

# Visual attention and the apprehension of spatial relations: The case of depth

CATHLEEN M. MOORE, CATHERINE L. ELSINGER, and ALEJANDRO LLERAS  
*Pennsylvania State University, University Park, Pennsylvania*

Several studies have shown that targets defined on the basis of the spatial relations between objects yield highly inefficient visual search performance (e.g., Logan, 1994; Palmer, 1994), suggesting that the apprehension of spatial relations may require the selective allocation of attention within the scene. In the present study, we tested the hypothesis that depth relations might be different in this regard and might support efficient visual search. This hypothesis was based, in part, on the fact that many perceptual organization processes that are believed to occur early and in parallel, such as figure-ground segregation and perceptual completion, seem to depend on the assignment of depth relations. Despite this, however, using increasingly salient cues to depth (Experiments 2–4) and including a separate test of the sufficiency of the most salient depth cue used (Experiment 5), no evidence was found to indicate that search for a target defined by depth relations is any different than search for a target defined by other types of spatial relations, with regard to efficiency of search. These findings are discussed within the context of the larger literature on early processing of three-dimensional characteristics of visual scenes.

Do we perceive where one object is relative to another if we have not specifically attended to those objects as a pair? Recent evidence suggests that we do not. In particular, data from visual search tasks suggest that the detection of targets defined by the spatial relations between objects may require the allocation of attention (Logan, 1994; O'Connell & Treisman, 1990; Palmer, 1994; Poder, 1999; Steinman, 1987). In the present paper, we are concerned with the question of whether this conclusion applies to all spatial relations, or whether some types of relations are privileged within the system so that they do not require the same level of attentional processing. In particular, we consider the possibility that depth relations may be extracted more efficiently than other spatial relations.

## **Evidence Concerning the Apprehension of Spatial Relations**

Using visual search performance as a diagnostic, Logan (1994) tested the hypothesis that the apprehension of spatial relations requires the allocation of attention. Targets were defined by the spatial relations between items within a pair. In one experiment, for example, the target was a-dash-to-the-left-of-a-plus pair (–+) among a-plus-to-the-left-of-a-dash distractor pairs (+–). In another experiment, the target was a-dash-above-a-plus pair

among a-plus-above-a-dash distractor pairs. In multiple experiments, Logan (1994) found that search for targets defined by spatial relations was highly inefficient. In particular, slopes relating search time to the number of items in the display were often greater than 30 msec per item for trials on which the target was present and greater than 60 msec per item for trials on which the target was absent. Search slopes of this magnitude suggest that the characteristic that defined the target from the distractors—in this case the spatial relation between items in a pair—was not extracted easily from the display. They are even consistent with the hypothesis that attention had to be allocated to each pair in turn in order to extract the spatial relations.

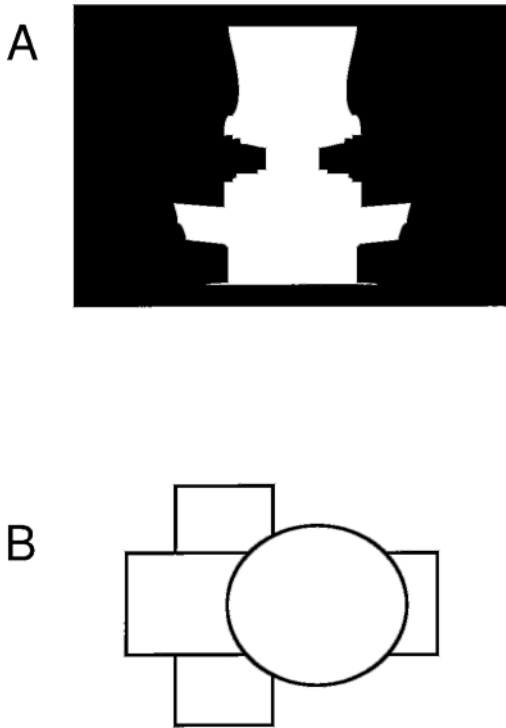
In an effort to further test the hypothesis that the apprehension of spatial relations requires attention, Logan (1994) sought converging evidence from a cuing paradigm. The task was again a visual search task in which subjects searched for a target that was defined by the spatial relations between items. In this case, however, one of the pairs of items was presented in a color different from that of the surrounding pairs. A validity manipulation was introduced so that on most of the trials, the target was the uniquely colored pair. Color then served as a cue to direct one's attention to that pair. If the detection of the target depended on the allocation of attention, this cue should have aided performance for those trials on which the target actually was the uniquely colored pair. Indeed, this was just what was found, thereby providing converging support for the conclusion that the apprehension of spatial relations involves the allocation of spatial attention.

