As the spatial frequency increases, there is a larger relative improvement in performance.

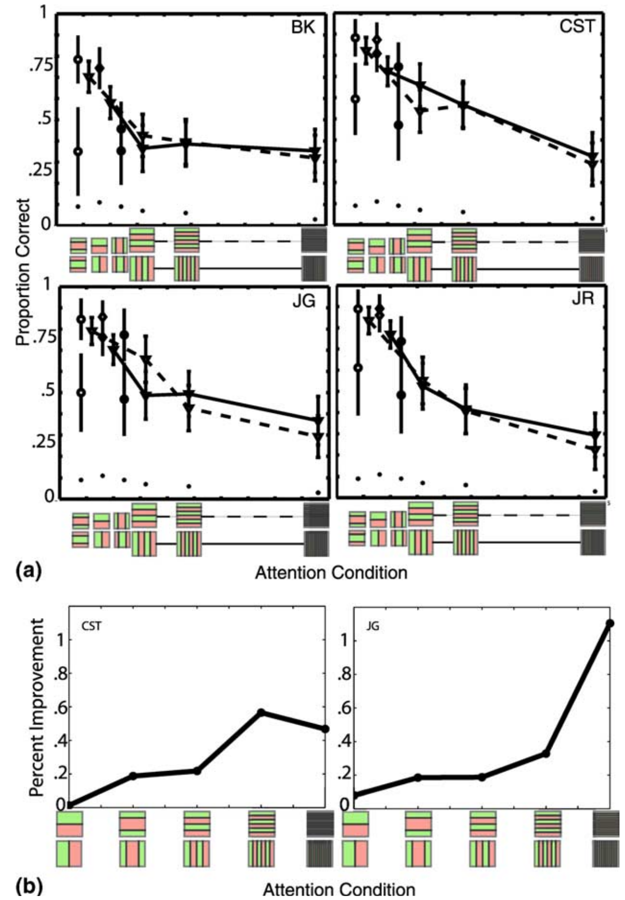

Fig. 6. (a) Accuracy by attention condition for four observers in attention-cued conjunction search. Dashed lines correspond to the horizontal condition and solid lines to the vertical. Dots indicate chance accuracy. Error bars indicate 95% confidence intervals. (b) Relative improvement in performance for two observers in the attention cued conjunction task as compared to the attention task without the conjunction search.

6.3. Discussion

When both the attention instructions and the search array are present, the observer can perform the task as a conjunction search with the two features of color and size. Comparing performance from the current experiment with that of Experiment 1, it is clear that performance improved when the color that defines an attended location was present during the display itself. This indicates that observers did indeed perform the conjunction task.

It is of interest that the performance in attention-cued conjunction search was qualitatively similar to performance in the attention tasks above. Following from this similarity, the model developed below will treat attention to a location as just another feature, like color, in a conjunction search.

The greatest benefit in attention-cued conjunction search relative to attention search was achieved in the most difficult condition—the highest spatial frequency of the attention/conjunction grating. In this condition, accuracy nearly doubles, whereas it improves by less than 10% in the easiest condition. This suggests that the feature map for color has a higher resolution than the attention control process.

7. Experiment 5: Internal noise via double-pass

Below, we propose a model of visual search. In a deterministic model, when the model performs the same trial twice, or better, the same sequence of trials twice, the predicted response will be exactly the same. We know that this never happens in human threshold performance. We propose to represent such indeterminacy in the model by internal noise. A double-pass method is used to estimate internal noise (Burgess & Colborne, 1988). The same observer performs precisely the same sequence of trials in two different sessions.

7.1. Methods

7.1.1. Stimuli and procedure

The stimuli and procedure are identical to those described in Section 2. Two observers participated in this experiment. Two observers (JG and CT) performed sequences of 1440 and 576 trials, respectively, that were identical to the sequences viewed in Experiment 1. Stimuli were exactly identical in every respect including for example, the jittering of each dot's position.

7.2. Results

As shown in Fig. 7, the general pattern of results from two identical sequences of trials is similar but obviously not identical. On a trial-by-trial basis, the observers' responses agreed in the sense of being scored correct or incorrect on 76% of trials (78% for JG, and 72% for CT).

7.3. Discussion

Overall, on 76% of trials the observer's response was the same for both instances of the trial (either correct

