



The Spatial Structure of Visual Attention

Cathryn J. Downing
Stanford University

Steven Pinker
Massachusetts Institute of Technology

ABSTRACT

In two experiments we examined the distribution of attention across visual space and the properties of the mental representation of space underlying visual attention. In Experiment 1, subjects focused their attention on various locations in a three-dimensional display, and we measured the "costs" in detection time for stimuli at unattended locations varying in horizontal distance and in depth from the attended location. Results suggest that attention falls off with depth from the focus of attention, and more steeply for stimuli that are farther than the focus of attention than for those nearer. Furthermore, attention falls off with lateral distance from the focus of attention according to a negatively accelerating gradient defined over visual-angle separation, not real-world distance. Control conditions confirm that these effects are not artifacts of ocular accommodation.

In Experiment 2, subjects attended to one of a set of 10 locations on a CRT. Detection times for stimuli at unattended locations reveal a gradient of two-dimensional attention whose slope is related to differences in receptive field size or cortical magnification at different retinal eccentricities. We suggest that visual attention is defined in a three-dimensional representational medium whose dimensions are horizontal and vertical visual angle, scaled by a cortical magnification factor, and depth.

INTRODUCTION

In this chapter we report an experimental investigation of the way that cognitive processes can access portions of the visual world by virtue of their locations. In particular, we attempt to assess the dimensionality and metric of the internal representation of visual space that underlies visual attention.

In our studies we exploit the fact that although humans usually focus their eyes and their attention on the same location in visual space, movements of attention can occur independently of eye movements (Eriksen & Hoffman, 1972; Posner, 1978, 1980; Sperling, 1960; Sperling & Reeves, 1980). A simple demonstration of this phenomenon can be found in a set of studies by Posner and his colleagues. They have shown that when subjects expect an event such as the illumination of a light to occur in a particular location, subjects can detect the light more quickly when it does occur in that location than when they have no prior expectation about its position. On the other hand, when the light occurs at a location other than the expected location, the subjects are slower to detect it than they are when they have no prior expectation about its location. The facilitation observed for stimuli at expected locations is often referred to as an attentional *benefit*; the inhibition observed for stimuli at unexpected locations is referred to as an attentional *cost* (Posner, 1978).¹ These attentional effects occur even when a person's eyes are fixated somewhere other than the expected location (Posner, 1978; see also Remington, 1980). Further support for the separability of the eye movement and attentional systems can be found from studies of primate electrophysiology (e.g., Wurtz, Goldberg, & Robinson, 1982) and human neuropsychology (Helman, 1979).

Investigations of the organization of cortical visual processing have revealed a number of retinotopically organized areas whose parts represent specific aspects of the visual world at particular locations (see Cowey, 1982). It seems reasonable to postulate that visual attention can operate by activating, priming, or selecting parts of these structures. If visual attention does work in this way, several questions arise. First, do the structures subserving attention represent three-dimensional space or two-dimensional retinal projections? Second, do the structures subserving visual attention represent the position of stimuli along the horizontal and vertical axes of visual space in terms of visual-angle separation or in terms of real-world distance? Third, when a particular location is attended to, what shape does the attentional "spotlight" take? Posner, Snyder, and Davidson (1980) and Hoffman and Nelson (1981) have shown that regions adjacent to attended locations can share some of the facilitation allocated to the attended location. However, we do not know whether the level of attentional facilitation for unattended locations declines as a function of their distance from the attended location, nor the shape of such a decline, nor whether such a decline is homoge-

¹There is disagreement over the extent to which attentional effects in the detection of luminance increments reflect differences in the quality of the internal representation of the stimulus, differences in the amount of information about the visual world that the visual system loses, or differences in the amount of information about the stimulus the observer accumulates before deciding that the event has occurred (see Bashinski & Bacharach, 1980; Posner, 1978; Shaw, 1984; and Sperling, 1984). However, the questions addressed in this chapter, concerning the metrics according to which visual locations are selected, are largely independent of the question of exactly what is done with the selected information.

neous over the entire retina or changes shape depending on where attention is centered.

EXPERIMENT 1

In this experiment, we use a modification of the attention task used by Posner et al. (1980) in conjunction with a three-dimensional display in order to address the first two questions raised in the preceding paragraph.

Method

Subjects. Sixteen subjects participated in the experiment either for course credit or for pay. An additional subject was eliminated because he did not show simple attentional costs and benefits. All subjects were members of the Stanford University community, and had either normal or corrected-to-normal vision.