## **Spatial Relations Defined by Depth**

As has been noted, Logan's (1994, 1995) studies addressed the question of the apprehension of left–right

---

Portions of this work were presented at the 1998 meeting of the Psychonomic Society in Dallas. We are grateful to Rich Carlson, Jim Enns, Gordon Logan, Toby Mordkoff, Ron Rensink, Kurt Schimmel, Rob Ward, and Jeremy Wolfe for helpful discussions and/or feedback on previous drafts of this paper. Support was provided by NSF Grant SBR-9728628. Correspondence concerning this article should be addressed to C. M. Moore, Department of Psychology, Penn State University, University Park, PA 16802 (e-mail: cmm15@psu.edu).



**Figure 1.** Examples of the apparent need for the assignment of depth relations in various perceptual organization processes. (A) Figure-ground segregation: Interpreting the image as a vase or two faces implies the assignment of which part, the white or the black, is in front of the other part. (B) Amodal completion: The completion of two rectangles behind the oval implies the assignment of relative depth positions—the oval in front of the horizontal rectangle in front of the vertical rectangle.

and up-down relations. These spatial relations were all defined between items that appeared on a single fronto-parallel plane, the computer monitor. Depth relations, in contrast, may be represented by the visual system differently than these other types of spatial relation, because so much of the information about the structure of a scene at least appears to depend on depth information. Figure-ground segregation, for example, can be conceptualized as the identification of which region is in front of the other region (see Figure 1A). Similarly, the parsing of a scene into objects when some objects occlude others appears to depend on the assignment of depth relations (see Figure 1B). To the extent that the visual scene is organized into perceptual objects—candidates for selection—it would seem that these objects would be established in an efficient spatially parallel manner for all parts of the scene simultaneously, rather than requiring the selective allocation of attention to different parts of the scene.

Consistent with these intuitions regarding the special status of depth information, Nakayama and Silverman (1986) found that targets defined by the conjunction of depth and color supported efficient visual search. Con-

junction targets vary in the efficiency of search that they support, depending on the saliency of the features making up the conjunction (Wolfe, 1994; Wolfe, Cave, & Franzel, 1989). That a conjunction involving depth yielded highly efficient search suggests that depth may indeed be something of a privileged feature. Nakayama and his colleagues (e.g., Nakayama, He, & Shimojo, 1995) have argued, in particular, that one of the earliest and most efficient processes within the stream of visual information processing is the representation of the surfaces that make up the scene. Attention, they argue, is then allocated separately within a surface, thus accounting for the efficient search that they observed for their depth  $\times$  color conjunction targets; subjects simply searched within a surface for a uniquely colored item. With regard to the present hypothesis, because surface information is usually dependent on depth information, it may be that depth relations are processed more efficiently than other spatial relations.

### Overview of the Present Study

Our purpose in the present experiments was to test the hypothesis that depth relations are different from other types of relations and that targets defined by depth relations will support efficient visual search. Experiment 1 provided a replication of Logan's (1994) basic up-down and left-right conditions, but with somewhat more salient stimuli. In Experiment 2, depth-relation targets were defined by using a single monocular depth cue (occlusion). This experiment yielded extremely inefficient searches. In Experiment 3, a second monocular depth cue (size) was added, but it failed to improve performance much. Finally, in Experiment 4, depth-relation targets were defined using binocular disparity, which is a very strong cue to depth (this was confirmed in Experiment 5). Still, search was highly inefficient.

Thus, despite the use of increasingly salient and compelling cues to depth, the present experiments provided no evidence to support the hypothesis that depth relations can support efficient search more effectively than other types of spatial relations. Some possible implications of these findings within the context of the larger literature on the preattentive processing of three-dimensional (3-D) characteristics of visual scenes are discussed in the General Discussion section.

## EXPERIMENT 1

Experiment 1 was designed to be a replication and small extension of Logan's (1994) original experiments. Logan (1994) used white pluses (+) and dashes (−) on a black background to form his spatial-relation pairs. It seemed possible that if the items were made more saliently different from each other, the apprehension of spatial relations between them might be easier and therefore allow for efficient search. With this in mind, we used items that differed from each other not only in form, but also in color (and, because the colors were not equiluminant, in lumi-