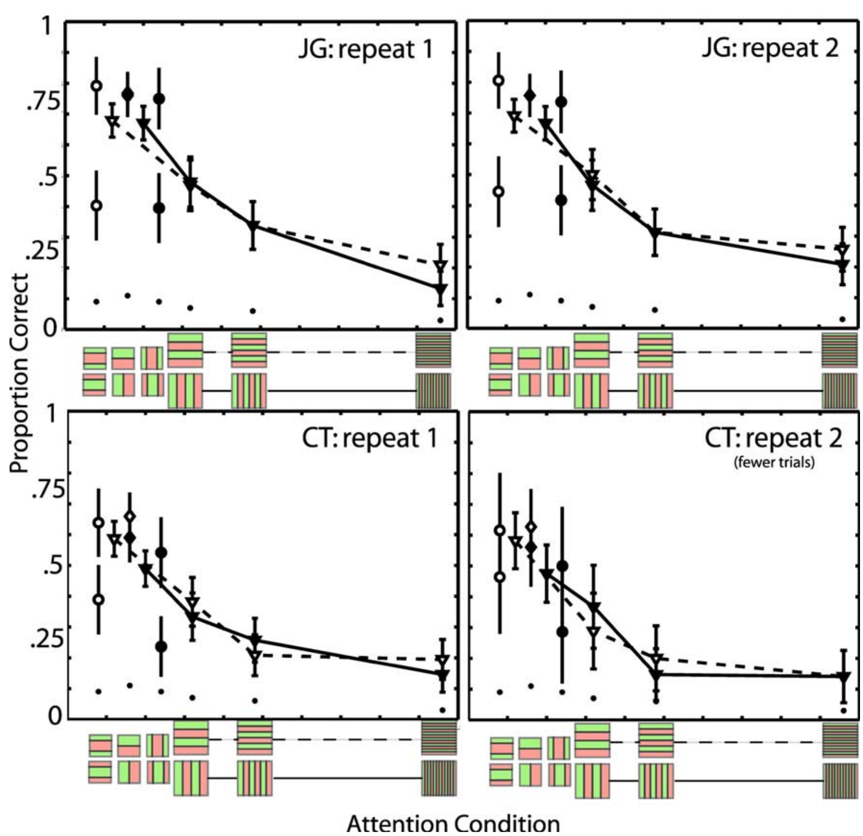

Fig. 7. Double-pass results. Two observers, JG and CT, completed two runs with exactly the same stimuli in the same sequence. The first run is on the left and the second run is on the right. JG completed 1440 trials, and CT completed 576 trials. Dots indicate chance performance. Error bars indicate 95% confidence intervals.

or incorrect). For any model that matches the average performance of the observer, this 76% serves as an upper bound on the predicted trial-to-trial correspondence between the model and data. This measure of internal noise does not distinguish between sensory noise, variable fatigue, occasional inattention, typing errors, and all other sources of trial–retrial discrepancy.

8. Model of the spatial distribution of visual attention: General framework

The framework of the model is the attentional modulation (by multiplication) of the input stimulus before it arrives at a decision stage (Reeves & Sperling, 1986). Such a process or framework, depicted in Fig. 8, is not unique to this work. It is a general framework that has been used to describe many different types of attentionally modulated visual information processing (for a review, see Shih & Sperling, 2002).

It is useful to conceptualize the search stimulus and the attention instruction as being processed by two streams although, obviously, much of the early processing is similar in both streams. The search stimulus is processed in a to-be-attended stream and the attention-instruction image is processed in an attention-control stream. After the search stimulus enters the system, it is represented in visual sensory memory (VSM). The VSM representation is acquired very quickly. Once the stimulus is turned off, the VSM representation decays with a longer time constant. The search stimulus is also perturbed by acuity constraints inherent to the visual system. As distance from fixation increases, acuity decreases; this decrease is somewhat slower along the horizontal than along the vertical dimension. The temporally and spatially modified search stimulus is then multiplied by the spatial attention pattern.

The attention-control stream interprets the attention-instruction image and generates the spatial attention modulation function (spatial attention pattern). Cue interpretation (determining what spatial attention pattern to generate) takes a period of time, τ, of about 50 ms. Following cue interpretation, an attention modulation function is generated quickly, reaching its maximum level in about 300 ms and sustaining this level until the search stimulus has been processed.

The output of the acuity function in the to-be-attended search stimulus path is multiplied by the output of the attention-control path. The information about each stimulus element (a disk in these experiments) is integrated over its space–time neighborhood and compared to the template of the target disk to arrive at the element strength. To this, internal noise is added, and the decision stage selects the location with the strongest element strength. This general outline must now be filled in with the model computations that are carried out at each stage.

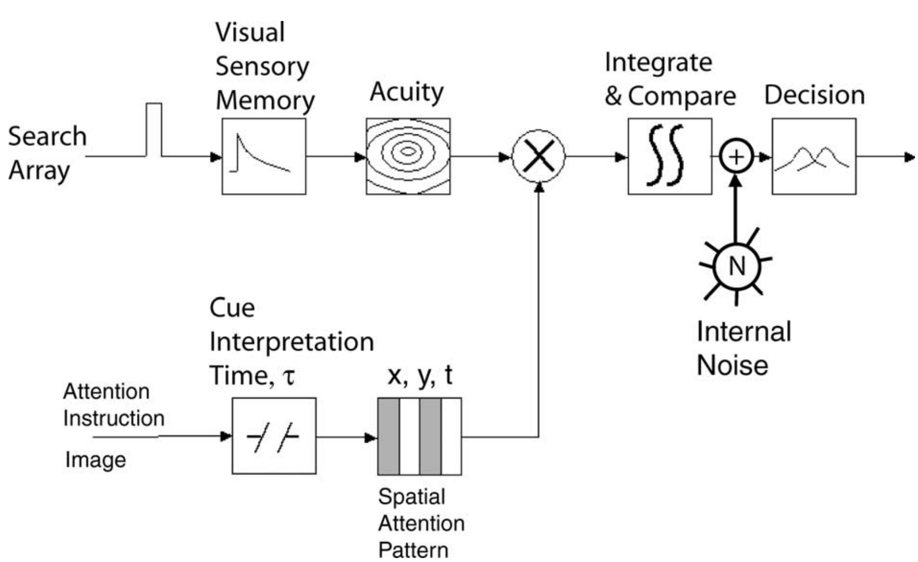

Fig. 8. General framework for attentionally-modulated information processing. There are two processing streams. The stimulus to be modulated (upper stream) enters visual sensory memory (VSM) and, when it turns off, begins to decay exponentially. The quality of information is constrained by acuity which decreases with eccentricity (as described by the concentric ellipses). The attention cue is processed in the lower pathway and, after a cue interpretation time τ, it generates a space–time attention function that multiplies the stimulus stream. Stimulus components are integrated over their spatial and temporal neighborhoods and compared with the target disk template to produce internal strength representation for each stimulus element. Element strength is determined by stimulus size, by acuity and by how much attention it has received. Internal noise is added to element strengths and the decision stage selects the location with the greatest element strength.

9. Model description

The more detailed model of Fig. 9 shows the basic computational structure. The input search array is acted on by three strength maps (acuity-, attention-, and conjunction feature-strength) to produce an element strength for each stimulus disk. Strength maps (to be described below) are encoded as 12×12 matrices of values. After internal noise is added, the decision stage selects the location (row and column) with the highest element strength as the first candidate response.