Apparatus. Each subject placed his or her chin and forehead in a restraint and with the left eye covered by an eyepatch, viewed a 102 cm × 91 cm rectangular platform lying horizontally on a table. Four small lights, .7 × 1.0 cm, were mounted on vertical stalks on the platform. There were eight possible positions for the lights, forming two parallel, curved rows of four positions each, one row behind the other. The near row was 101 cm from the subject; the far row was 171 cm. During each of the four sessions of the experiment, the subject saw four lights: Half of the subjects saw the lights at 10° left, 5° left, 5° right, and 6° right of fixation, and the other half saw them at 6° left, 5° left, 5° right, and 10° right. This yielded retinal separations of 1°, 5°, 10°, 11°, 15°, and 16° of visual angle between pairs of lights. Although the retinal position of the lights remained fixed across sessions, the lights' distance in depth along the subject's line of sight was varied from session to session. In the four configuration conditions that we used, the positions of the four lights (listed from left to right) in depth were as follows: (1) near, far, near, far; (2) far, near, near, far; (3) far, far, near, near; and (4) near, near, near, near. These particular configurations were chosen because they satisfied two constraints: that each pair of lights be seen equally often with the two lights at the same depth and with one of the lights at the near depth and one at the far depth; and that the positioning of the lights would not encourage subjects to favor one depth over the other, as might have been the case if three lights had been positioned at one depth and the fourth had been positioned at another.

A 1 × 2 cm LED chip that could display a digit was mounted on the center of the platform at a depth of 127 cm. The positions of the near lights, digit display, and far lights in depth were chosen so that the near and far lights were equally in focus when the digit display was fixated. The lights and the digit display were

positioned vertically so that their projections would all lie along the same horizontal line when viewed by the subject. Because the lights fell outside Panum's area, binocular viewing would lead to double images, so subjects viewed the display monocularly. The platform was covered with wide-wale corduroy and illuminated with spotlights so that linear perspective and a texture gradient could serve as depth cues. The apparatus was controlled in real time by a microcomputer.

Procedure. On each trial, subjects fixed their eyes on the central chip, which displayed a digit from 0 to 4 indicating either that a certain light (1, 2, 3, or 4) would subsequently be illuminated with high probability, or that the illumination of any of the four lights was equiprobable (0). The subjects' task was to attend to the light indicated, or to none of the lights if they were equiprobable, without moving their eyes from the central chip, and then to press a response key whenever *any* of the four lights was lit. On most of the trials, the light to which subjects were attending was the light illuminated, but on a small percentage of the trials, one of the three unattended lights was illuminated instead.

Four blocks of 185 trials were run with each of the four configurations of the lights. Sessions using a given configuration lasted 1 hour; subjects were run in either 1- or 2-hour sessions. Within each block, 80% of the trials cued the subject to attend to one of the lights and 20% cued him or her not to attend to any particular light. A light was turned on following the cue on only 76% of the trials (the remaining trials were catch trials). When a light was turned on, 79% of the time it was the cued light, and 21% of the time it was one of the other three lights, each of the three occurring with equal likelihood. Thus, for each subject, there were 32 trials for each of the 12 combinations of retinal separation and depth separation between cue and stimulus. Each light was cued equally often, and catch trials were evenly distributed across cue types. Each subject received a different random ordering of the trials for each block of trials.

An "error" tone informed the subjects when they made either anticipatory responses or eye movements (discussed later), and the digit display flickered to reward quick responses (i.e., responses that were as fast or faster than the mean reaction times of trials in which the stimulus had occurred at the expected location and trials in which no expectancy had been set up). Trials on which errors were made were rerun at a randomly selected time later in the block.

Half the subjects wore skin electrodes that detected lateral eye movements in either direction of 5° or more. The order of configuration conditions, type of retinal projection (whether the 10° light was to the left or right of fixation), and monitoring of eye movements were counterbalanced across subjects.

Each trial began with a 1000 msec intertrial interval, followed by presentation of the cue on the central chip. The interval between the presentation of the cue and the illumination of a light varied randomly (according to a rectangular distribution) across trials, with cue lengths ranging from 400 to 800 msec in

noncatch trials, and lasting a fixed length of time beyond that in catch trials. (For each subject, the additional fixed length of time on catch trials was equal to the subject's mean reaction time on the practice trials plus two standard deviations). A response or an eye movement ended both the display of the cue and the illumination of the light, and was followed by the feedback period, lasting 500 msec.

Subjects initiated each block of trials by pressing a separate "start" key. Within each block of trials the subject was given an opportunity to take a short break from the task after every tenth trial. Trials resumed after these breaks when the subject pressed the "start" key again. Otherwise, trials were initiated automatically.

Results and Discussion

Attentional costs were calculated as the amount of time that subjects required to detect a light in a particular position when they had been attending to some other location (the "unexpected" reaction time), minus the time required to detect that light when attention was not directed to any particular location (the "neutral" reaction time). These costs were computed separately for cue-stimulus pairs corresponding to different retinal separations and different separations in depth. Because retinal separations of 10° and 11° and those of 15° and 16° were so close, we averaged their costs, so that our analyses were of costs for separations of 1°, 5°, 10.5°, and 15.5°. Reaction times longer than 1000 msec or shorter than 100 msec were discarded.

By analyzing costs rather than detection times per se, we removed the component of the detection times attributable to perceptual properties of the stimulus, such as those related to the retinal eccentricity or distance in depth of the particular light illuminated. This assumes that effects of attentional activation and intrinsic properties of the stimulus are additive, an assumption that Posner's (1978) findings suggest is true, and that we examine later in this section and in Experiment 2.