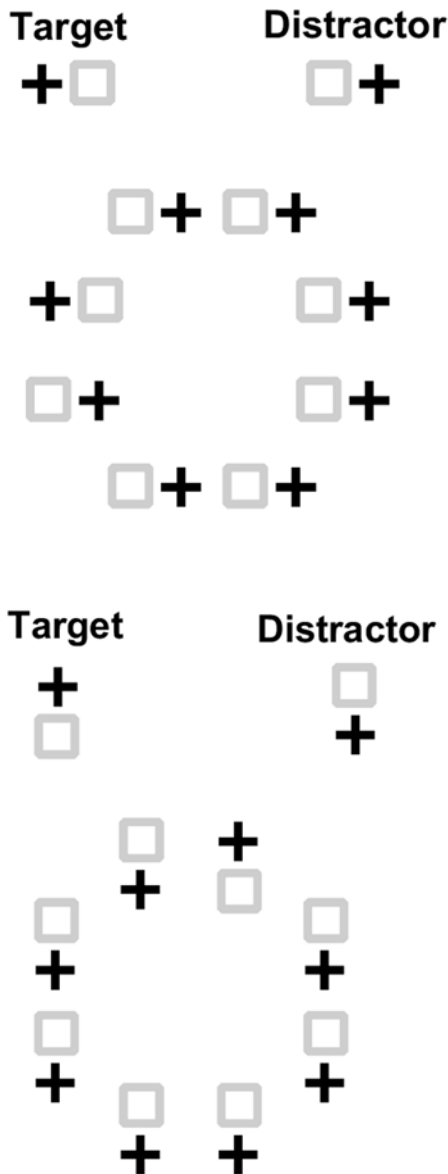


Figure 2. Illustrations of displays used in Experiments 1A (A) and 1B (B). (See text for details.)

nance as well). We used purple plusses and green squares to make up our spatial-relation pairs (see Figure 2).

## Method

**Subjects.** Thirty individuals from the Pennsylvania State University subject pool were tested—15 in Experiment 1A and 15 in Experiment 1B. All reported normal or corrected-to-normal visual acuity and color vision, and all were naive as to the purpose of the experiment before being tested. The subjects received extra credit in a psychology course for participating in the experiment.

**Equipment.** Stimuli were presented on a 21-in. Nanao FlexScan F2-21 color monitor controlled by a Number 9 Imagine Series 128 graphics board in a pentium-based computer and were viewed from a distance of approximately 70 cm. The room was dimly lit with indirect incandescent lighting (7.30 cd/m<sup>2</sup>).

**Stimuli.** Displays consisted of one to eight pairs of purple plusses (0.98° × 0.98°) and green squares (0.98° × 0.98°), which were drawn with lines of 0.11° width. In Experiment 1A, the circles and squares within a pair were separated by a 0.20° gap along the *x*-axis and were aligned along the *y*-axis (see Figure 2A). In Experiment 1B, they were separated by a 0.20° gap along the *y*-axis and were aligned along the *x*-axis (see Figure 2B). Each pair was centered on one of eight equidistant locations along an imaginary circle that had a diameter of 11.20°. The position in which a given pair appeared was determined randomly, with the exception that the target pair (defined in the Task section) appeared in each of the eight locations an equal number of times within a block of trials.

**Task.** In Experiment 1A, the task was to search for a pair of items in which the plus was to the left of a square among pairs in which the square was to the left of the plus. In Experiment 1B, the task was to search for a pair of items in which the plus was above the square among pairs in which the plus was below the square. The subjects responded on the “z” and “/” keys on the keyboard with their left and right forefingers, respectively. They indicated “target present” by pressing the key with their dominant hand and indicated “target absent” by pressing the key with their nondominant hand. The subjects were asked to make their responses as quickly as possible while maintaining an accuracy level of 95% or higher.

**Design.** A 2 (target: present, absent) × 4 (set size: 1, 2, 4, 8) within-subjects design was used. The first variable, target, referred to whether the target pair was present within the display or absent. The second variable, set size, referred to the total number of pairs within the display. Both variables were manipulated pseudo-randomly within blocks of trials. The subjects participated in eight blocks of 64 correct trials each (error trials were repeated at a random point later in the block). This resulted in 64 observations in each of the eight conditions.

**Procedure.** Each subject participated in a single 1-h session that began with a set of written instructions that described the task. After the instructions, the subjects completed a 64-trial practice block, followed by the eight blocks from which the data were collected.

Each trial began with the presentation of a fixation cross. After 500 msec, the search display appeared and remained present until a response was made. Following the response, the screen went blank and remained so for 1,500 msec, at which point the fixation cross for the next trial appeared. Trials on which the wrong button was pushed or responses were longer than 5,000 msec or shorter than 150 msec were flagged as errors. Each error was followed by a 100-msec 500-Hz tone. The mean reaction time (RT) and percent of trials correct for a given block were displayed on the monitor following each block. If the accuracy for a given block was lower than 95%, the following message accompanied the end-of-block feedback: “You are making too many errors. Please slow down and increase accuracy.”