The first strength map is a representation of the stimulus input. Namely, locations which contain objects with features consistent with the target are assigned a higher strength (T) than those locations with features inconsistent with the target (D). Thus, the stimulus input map is a 12×12 matrix, where each cell has a value of either D or T. In the current experiments, this map could be said to be based on a bottom-up saliency derived entirely from integrated local luminance because the targets are larger than the distractors. One can easily conceive of situations in which target identification requires more complicated operations or more top-down processing.

The second strength map is a representation of attended and unattended regions. The decrease in proportion correct with an increase in spatial frequency of the attention instruction observed in all the experiments reported here is suggestive of a low-pass attention modulation transfer function. As the spatial frequency of the attention instruction increases, the amplitude of attention modulation decreases. In other words, in conditions where there are many narrow stripes there will be less attention allocated to attended areas and less suppression of information in unattended areas than in conditions where there are only a few wide stripes.

Linear systems analysis was employed to model such an attention modulation transfer function, using the sum of sinusoidal components to construct a distribution for each of the five attention instructions. The amplitude of each of the harmonics in the sum is determined by an attention modulation transfer function which takes the form of an exponential decay function, $A(f) = e^{-rf}$, where A is the amplitude of that harmonic, r is the rate of fall-off, and f is a spatial frequency. Fourier theory provides a well-defined foundation on which to build this model.

For each attention condition the number of harmonics included beyond the fundamental frequency is determined by the number of possible target locations within one row or column of each region (e.g., for ATT_4, only the fundamental frequency is included in the calculation because there is only one possible target location within a row or column of each region, while for $ATT_{1,0}$, the third, fifth, seventh, ninth, and eleventh harmonics are also included because there are

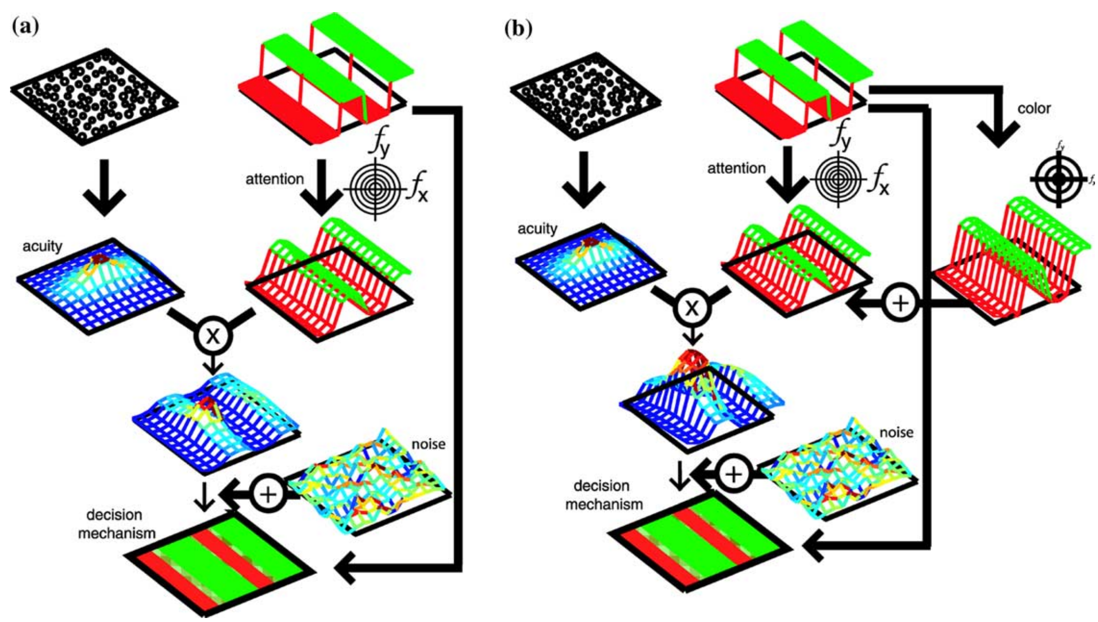

Fig. 9. (a) Structure of the model and its parameters. Each stage is a 12×12 strength map. The stimulus input map values are higher at target T and false target locations and lower at distractor D locations. Acuity is highest at fixation, and diminishes as eccentricity increases with a slower decrease along the horizontal (relevant parameters: α, β, and γ). The attention map strengths are higher at attended locations and lower at unattended locations (fall-off rate r of the spatial attention modulation function as a function of spatial frequency, indicated by concentric circles). The output map is simply the location-by-location product of the stimulus with the acuity function the attention modulation map, plus the addition of internal noise. The decision stage selects one location as the response. (b) Structure of the model with the addition of the color information. The structure is identical to that of (a) except that in this case the attention instructions remain on during the search. The additional information available via the color map adds to the attention map.

six possible target locations within a row or column of each region). Although all the relevant harmonics are included, the higher frequency harmonics will have negligibly small amplitudes.

To avoid inconvenient complications of negative values, a positive constant 1.0 is added to each attention function described above. Because only the differences between target, false target, and distractor strengths matter in this model, an added constant has no effect on model predictions.

The specific formulas are

$$\mathrm{ATT}_{1,0}(x) = 1 + \mathrm{e}^{-rf_1} \sum_{k=1}^{6} \frac{1}{2k-1} \sin(2k-1) f_1 2\pi x \quad (1)$$

$$\mathrm{ATT}_{1,\pi/2}(x) = 1 + \mathrm{e}^{-rf_2} \sum_{k=1}^{6} \frac{1}{2k-1} \sin\left[(2k-1) f_2 2\pi x + \frac{\pi}{2}\right] \quad (2)$$

$$\mathrm{ATT}_{2}(x) = 1 + \mathrm{e}^{-rf_3} \sum_{k=1}^{3} \frac{1}{2k-1} \sin(2k-1) f_3 2\pi x \quad (3)$$

$$\mathrm{ATT}_{3}(x) = 1 + \mathrm{e}^{-rf_4} \left[\sin f_4 2\pi x + \frac{1}{3} \sin 3 f_4 2\pi x\right] \quad (4)$$

$$\mathrm{ATT}_{4}(x) = 1 + \mathrm{e}^{-rf_5} \sin f_5 2\pi x \quad (5)$$

where $f_1 = f_2 = 1$ cpi, $f_3 = 2$ cpi, $f_4 = 4$ cpi, and $f_5 = 6$ cpi; cpi is cycles per image.