We expected costs to increase as the retinal separation between cued and illuminated lights increased, reflecting the shape of the fall-off of attentional facilitation with increasing horizontal distance in the mental representation of space subserving visual attention. More importantly, we expected that if the structure or structures subserving attention represent depth, costs would be greater when the two lights were separated in depth than when they were at the same depth; conversely, if depth is not represented, costs would vary only with retinal separation, and not with separation in depth. Finally, by comparing the increase in costs with lateral distance between illuminated and attended lights at near and far positions in depth, we hoped to determine whether distance along the horizontal axis of visual space is represented in units of retinal or real-world distance.

Figure 8.1a shows that attentional costs increase according to a negatively accelerating function as the retinal separation between cued and illuminated lights increases. Furthermore, costs are greater when the lights are at different distances in depth than when they are at the same distance; this effect, however, is pronounced only at the two largest retinal separations. An analysis of variance shows that the effects of retinal separation— $F(3, 45) = 89.16, p < .001$ —

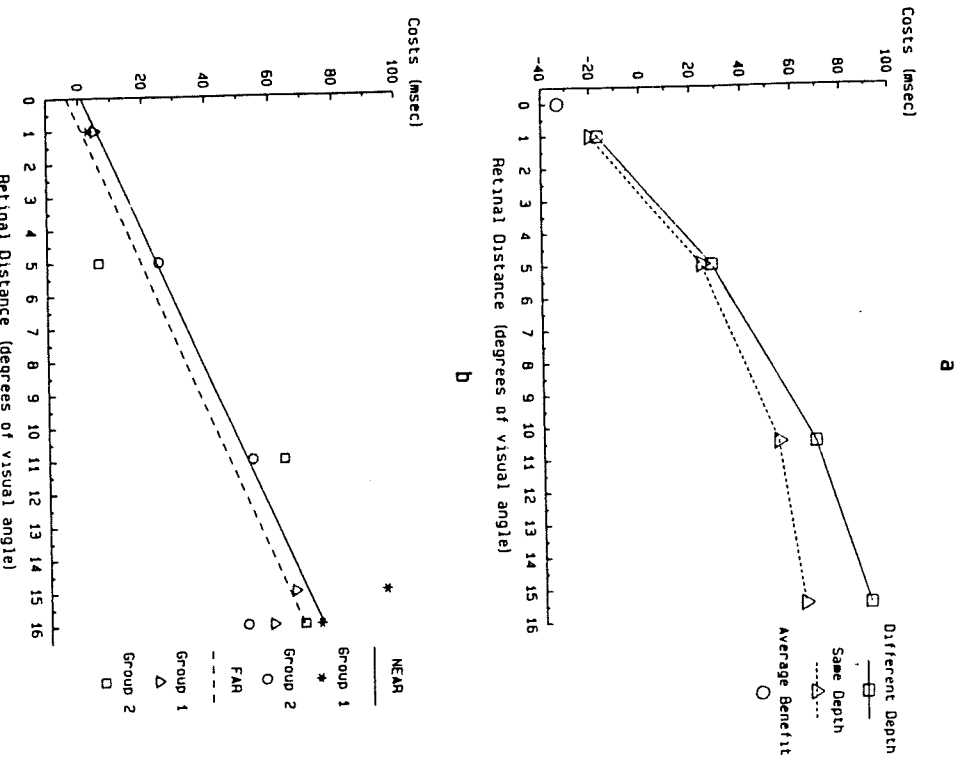


FIG. 8.1. (a) Attentional costs as a function of the retinal separation and separation in depth of the attended and illuminated lights. The average benefit is plotted as the cost for lights separated by 0° . (b) Same-depth costs as a function of the retinal separation and position in depth of the attended and illuminated lights. Group 1 and 2 saw the lights at different absolute retinal positions (see text for description of these two retinal configurations).

separation in depth— $F(1, 15) = 7.28, p < .05$ —and the interaction of these two effects— $F(3, 45) = 3.22, p < .05$ —are significant; a trend analysis shows significant linear and quadratic trends for retinal separation— $F(1, 15) = 132.86, p < .001$ and $F(1, 15) = 12.04, p < .01$, respectively—and a significant interaction between depth separation and the linear component of retinal separation— $F(1, 15) = 7.38, p < .05$. In addition, costs were significantly greater when the cued light was near and the illuminated light far (mean cost = 48 msec) than vice versa (mean cost = 37 msec)— $F(1, 15) = 6.89, p < .05$.

Figure 8.1b breaks the "same-depth" data from Fig. 8.1a into two parts, corresponding to whether the attended and illuminated lights were both near or both far.² With fewer data, the cost function is not as smooth, but there are no apparent differences between near–near and far–far costs. One way to assess possible differences statistically is to perform a trend analysis. When this is done, the linear component of the increase in attentional cost over visual angle is the same for near and far depths (for near depths: slope = 4.67 msec/deg, $r = .88$; for far depths: slope = 4.58 msec/deg, $r = .94$), as can be seen from the parallel regression lines plotted in Fig. 8.1b. The difference in slopes is not significant. Furthermore, there is no significant interaction between position in depth and the quadratic component of retinal separation. Thus, although we found that *separation* in depth had a significant effect on attentional costs, we found no effect of the *absolute position* in depth. Attentional costs increased at a greater rate for different-depth cue-stimulus pairs than for same-depth cue-stimulus pairs, but they increased at the same rate for all same-depth cue-stimulus pairs, regardless of whether the cued location and stimulus were both near or both far.