The third strength map describes the change in acuity with a change in location. The data indicate a clear decrease in proportion correct with an increase in the eccentricity of the target, consistent with decreased visual acuity with increased eccentricity. Here we use "acuity" as a abbreviation for "density of processing resources per square degree of visual angle" which is what determines visual acuity (except possibly in the central fovea). This aspect of the data (which results from known early limitations of the visual system) is modelled by using a modification of the derivative of the V1 cortical magnification function $\log(z + a)$ (Schwartz, 1984). The function used here is

$$\mathrm{ACUITY}(x, y) = \left(\frac{1}{\sqrt{x^2 + \alpha y^2 + \beta}}\right)^{\gamma} \quad (6)$$

where α determines the degree to which acuity differs along the horizontal versus the vertical, β determines the function's behavior at zero, and γ controls the rate of fall-off in acuity as eccentricity increases. The parameter γ is important here because some researchers (e.g., Stevens, 2002) report that acuity does not decline as rapidly as would be expected by the cortical magnification function.

The final strength map is calculated by multiplying the strength from each of the previous three maps at each location, resulting in a 12×12 matrix of strength values.

Up to this point the model is completely deterministic; given the same stimulus, the model will always produce the same response. The human observer will not. As estimated in Experiment 6, 76% of observer responses corresponded (i.e., the observer was either correct or incorrect on both presentations of the trial). For any model that matches the average performance of the observer, this 76% serves as an upper bound on the predicted trial-to-trial correspondence between the model and data. To model response variability, independent normally distributed noise with mean 0 and standard deviation σ_n is added to each of the 144 locations in the final strength map.

The decision phase of the model, with rare exceptions to be described below, selects the location with the highest strength as the target location. However, according to observer reports, there are trials in which the first choice for a target location in the search array is known to be in an unattended region. On these trials, observers report choosing one of three strategies. If the initial candidate location is close to an attended region, they select a nearby attended location as the response. Or, they may select a second location that they suspect may have contained the target. Otherwise, they report choosing "at random" another location from the attended regions. Observers report this happening rarely, but noticeably. In the model, such a higher level cognitive "veto" mechanism (corresponding to the subjective experience of which is described above) is assumed to operate after the initial selection of the location with highest strength.

Given this idea of a veto mechanism, why does the observer not do away with the attempt to suppress or attend specific regions, and simply perform a comparison of each target or false target location with the attention instructions and select the one which falls on an attended location? The answer is straightforward: observers are able to maintain only very few locations in memory for sufficient time to complete the comparison. Observer reports indicate that remembering the location of two target/false target items is feasible, but nothing much beyond this is available in the response phase. In order to have a reasonable chance that one of the very few locations that can be stored in short-term visual memory will be in the attended region, the attention system needs to suppress the irrelevant information at an early stage.

The model's decision mechanism operates as follows. If the location with the highest strength falls on an attended location, it is selected as the response. If not, the model checks if this unattended location is adjacent to an attended location. If it is adjacent, the model reduces the strength at the candidate unattended location by a small amount. If the unattended location is still the strongest, an adjacent candidate location is chosen as the response. Otherwise, the new strongest location becomes the candidate response. If this second location is again in an unattended region, the same steps are again followed. If the third strongest location is still in an unattended region, a location is selected at random. This veto mechanism only changes a handful of responses (always less than 5% of the data) and is of insignificant impact because it seldom arrives at a correct response.

Conceptually, the veto mechanism utilizes a memory that remembers only the three strongest elements. It is retained for two reasons: (1) A veto mechanism corresponds with subject experience. (2) A model that represents only early visual processing might, just like human perceptual processing, propose a candidate location that is far inside an unattended area. However, a human observer would never report such a location. If the model ever reported such a location, it would be powerfully rejected by a likelihood test. Ultimately, as models correspond more and more closely to human performance, it is necessary to incorporate into them more and more complex aspects of performance, such as the rare veto mechanism and, ultimately, typing mistakes, to achieve reasonable fits to the data. Or, one may discard a fraction of the data, so-called robust estimation. See Fig. 9(a) for an overview of the model.

10. Model implementation

The model described above was used to predict performance on a trial-by-trial basis (1440 trials) for each of three observers (CT, JG, and GA—data from Experiment 1 when only viewing ten-false-target trials). For each trial, the stimulus presented to the observer was used by the model to generate a response in the manner described above. The model parameters were optimized such that the 1440 predictions best accorded with the responses of an observer who had viewed these same stimuli, by maximizing the likelihood that the model's response was correct when the observer's response was correct, and incorrect when the observer's response was incorrect. In other words, for trials in which the observer was correct, the probability p that all ten false targets had a strength less than the target strength was maximized. For incorrect trials, $1 - p$ was maximized.

Responses for both the model and observer were classified as either correct or incorrect. The estimated model parameters were the distractor and target strengths for the stimulus map (D and T), the rate of frequency falloff r of the attention modulation transfer function, the shape of the acuity map (α, β, and γ), and the standard deviation of the noise (σ_n). The random jitter of each disk position in the display is not modelled here, as it is not anticipated to affect the process. The model parameters were estimated without the inclusion of the veto mechanism. Subsequently the veto mechanism was implemented to obtain the trial-by-trial correspondence. This mechanism was not included in the parameter fit because it is a relatively minor element of the decision phase, affecting decisions on a small (less than 5%) number of trials. Thus, 1440 individual predictions were generated and compared to the trial-by-trial performance data of the observer.

Table 1 displays the model performance for each of the three observers, as well as the parameter values giving rise to that performance. Note that parameters are quite similar across observers. For observer CT, the center of the acuity function was shifted to accommodate the shift in her peak performance in Experiment 1. Given her shift in performance by target location, it is likely that she was fixating at this shifted location rather than at the fixation point.

As displayed in Table 1, the model and observer behavior corresponded on approximately 70% of trials. It is important to note that this correspondence arises from both correct and incorrect responses.

While the parameters were optimized based on trial-by-trial comparisons, it is important to also compare the model performance across attention conditions with that of the observer. As shown in Fig. 10(a), the model performance is similar in nature to the observer performance. One consistent difference is in the $ATT_{1,0}$ condition; observers do better on this condition than on

Table 1
Model performance and parameter values for each observer

Obs	%	(c,c)	(i,c)	(c,i)	(i,i)	α	β	γ	D	T	r	σ_n
CT	70%	340	189	242	669	0.27	0.16	0.75	0.33	1.23	0.55	0.05
GA	70%	404	197	243	596	0.26	0.17	0.67	0.35	1.23	0.54	0.05
JG	69%	404	270	186	580	0.26	0.16	0.71	0.35	1.24	0.54	0.05

Each ordered pair, e.g., (c,c) represents (model performance, observer performance), where c correct and i incorrect, e.g., (c,i) is the number of trials in which the model was correct and the human observer was incorrect. % represents the average correspondence between consecutive runs of the model, using the parameters given here; α represents the weight of vertical relative to horizontal eccentricity in acuity; β, in cpd, indicates the size in degrees of visual angle of a central area of unchanging acuity; γ represents the rate of acuity fall-off with eccentricity; D indicates the strength of distractors; T indicates the strength of targets and false targets; r determines the amplitude of the modulation of attention at each frequency; σ_n is the standard deviation of internal noise.

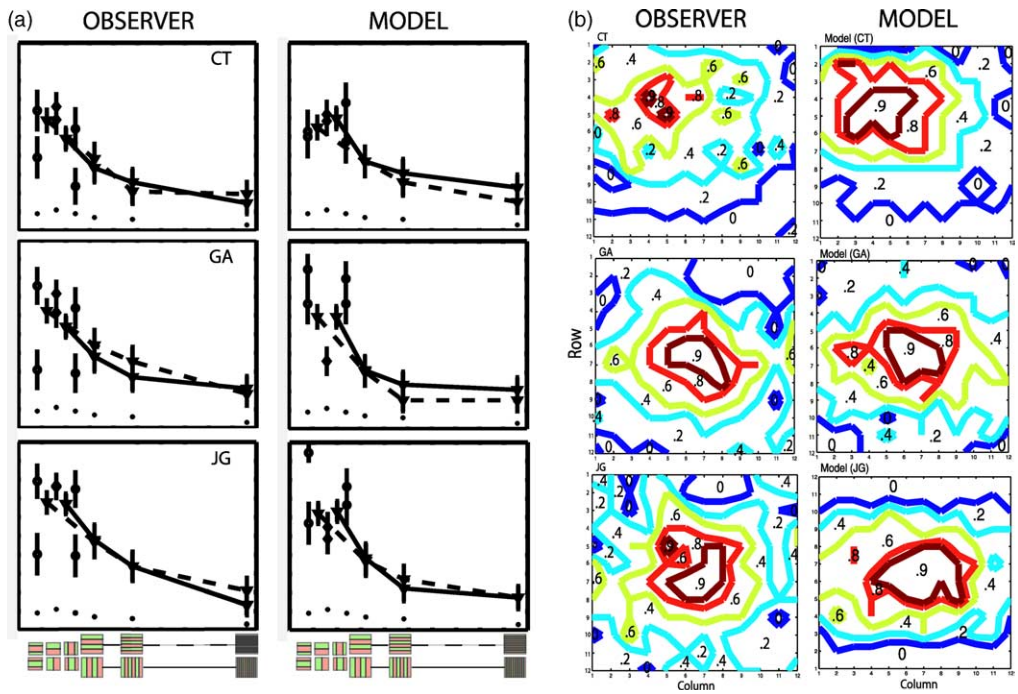

Fig. 10. (a) Observer and model performance across the attention conditions for three observers. The model runs shown are one instance for each subject, using the parameters as shown in Table 1. Dots indicate chance performance. Error bars indicate 95% confidence intervals. (b) Observer and model performance across the 144 possible target locations. Plotting conventions are identical to those in Experiment 1, Fig. 2.

the $ATT_{1,\pi/2}$ condition, while the model does not consistently show this pattern. One possible explanation for this difference is eye movements. The attention instruction is presented 750 ms prior to the onset of the search array. This is sufficient to allow observers to refixate to an optimal position. This optimal position is near the center of the display in all conditions except for $ATT_{1,0}$ where half of the visual field is to be attended and the other half is to be ignored. In this condition, it would be obviously optimal to refixate at the center of the attended region, placing more of that region in central vision and moving the unattended region farther into the periphery. Doing so would improve performance, yet the model does not do this. Observers were instructed to maintain central fixation throughout the task, but such maintenance was not independently verified.

Fig. 10(b) displays the model performance by target eccentricity, along with the corresponding observer performance. The nature of the contours is very similar, with a falling of performance with an increase in target eccentricity and a slower fall-off along the horizontal than along the vertical. Compared to the observer, the model does seem to perform better in the center of the display and worse along the edges, but given that the model was not optimized to maximize its correspondence to the target eccentricity data, but rather the trial-by-trial response, the similarity is quite encouraging.

The parameter values in Table 1 are similar across observers, and also have reasonable interpretations. The decay rate of the attention modulation transfer function, r, for instance, indicates a clear fall-off in attention modulation with an increase in the spatial frequency of the attention instructions. The target strength, T, is substantially higher than the distractor strength, D. The acuity parameters, α, β, and γ, produce an acuity function that is elliptical in nature, with a decrease in acuity with an increase in eccentricity as well as a slightly slower fall-off in acuity along the horizontal as compared to the vertical.