The results of this experiment suggest two conclusions. First, the representation of space underlying visual attention contains depth information.³ Second, it appears that in this representation, the depth information is not used to convert the retinal metric of visual angle into a metric preserving size constancy.

There were no significant differences between the results with and without eye-movement equipment. Although costs were greater for subjects who saw the

²Note that not all of the same-depth data from Fig. 8.1a were used in this comparison, because same-depth retinal separations involving the position 5° to the right of fixation were always near. For this comparison, we had to include only same-depth retinal separations that appeared at both the near and far depths. Note also that before plotting Fig. 8.1b, we regressed the costs against visual angle separately for the two groups of subjects who viewed different configurations of lights, and subtracted the intercept for each group from each data point for that group, so as to equate for intergroup differences in gross overall costs. This subtraction does not affect our conclusions. The slopes for the near and far regression lines did not differ from one another either before or after equating group means. Tests of the interaction between the polynomial trends and position in depth were performed separately for the two groups.

³In an unpublished experiment with nine subjects we have replicated the finding that there are significantly greater costs when a stimulus appears at an unattended depth than when it appears at an attended depth.

most peripheral light in the left visual field than for those who saw it in the right visual field— $F(1, 14) = 5.30, p < .05$ —this effect did not interact with any of the other effects in the study. The finding that costs are greater when peripheral stimuli are presented in the left visual field may be a reflection of hemispheric asymmetries in visual attention (see Heilman, 1979, p. 298), but because all subjects viewed the display with the right eye only, the difference could be attributable to temporal-nasal differences in attention instead.

One concern over the results we report is that they may in part reflect a nonadditive combination of attentional facilitation and differences in intrinsic detectability due to retinal eccentricity. This concern arises because the degree of retinal separation between cue-stimulus pairs was not completely orthogonal to the retinal position of the stimulus in this experiment. However, we can rule out the effects of such interactions in producing our results in several analyses.

First, in the condition in which attention was not directed to any particular location prior to the onset of the light, we found no significant differences in the detection times for lights at different eccentricities and distances in depth (all p 's $> .10$). Second, when we examined the effect of depth separation separately for each of the four absolute retinal positions ($10^\circ, 5^\circ$ on the side containing the light at $10^\circ, 5^\circ$ on the side containing the light at $6^\circ, 6^\circ$), collapsing across the two groups of subjects, we found: (a) greater costs for different-depth trials than for same-depth trials in all four cases (this difference was significant in three out of four cases— $p < .05$); (b) significantly greater costs for greater retinal separations in all four cases ($p < .05$); and (c) an interaction between retinal separation and depth similar to that found in the mean data in three cases (this was significant in two of those cases— $p < .05$). Third, we conducted an analysis using absolute retinal position ($5^\circ, 6^\circ$, or 10° eccentricity) and depth separation as factors (holding retinal separation between cued position and stimulus position constant at 15° or 16°), and found a significant effect of depth separation— $F(1, 14) = 10.44, p < .01$ —but not an effect of absolute retinal position, nor an interaction between depth separation and retinal position (all p 's $> .10$).

It is also worth noting that the overall interaction whereby depth separation affected costs for the large but not the small retinal separations (see Fig. 8.1a) cannot be attributed to differences in absolute eccentricity of the stimuli. For 5° and $15^\circ/16^\circ$ retinal separations there were both relatively central and relatively peripheral stimuli; for 1° and $10^\circ/11^\circ$ retinal separations there were only relatively central stimuli. Thus, the eccentricity of the stimulus does not predict whether or not a depth effect will be found; of the retinal separations involving relatively central stimuli, one showed a depth effect ($10^\circ/11^\circ$) and the other did not (1°), and of the retinal separations involving both relatively central and relatively peripheral stimuli, one showed a depth effect ($15^\circ/16^\circ$) and the other did not (5°).

Two other alternative explanations for the depth selectivity that we have found arise from the possibility that when a subject attends to a particular depth

in response to the digit cue, he or she also accommodates the lens of the eye to that depth. However, these explanations can be shown to be highly unlikely.

One of these explanations is that when the stimulus is presented at a depth other than the attended/accommodated depth, subjects reaccommodate to the depth of the stimulus, and that the reaccommodation process interferes with the response process, thereby slowing responses to stimuli at the unattended depth. This alternative is unlikely because we found significant depth effects for stimuli too peripheral to elicit accommodative responses. In particular, we found a significant depth effect for cue-stimulus pairs separated by 15° and 16° when the stimulus was presented at 10° eccentricity— $F(1, 7) = 6.09, p < .05$ —even though no accommodative response occurs to stimuli at 10° in situations, such as our experiment, in which the only accommodative stimulus is blur (see Ciuffreda & Kenyon, 1983, pp. 112–114).