In addition to testing the model performance by optimizing the trial-by-trial correspondence with each of three observers, the model was also tested using the one-false-target trials from Experiment 3 as input. Using the parameters as optimized for JG (similar results obtain for the others) the model's performance on one-false-target trials is computed. Fig. 11 displays these results, which can be compared to corresponding observer performance in Fig. 3(a) and (b).

Lastly, the model's double-pass performance was evaluated using the optimized parameters twice and calculating the trial-by-trial correspondence of the two runs. Average trial-by-trial correspondence was around 82% for the model. This correspondence is higher than that of 76% found for observer correspondence, but further improvements currently underway to the methods of fitting the parameters are expected, in part, to

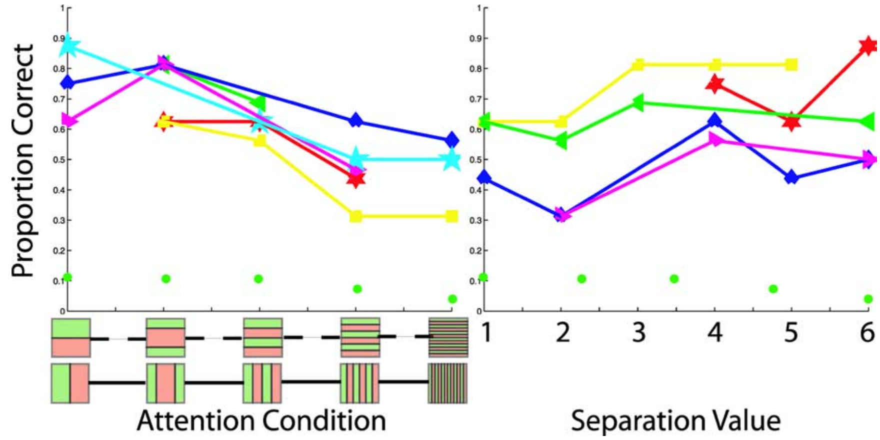

Fig. 11. Model predictions of performance in the one-false-target conditions. The left graph illustrates model performance as a function of attention condition. Each line represents a separation value: 6-point stars = 1, squares = 2, leftward triangles = 3, diamonds = 4, rightward triangles = 5, 5-point stars = 6. The right graph depicts model performance by separation value, where each line represents an attention condition: 6-point stars = $ATT_{1,0}$, squares = $ATT_{1,\pi/2}$, leftward triangles = ATT_2, diamonds = ATT_3, and rightward triangles = ATT_4. Dots indicate chance performance. Compare these model predictions to data of Fig. 3, leftmost panels.

improve this performance. Nevertheless, even at this stage, the model's performance is qualitatively similar to that of the observers and it exhibits most of the important features of the data.

11. Review and general discussion

The experiments implemented above were designed to investigate the characteristics and limitations of the spatial distribution of visual attention. By requiring observers to attend disjoint regions while suppressing intervening regions, the tasks provide data necessary to develop and test a general theory of visual attention.

A novel search task. In Experiment 1, a novel search task was implemented in which the presence of numerous false targets populated the unattended areas. The unique design of this task strongly penalized non-disjoint distributions of spatial attention. The false targets were critical in forcing attention towards various disjoint distributions. The data we obtained indicate how successfully spatial attention was able to conform to the requested distributions.

Characterizing disjoint distributions of attention. Previous research had tended to focus on whether or not attention could be distributed disjointly. While the answer to this question was mixed, studies did not arrive at a quantitative characterization of the extent to which spatial attention could conform to various requested disjoint distributions. Numerous studies found that observers did not seem able to simultaneously attend two or more locations while ignoring intervening locations (Eriksen & Yeh, 1985; Heinze et al., 1994; McCormick et al., 1998; Pan & Eriksen, 1993; Posner et al., 1980). Still others found that, while there were costs to disjointly distributing attention, observers could attend separate locations without necessarily processing regions in between (Awh & Pashler, 2000; Bichot et al., 1999; Castiello & Umilta, 1992; Eimer, 2000; Hahn & Kramer, 1998; Melchner & Sperling, 1978; Schmidt et al., 1998). Here we sought to use a systematic change of the required distribution of attention to gain insight into the ability of spatial attention to conform to these distributions. The general pattern of results here was that the ability of spatial attention to conform to requested disjoint distributions decreased with an increase in the spatial frequency of the requested distribution and decreased with an increase in target eccentricity.

To a considerable extent, observers were able to attend to multiple locations while strongly suppressing the intervening areas for low-spatial-frequency requested distributions and to a lesser extent for high-spatial-frequency distributions. The data demonstrate that a quantitative approach, rather than an all-or-none approach to the question "can attention be distributed disjointly" is more likely to lead to a formal theory of spatial attention.

Target to false-target separation: A possible confound. The possible confound in Experiment 1 was false target proximity, which covaried with attention condition. As the spatial frequency of the attention-instruction images increases, performance decreases. With ten-false-target stimuli this performance decrease could be explained, at least in part, by the increase in proximity of false targets to the true target (separation) that occurs when the spatial frequency requested by the attention instructions increased. To draw conclusions about the effect of the attention distribution, per se, it was important to rule out the effect of false target proximity.

Embedded, infrequent one-false-target trials. To separate the attention and target to false target separation effects, trials in which there was only one false target were created. The separation between the target and false target was varied, as was the attention condition under which each target-false target configuration was presented. Because the one-false-target trials were presented randomly on only 17% of trials and because they were unnoticed amongst ten-false-target trials, observers did use the same strategy in processing them as the ten-false-target stimuli.

The results were that, regardless of the target to false target separation, proportion correct in one-false-target trials declined with the increases in the spatial frequency of the attention instructions. These results show that the pattern of results observed with the ten-false-target stimuli cannot be explained by false target proximity. Additionally, performance on the one-false-target trials was consistently better than performance on the ten-false-target trials. This indicates that suppression of the false targets in unattended regions is not complete.

The spatial characterization of "pure", unconfounded, spatial attention. The comparisons of one-false-target and ten-false-target stimuli, indicate that while the number of false targets somewhat affects search accuracy, the spatial distribution of false targets does not (over the range of separations explored). Most important, there is a large residual effect of attention, per se. Identical stimuli were responded to quite differently when the observer was attempting to conform to different requested attention distributions.

Foreperiod variations distinguish early low-level from high-level attentional processing. Experiment 2 provided evidence that the search task involved familiar attention mechanisms as follows. The ideal attention task uses identical stimuli under different attention conditions. When the different attention conditions produce different performances while the same stimuli and response alternatives remain the same, it clearly defines an effect that can be attributed only to attention. We wished to compare spatial attention that is attempting to conform itself to spatial locations with neutral spatial attention. This requires a different paradigm than that of Experiment 1. In Experiment 2, the foreperiod between attention instructions and the search array was varied. It included a condition in which the attention instructions occurred after the search array was turned off. The large performance difference between trials with identical attention instructions, some of which appeared before the search array and others after, indicated clearly that observers entered a non-neutral attentional state (e.g., attending a particular spatial configuration) when the instructions were presented first. When they were presented after the search array, observers remained in a neutral state and performed very poorly. The different foreperiods (from attention instructions to search array) allowed for the estimation of the length of time required to switch from the neutral attention state to the specific state (e.g., attend to a particular configuration). Attentional preparation (conforming to a particular requested spatial distribution of attention) was initiated about 50 ms after an attention instruction was received and it was complete in about 350 ms.

Scale invariance and its implication for attentional control processes. Results from the first two experiments explored the spatial frequency of the attention instructions. Experiment 3 sought to specify whether the results should be understood in terms of absolute or relative spatial frequency. To accomplish this, the task from Experiment 1 was repeated at two viewing distances: 120 cm (the original distance) and at 60 cm. If performance depends on absolute spatial frequency, this change in viewing distance should alter performance since all spatial frequencies—attention instructions and stimulus frequencies—would be doubled by halving the viewing distance. The results showed no significant change in performance with the change in viewing distance. This suggests that the distribution of visual attention may be understood in terms of object spatial frequency versus retinal spatial frequency. Dependence on object versus retinal spatial frequency in turn suggests that the attention control signal may be generated at a high level of perceptual or cognitive processing. Alternatively, it is possible that appropriate spatial inhomogenities could result in scale invariance in the detection task even if attention control were generated at a lower level of processing than an object level.

Is attention-cued search an abstract form of a conjunction search? Experiment 4 was designed to compare an explicit attention-cued conjunction task with the attention tasks of Experiments 1–3. The red-green grating that defined the attention instruction remained on throughout the duration of the trial. Thereby observers could use the color of an attended area to perform a conjunction search in which a target was defined by its large size and surrounding color, false targets by large size and the opposite surrounding color, and distractors by their small size on either color. Experiments 1–3 had involved an implicit conjunction of "attended area" with disk size versus the explicit conjunction of color with disk size. Performance in the attention-cued conjunction task was consistently better than on the attention-cued tasks, with most of this improvement occurring at the highest spatial frequencies. Other than this, the general pattern of results did not change.

This similarity of attention-cued conjunction results to simple attention cuing suggests two things. First, it might be useful to consider that, at the point in the brain where the salience of a location is being computed, attention contributes similarly to a target-defining feature. Attention adds to the salience or strength at that attended location in the same way that a target-defining color does. Second, it suggests that the resolution of the color feature map is better than the resolution of the attention feature map, thereby producing much more benefit at the highest spatial frequencies. Fig. 9(b) shows how the model structure can be augmented to incorporate both attention and feature information.

The three main conclusions. Overall, Experiments 1–4 show that observers can distribute attention in accordance with various requested configurations, but that there are severe limits to this distribution. These limits on the distribution depended on a number of factors: (1) the relative spatial frequency of the distribution, (2) the eccentricity of the target, and (3) the amount of interfering information in the unattended regions, specifically, number of false targets. There are certainly other factors, but these three are important and they were studied by means of the novel multiple-false-target search task described above.

A formal computational model for spatial attention? The goal of developing such a task was to find and explore those factors involved in the limiting of the spatial distribution of visual attention. From the identification of these factors, a general model of visual attention was developed and implemented. While there are certainly further developments and improvements to be made, the model has the promise to simply and systematically describe the spatial distribution of visual attention in human observers.

Combining three strength maps. The current version of the model combines three strength maps to arrive at a final strength map from which a response is chosen (most often the location with the highest strength): an attention modulation map, an acuity map, and a conjunction-feature map. This sort of structure is not novel; a similar structure can be found, for example, in the Guided Search Model 2.0 of Wolfe (1994). That attention may itself be considered as a feature is, however, a promising and interesting addition to this type of structure. The specific structure of the three strength maps in the present model, their combination, and the decision process together incorporate the important factors relating to visual attention identified above.

The attention map addresses the reduced performance with an increase in spatial frequency through the use of an attention modulation transfer function and linear systems analysis. Attended regions are enhanced and unattended regions are suppressed, more so at lower spatial frequencies and less so with higher spatial frequencies. The relationship between the spatial frequency and the amplitude of attentional modulation at that frequency is well described by a function in which the modulation amplitude of attention decays exponentially with frequency. To compute the attentional modulation in response to a requested distribution of attention, first analyze the requested distribution into its component sine waves, look up or compute the response to each of these sine-wave components for the attention frequency–response function, and then sum the component responses.

The acuity map addresses the reduced performance with an increase in target eccentricity. Each of the 144 locations is assigned a strength using a function with three parameters—lower strengths for the more eccentric locations and higher strengths for those nearest fixation. Additionally, the strengths decline more slowly along the horizontal dimension of the display than along the vertical.

Model predictions. The reduction in performance when presented with ten false targets as compared to one false target is automatically incorporated into the model in a relatively simple manner: It is more likely that the strength at a false target location is higher than that of the target location when there are ten false targets than when there is only one. Therefore, the model is more likely to select an incorrect response when there are more false targets.