The second alternative explanation is that responses to stimuli at the unattended depth are slower because subjects accommodate to the attended depth, thereby making stimuli at the unattended depth more blurred than those at the attended/accommodated depth. Because blurring can decrease contrast, the greater costs at unattended depths could be attributed to slower detection of low-contrast or poorly resolved stimuli. This possibility was ruled out by a control experiment in which we told people to accommodate to a particular depth (either 101 cm or 171 cm away) and then to detect stimuli either at the same depth or at a different depth; no attentional instructions were involved. We found that the blur produced at the unaccommodated depth had no effect on reaction times (mean for same depth = 331 msec, mean for different depth = 333 msec, $F(1, 7) = 1.29, p > .10$). Thus, it appears that even if subjects did accommodate the lens of the eye to the depth where they expected a light to occur in Experiment 1, the blur that this produced at the unaccommodated depth would not have been sufficient to account for the depth effect that we found. Presumably, this is because the poor resolution in peripheral regions of the retina makes the increase in sharpness of the retinal image brought about by accommodation negligible.

EXPERIMENT 2

In Experiment 1, we assumed that the retinal position of a stimulus would not have any effect on the magnitude of attentional costs or benefits for that stimulus (see Posner, 1978). Although the analyses we performed on the data from Experiment 1 did confirm this assumption, we wanted to test it more explicitly. In particular, we wanted to determine whether the two-dimensional gradient of visual attention is identical regardless of the retinal position where it is centered; in other words, whether the visual field is homogeneous with respect to attentional selection of information by location.

There are several ways this null hypothesis of visual field homogeneity could be false. Costs and benefits could be multiplicative with differences in intrinsic

detectability arising from eccentricity. In addition, overall costs and benefits could change depending on where on the retina the gradient is centered: Some parts of the retina might be intrinsically more "attendable" than others. Changes in resolution with retinal eccentricity (arising from the change in receptive field size and the related change in cortical magnification with eccentricity) could also make the gradient appear more or less sharply peaked depending on whether it lay in highly resolved or poorly resolved regions. Finally, the gradient of attention could be defined smoothly over both visual fields, or it could have different properties in the visual field where attention is focused and in the opposite field.

The only data we know of relevant to these issues come from three standard cost-benefit attention experiments conducted by Posner and his colleagues (see Posner, 1978, pp. 198-202). In these experiments, the researchers manipulated the eccentricity of two possible stimulus locations, and found that costs and benefits were the same whether stimuli were at 0.5°, 6.9°, or 25° from fixation. The similarity of costs and benefits for foveal and peripheral stimuli allowed them to conclude that (Posner, 1978) "costs and benefits from a voluntary shift of visual attention are the same, regardless of the eccentricity of the events" (p. 202).

In this experiment, we address this issue in more detail by comparing costs and benefits for various retinal separations across a wide range of positions of the cued location and stimulus. The method is similar to that of Experiment 1, except that our display was two-dimensional rather than three-dimensional (i.e., all stimulus locations were at the same depth), and contained 10 equally spaced stimulus locations rather than four unequally spaced locations.

Method

Subjects. Twelve Stanford University students participated in this experiment for either course credit or pay. As in the previous experiment, all subjects had either normal or corrected-to-normal vision. An additional four subjects did not complete the experiment because of problems in calibrating the eye-movement equipment or in the subjects' ability to follow the task instructions.

Apparatus. Subjects viewed a CRT monitor screen that measured approximately 37 cm horizontally. We again used a chin rest with a forehead restraint to position subjects' heads properly, and an eyepatch to cover each subject's left eye. Subjects fixated a cue location at the center of the screen that was positioned directly in front of the right eye. Ten unfilled boxes measuring about $1^\circ \times 1^\circ$ were displayed in a horizontal row on the screen, located at 1.25°, 3.75°, 6.25°, 8.75°, and 11.25° from fixation in both directions. Each box was thus 2.5° from its neighbor. Above each box, a number from 1 to 10 was displayed, corresponding to the position of the box along the horizontal axis; the leftmost box was labeled "1" and the rightmost box "10." This display was present at all times during the experiment. As before, a microcomputer controlled the experiment.

Procedure. The procedure for this experiment was similar to that for Experiment 1. Subjects maintained eye fixation on the central cue location, where they saw a digit from 0 to 10 that indicated to which box, if any, they should attend. Their task was simply to press a response key as quickly as possible, without making an eye movement or an anticipation, whenever any one of the boxes was filled in.

Subjects participated in four experimental sessions of 1 to 1½ hours each. In each session, there were two blocks of 440 trials. In 86% of the trials in each block, the subject was cued to attend to one of the 10 boxes; in the remaining 14% the subject was cued to attend to none of the boxes. A stimulus was actually presented following the cue on 82% of the trials. The remaining trials were catch trials. The stimulus was presented in the attended box on approximately 71% of the noncatch trials, and was presented in one of the other nine boxes on the remainder of the noncatch trials. Each of the 90 possible cue-stimulus mismatches occurred only once in each block of trials. Over the course of the experiment, however, each mismatched cue-stimulus pair occurred eight times for each subject.

We monitored eye movements of 3.75° or more on two of the four experimental sessions for each subject. Each of the 12 subjects was randomly assigned to one of the six possible combinations of two sessions with eye-movement equipment and two sessions without. Eye movements and anticipations were handled just as in Experiment 1: Error feedback was given and the trial was rerun at a randomly selected time later in the block. If the subject's reaction time was as fast as or faster than the mean reaction time for the previous 40 trials in which the stimulus had occurred in the attended box, the subject was presented with a flickering asterisk at the center cue location.

The sequence of events within a trial was identical to that in Experiment 1, with the exception that the cue interval ranged from 600 to 1000 msec on noncatch trials, and lasted a fixed 1000 msec beyond that on catch trials. Subjects were paid 50 cents for each block on which they had "attended closely," and were informed after each block whether they had earned the bonus. The criterion for earning a bonus, which was not revealed to the subjects, was that both the mean cost and the mean benefit for the block be 5 msec or greater.

Results and Discussion

We calculated the attentional costs for all trials in which the subject had been attending to one location and the stimulus had occurred at another, as well as the benefits for all trials in which the stimulus had occurred at the attended location.⁴

⁴As in Experiment 1, there were practically no effects of absolute stimulus position on the detection times for the neutral trials (the slowest mean detection time was only 14 msec slower than the fastest). This makes the additivity issue nearly moot in interpreting attentional effects in this experiment.

Before calculating costs and benefits, we removed all reaction times that were at least two standard deviations above or below each subject's mean reaction time for that particular cue and stimulus combination.

The mean costs and benefits are presented in Fig. 8.2, in which stimulus position is shown along the abscissa and cued position is the parameter distinguishing the curves. One way of interpreting the graph is to think of each curve as an approximation to the shape of the attentional gradient when it is centered on different retinal locations.

As in Experiment 1, attentional facilitation falls off (i.e., costs increase) with increasing distance from the cued location. However, the shape of this fall-off is not homogeneous across the retina. The gradient is most sharply peaked, with a steeper slope, when it is centered on a cued position near the fovea than when it is centered on one of the next three positions moving toward the periphery; however, it is again more sharply peaked at the most peripheral position on each side. The benefit for expected locations (i.e., the negative peak of the gradient) changes in a similar way, with greatest benefits at the most foveal and the most peripheral cued locations. The gradient becomes very steep as it passes over near-foveal locations, regardless of where it is centered. All the gradients flatten

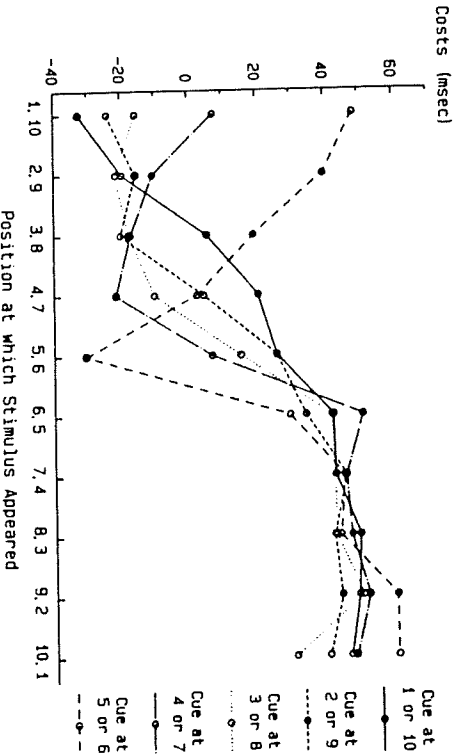


FIG. 8.2. Costs and benefits from Experiment 2. Stimulus position is shown along the abscissa and cued position is the parameter distinguishing the curves. Costs and benefits from cue-stimulus pairs in corresponding positions on either side of the midline were averaged before these data were plotted; thus, the abscissa does not represent stimulus position in left to right order but in an order from the most peripheral stimuli on the same side as the cued location (1 and 10) to the most peripheral stimuli on the side opposite to the cued location (10 and 1). The point on each curve corresponding to the benefit for the cued location is filled in (costs and benefits are plotted on the same scale).

out at a constant cost level at stimulus locations across the midline from the cued location.

We conducted two tests of the null hypothesis that the gradient is identical regardless of where it is centered: one for benefits, one for costs. Although there are benefits for stimuli at all cued locations, the amount varies significantly across different locations— $F(9, 99) = 3.60, p < .001$. To test the effect of retinal position on the overall costs and on the increase in costs with increasing cue-stimulus separation, we examined data from trials with separations of 12.5° or less; only within this range were the two factors of retinal separation and stimulus position orthogonal. A repeated-measures analysis of variance with factors for retinal separation (2.5°, 5°, 7.5°, 10°, and 12.5°), stimulus eccentricity (1.25°, 3.75°, 6.25°, 8.75°, or 11.25° from fixation), and stimulus field (left visual field or right visual field) showed that costs increased with increasing retinal separation— $F(4, 44) = 58.51, p < .0001$. This increase included a significant linear trend— $F(1, 11) = 99.39, p < .0001$ —and a significant quadratic trend— $F(1, 11) = 14.54, p < .005$ —but no significant higher-order trends. There was also a significant effect of stimulus eccentricity— $F(4, 44) = 2.65, p < .05$ —and a significant interaction between stimulus eccentricity and retinal separation— $F(16, 176) = 9.13, p < .0001$. Thus, contrary to the null-hypothesis assumption, costs for a given retinal separation appear to depend on the retinal eccentricity of the stimulus. Costs are greater on the whole for stimuli closer to the fovea, and this effect is found mainly with smaller retinal separations (2.5°, 5°, and 7.5°). The main effect of stimulus position and its interaction with retinal separation remain significant even when the linear and quadratic components of cued location are partialled out— $F(4, 44) = 7.96, p < .0001$ and $F(16, 176) = 2.67, p < .001$, respectively.