After parameter optimization, the model performance was evaluated by attention condition and by target eccentricity. In addition to the trial-by-trial correspondence, similar overall patterns were also observed. Parameter values were similar across observers, and their relationship to one another can be understood in a reasonable manner, reflecting the theory underpinning the model.

The future of a linear systems approach to the characterization of spatial attention remains promising given the foundation provided here. Even with the current model, it is possible, in principle, to predict performance on tasks with almost any novel or more complex request distribution of attention. The circumstances under which these predictions work remain to be determined.

12. Conclusion

The search paradigm developed here allows for the investigation of the spatial distribution of visual attention. By manipulating the stimulus configuration, the ability of the attention distribution to conform to a requested distribution can be systematically studied. By presenting "false targets" in the unattended areas, the paradigm forces observers to ignore those areas and to more closely conform to the requested attention distribution. Successful performance in the search task hinges upon the ability to "split the attention beam."

Results show that the modulation amplitude of attention decreases with an increase in the requested spatial frequency. Here, attention was modelled as a sum of sine-waves with fundamental frequency equivalent to the spatial frequency of the requested attention distribution. With this relatively simple, theoretically based structure, agreement between observer data and model behavior on a trial-by-trial basis, as well as over attention conditions and target location, was quite reasonable. This agreement lends support to this approach to the modelling of spatial attention.

The linear systems approach to the modelling of spatial attention has considerable promise. This approach enables a prediction of the achievable distribution of attention in response to any requested distribution. In a linear systems approach, any requested distribution of attention can be described mathematically by a sum of sine-wave components with different frequencies, amplitudes, and phases. Once the distribution of attention in response to each of these sine-wave components has been determined, the predicted attention distribution in response to any arbitrary requested distribution becomes a routine computation.

The experiments described here dealt with simple gratings and provide the foundation for a Fourier-based theory of attention. Fourier theory provides a well-established basis for a comprehensive theory of spatial attention. Future experiments will test how generalizable the theory is to arbitrary distributions of attention.

Acknowledgements

This research was supported in part by the US Air Force Office of Scientific Research, Life Sciences Directorate.

References

Awh, E., & Pashler, H. (2000). Evidence for split attentional foci. *Journal of Experimental Psychology: Human Perception and Performance, 26*(2), 834–846.

Bichot, N., Cave, K., & Pashler, H. (1999). Visual selection mediated by location: feature-based selection of noncontiguous locations. *Perception & Psychophysics, 61*(3), 403–423.

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Burgess, A. E., & Colborne, B. (1988). Visual signal detection. Observer inconsistency. *Journal of the Optical Society of America, 5*(4), 617–627.

Carrasco, M., Talgar, C., & Cameron, E. L. (2001). Characterizing visual performance fields: effects of transient covert attention, spatial frequency, eccentricity, task and set size. *Spatial Vision, 15*(1), 61–75.

Castiello, U., & Umilta, C. (1992). Splitting focal attention. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 837–848.

Eimer, M. (2000). An erp study of sustained spatial attention to stimulus eccentricity. *Biological Psychology, 52*, 205–220.

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

Eriksen, C., & Yeh, Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 583–597.

Hahn, S., & Kramer, A. (1998). Further evidence for the division of attention among non-contiguous locations. *Visual Cognition, 5*(1–2), 217–256.

Heinze, H.-J., Luck, S., Munte, T., Gos, A., Mangun, G., & Hillyard, S. (1994). Attention to adjacent and separate positions in space: an electrophysiological analysis. *Perception & Psychophysics, 56*(1), 42–52.

Kramer, A., & Hahn, S. (1995). Splitting the beam: distribution of attention over noncontiguous regions of the visual field. *Psychological Science, 6*, 381–386.

LaBerge, D. (1995). *Attentional processing: the brain's art of mindfulness. Perspectives in cognitive neuroscience*. Cambridge, MA: Harvard University Press.

LaBerge, D., & Brown, V. (1989). Theory of attentional operation in shape identification. *Psychological Review, 96*, 101–124.

McCormick, P., Klein, R., & Johnston, S. (1998). Splitting versus sharing focal attention: comment on Castiello and Umilta (1992). *Journal of Experimental Psychology: Human Perception and Performance, 24*(1), 350–357.

Melchner, M., & Sperling, G. (1978). The attention operating characteristic: examples from visual search. *Science, 202*, 315–318.

Pan, K., & Eriksen, C. (1993). Attentional distribution in the visual field during same–different judgments as assessed by response competition. *Perception & Psychophysics, 53*(2), 134–144.

Parish, D., & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research, 31*(7–8), 1399–1415.

Pelli, D. (1997). The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Podgorny, P., & Shepard, R. (1983). Distribution of visual attention over space. *Journal of Experimental Psychology: Human Perception and Performance, 9*(3), 380–393.

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

Posner, M., Snyder, C., & Davidson, B. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*(2), 160–174.

Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review, 93*(2), 180–206.

Schmidt, W., Fisher, B., & Pylyshyn, Z. (1998). Multiple-location access in vision: evidence from illusory line motion. *Journal of Experimental Psychology: Human Perception and Performance, 24*(2), 505–525.

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

Shih, S., & Sperling, G. (2002). Measuring and modelling the trajectory of visual spatial attention. *Psychological Review, 109*(2), 260–305.

Sperling, G., & Dosher, B. (1986). *Strategy and optimization in human information processing*. New York: Wiley (pp. 1–65, Chapter 2).

Sperling, G., & Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. *Psychological Review, 102*(3), 503–532.

Stevens, C. F. (2002). Predicting functional properties of visual cortex from an evolutionary scaling law. *Neuron, 36*, 139–142.

Tsal, Y. (1983). Movements of attention across the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 523–530.

Wolfe, J. (1994). Guided search 2.0: a revised model of visual search. *Psychonomic Bulletin & Review, 1*(2), 202–238.