A large part of the interaction in Fig. 8.2 may be accounted for with a single assumption: that the shape of the gradient is invariant, but that the distance metric it is defined over is not invariant across the visual field. The gradient could be defined over visual-angle units scaled by receptive field size, or, equivalently, it could be defined over the cortical representation of the visual input, where a given range of visual angle maps onto a greater extent of cortex the closer it is to the fovea. This would result in attention falling off more rapidly for retinal regions where resolution is fine than for regions where it is coarse.⁵ This, in turn, could account both for the asymmetry in each of the gradients in Fig. 8.2 and for their increasing bluntness from the fovea to the periphery. It could also account for the flattening of the gradients as they cross the midline: In crossing the midline the gradient would pass through the densest region of the retina, and so the portion lying in the contralateral visual field would tend to be the

⁵Because all the boxes on the screen were clearly resolvable, this effect is not reducible to subjects' being unable to discriminate differences among stimulus positions in eccentric regions due to the physiological resolution of the retina.

asymptotic tail. In this simple model, there would be no factor specific to stimulus position on the retina per se, and no factor specific to crossing versus not crossing the midline.

To test this possibility, we ran several hierarchical regression analyses (see Cohen & Cohen, 1975) on the costs and benefits. A model with the linear and quadratic components of retinal separation (a crude approximation to the assumed shape of the moveable gradient), stimulus eccentricity, and the interactions of the linear and quadratic components of retinal separation with stimulus eccentricity accounted for 76.6% of the variance among the 100 means that went into Fig. 8.2. All of these regressors, except the interaction between stimulus position and the quadratic component of retinal separation, accounted for significantly more of the variance when included in the regression than when excluded, and accounted for enough additional variance to justify inclusion in the model according to the criteria suggested by Cohen and Cohen (1975). However, a simpler model, including *only* the linear and quadratic effects of retinal separation, but expressing retinal separation in terms of millimeters of cortex according to the magnification formula of Rovamo and Virsu (1979), accounts for more variance (86.4%) with three fewer free parameters. Adding the following regressors did not increase the amount of variance accounted for by an amount that would justify their inclusion in the model (see Cohen & Cohen, 1975): stimulus eccentricity; cue eccentricity; the interactions of these effects with the trend components of cortically magnified retinal separation; a variable representing whether the stimulus projected to the same hemisphere as the cued location or to the opposite hemisphere; or the interaction of this variable with scaled retinal separation. Thus, much of the interaction in Fig. 8.2 appears to be explicable in terms of the simple assumption that attention makes sharper distinctions for retinal regions with finer resolution.

In addition to effects of retinal resolution or cortical magnification, our data may show an endpoint anchoring effect. Benefits are greater and the gradient is steeper for cued locations at both ends of the series of possible stimulus locations and for locations adjacent to the fixation point. Hierarchical regression analyses testing the contribution of a variable representing the proximity of the cued location to these "landmarks" (and its interactions) revealed that this effect was significant and larger than that of the other additional regressors we had tested, although not large enough to justify inclusion in the model (see Cohen & Cohen, 1975).

It is not completely clear why we observed effects of stimulus position and Posner (1978) did not. Our data would lead one to predict that costs would be smaller for the retinal separation of 1° than for the retinal separation of 50° , even though the cued and stimulus positions were more central for the small separation. Perhaps a key difference is that in Posner's experiments, subjects in any one experiment had to detect stimuli at one of two positions, whereas in ours, 10 positions were possible. Posner's subjects may have narrowed their gradients to a

very thin peak at the cued location, such that noncued positions would all fall on the tail of the gradient. Another possibility stems from the fact that the possible stimulus locations in Posner's experiments were on either side of the midline. Costs in Experiment 2 appeared to asymptote for all positions across the midline from the attended locus (see Fig. 8.2). Although our analyses did not discern a contribution of this midline-crossing factor beyond the contribution of scaled retinal separation, this question deserves further investigation.

GENERAL DISCUSSION

The results reported in this chapter suggest that the mental representation underlying visual attention has as its dimensions a visual-angle scale (presumably, both horizontal and vertical), distorted by something similar to cortical magnification or change in receptive field size, and a scale representing distance in depth. Such a representation of space is similar to the $2\frac{1}{2}$ -D sketch or depth map proposed by researchers in computer vision as an intermediate stage of visual representation in the shape-recognition process (Marr, 1982; Marr and Nishihara, 1978). According to Marr, the $2\frac{1}{2}$ -D sketch is the first representation of the visual world from which cognitive processes can read information (see also Pinker, 1984); its dimensions are visual angle and depth, and its resolution is nonhomogeneous and proportional to the resolution of the corresponding areas of the retina. Though we would not claim that visual attention selects information directly from the $2\frac{1}{2}$ -D sketch, our results do suggest that the representation underlying attentional selection is organized similarly to the $2\frac{1}{2}$ -D sketch.

Our findings also bear out Ullman's (1984) conjecture that the "visual routine" corresponding to shifting the locus of visual processing should be applicable to specific three-dimensional regions of the visual world and not just to retinally defined regions.

In addition to suggesting the dimensions of the space represented in the areas in which visual attention exerts its effects, these data suggest what the shape of the gradient operating on these areas might be, on the assumption that attentional costs are linearly related to the amount of activation defined by the gradient.⁶ Specifically, attention falls off with retinal separation according to a gradient that decreases with a negative acceleration and that can appear sharply peaked (as in Experiment 1) or relatively blunt (as for the middle locations of Experiment 2). These differences in bluntness appear to depend on the eccentricity of the attended location, and possibly on its proximity to a perceptual landmark and on the number of potential targets as well. For small retinal distances from the

⁶Of course, it is difficult to discriminate experimentally between a gradient of simultaneous attention applied in parallel over the entire visual field and a probability distribution governing where a circumscribed attentional "peephole" is centered.

attended locus, the height of the central peak and its immediately surrounding region changes relatively little with increasing depth from the attended region. However, at retinal distances greater than 5°, the gradient falls off more steeply with increasing depth from the attended locus. Furthermore, the fall-off of attentional activation with separation in depth may be asymmetric along the depth dimension, being steeper on the far side of the plane of maximum activation than on its near side.

Although the gradient model we have proposed seems the simplest way to explain our data, other models could be devised by trading off properties of the representation subserving attention and the shape of the gradient defined over that representation, by positing the involvement of multiple representations instead of a single one, or by assuming a nonlinear relation between measured costs and the underlying gradient of attention. For example, one might account for the data in Fig. 8.1a by positing separate structures representing two-dimensional space and depth, each with a different gradient of attention, and a nonlinear combination function that determines the detectability of an event as a joint function of its activation in the two structures. However, our data do strongly suggest that visual attention is sensitive to depth, visual angle, and retinal or cortical resolution, and these are findings that any model of visual attention must account for.

ACKNOWLEDGMENTS

This research was supported by National Science Foundation grants BNS81-14916 and 82-16546 and NIH grant ROI HD 18381 to the second author and by the MIT Center for Cognitive Science under a grant from the Sloan Foundation, and was conducted during the first author's tenure as an NSF fellow. Portions were presented at the meeting of the Psychonomic Society in 1982. We wish to thank Larry Maloney, Misha Pavel, Molly Potter, Whitman Richards, Roger Shepard, Ewart Thomas, Barbara Tversky, Brian Wandell, and Jeremy Wolfe for helpful comments on earlier versions of this manuscript, and Roger Shepard for use of his laboratory facilities for Experiment 2.

REFERENCES

- Bashinski, H. S., & Bacharach, V. R. (1980). Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations. *Perception and Psychophysics*, 28, 241-280.
- Ciuffreda, K. J., & Kenyon, R. V. (1983). Interactions between accommodation and vergence. In C. M. Schor & K. J. Ciuffreda (Eds.), *Vergence eye movements* (pp. 99-174). Woburn, MA: Butterworth.
- Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cowey, A. (1982). Sensory and non-sensory visual disorders in man and monkey. *Philosophical Transactions of the Royal Society of London*, B298, 3-13.
- Eriksen, C. W., & Hoffman, J. E. (1972). Some characteristics of selective attention in visual perception determined by vocal reaction time. *Perception and Psychophysics*, 11, 169-171.
- Heilman, K. M. (1979). Neglect and related disorders. In K. M. Heilman & E. Valenstein (Eds.), *Clinical neuropsychology* (pp. 268-307). New York: Oxford University.
- Hoffman, J. E., & Nelson, B. (1981). Spatial selectivity in visual search. *Perception and Psychophysics*, 30, 283-290.
- Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.
- Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society*, 200, 269-294.
- Pinker, S. (1984). Visual cognition: An introduction. *Cognition*, 18, 1-63.
- Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3-25.
- Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109, 160-174.
- Remington, R. W. (1980). Attention and saccadic eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 726-744.
- Rovamo, J., & Virsu, V. (1979). An estimation and application of the human cortical magnification factor. *Experimental Brain Research*, 37, 495-510.
- Shaw, M. L. (1984). Division of attention among spatial locations: A fundamental difference between detection of letters and detection of luminance increments. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 109-120). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sperling, G. (1960). Information available in brief visual presentation. *Psychological Monographs*, 74(11, whole number 498).
- Sperling, G. (1984). A unified theory of attention and signal detection. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 103-181). New York: Academic.
- Sperling, G., & Reeves, A. (1980). Measuring the reaction time of a shift of visual attention. In R. W. Nickerson (Ed.), *Attention and performance VIII* (pp. 347-360). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ullman, S. (1984). Visual routines. *Cognition*, 18, 96-159.
- Wurtz, R. H., Goldberg, M. E., & Robinson, D. L. (1982). Brain mechanisms of visual attention. *Scientific American*, 246, 124-135.