## Results and Discussion

Figure 3 shows the mean correct RTs for the target-present and target-absent trials as a function of set size for Experiments 1A and 1B. Table 1 gives the slopes and *R*<sup>2</sup> values of the best-fitting lines (shown in Figure 3) for target-present and target-absent trials separately. Overall, the results provide a replication of Logan’s (1994) basic finding. Targets defined by the spatial relationship between objects yielded highly inefficient search performance, and making the items in the pairs more saliently different from each other did not allow the spatial-relation targets to be detected efficiently.

The RTs were submitted to separate 2 (target: present, absent) × 4 (set size: 1, 2, 4, 8) within-subjects analyses of variance (ANOVA) for the two experiments, with alpha

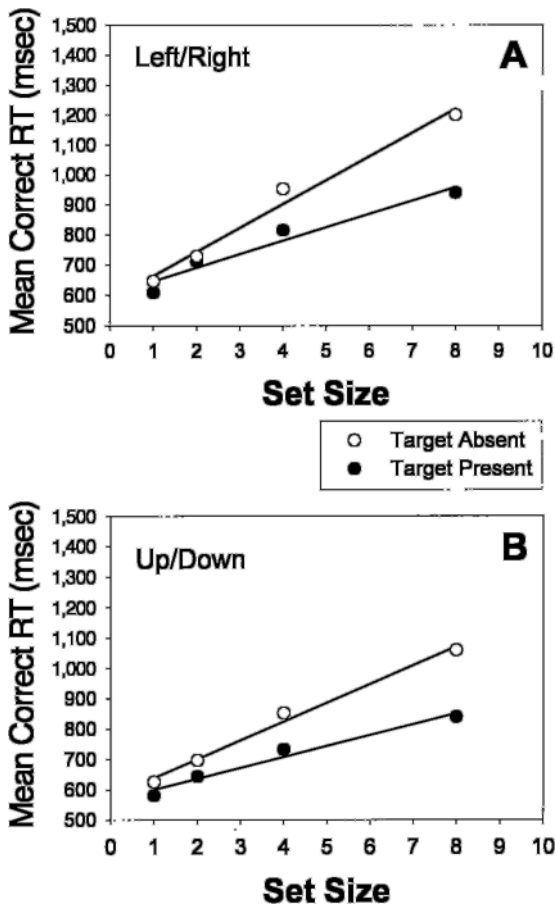


Figure 3. Mean correct RTs for Experiments 1A (A) and 1B (B). The best-fitting lines to the target-present and target-absent data are shown in the figures. Slopes and  $R^2$ s are given in Table 1.

set at .05. All sources of variance were significant in Experiment 1A [target,  $F(1,14) = 22.40$ ,  $MS_e = 17,210$ ,  $p < .001$ ; set size,  $F(3,42) = 43.14$ ,  $MS_e = 26,399$ ,  $p < .001$ ; target  $\times$  display size interaction,  $F(3,42) = 14.58$ ,  $MS_e = 6,385$ ,  $p < .001$ ]. All sources of variance were significant for Experiment 1B as well [target,  $F(1,14) = 37.74$ ,  $MS_e = 9,412$ ,  $p < .001$ ; set size,  $F(3,42) = 67.602$ ,  $MS_e = 10,248$ ,  $p < .001$ ; target  $\times$  display size interaction,  $F(3,42) = 26.35$ ,  $MS_e = 1,873$ ,  $p < .001$ ].

All of the same analyses were conducted on the arcsine transformations of the error rates (ERs, given in Table 2). For Experiment 1A, more errors were made on target-present trials than on target-absent trials [ $F(1,14) = 5.397$ ,  $MS_e = 0.0108$ ,  $p < .05$ ]. In addition, errors increased more steeply as a function of set size for target-present trials than for target-absent trials, as was confirmed by the significant interaction term [ $F(1,14) = 5.397$ ,  $MS_e = 0.0108$ ,  $p < .001$ ]. The same effects occurred in Experiment 1B: More errors were made on target-present trials than on target-absent trials [ $F(1,14) = 23.60$ ,  $MS_e = 0.0063$ ,  $p < .001$ ], and errors increased more steeply as a function of set size in the target-present condition than in the target-

absent condition [ $F(3,42) = 11.48$ ,  $MS_e = 0.0040$ ,  $p < .001$ ]. These effects are not surprising, because subjects rarely make false alarms in visual search tasks. Instead, the errors tend to be concentrated in the target-present trials as misses. No other effects that were in a different direction from those of the RTs were revealed, indicating that the pattern of RT results was not obscured by a speed-accuracy tradeoff.<sup>1</sup>

## EXPERIMENT 2

Experiment 2 provided our first test of the hypothesis that targets defined by a depth relation would allow for efficient visual search. Depth was defined by a single monocular cue—occlusion. Targets were plus-in-front-of-square pairs and distractors were square-in-front-of-plus pairs (see Figure 4). At the level of featural analysis, targets differed from distractors only in terms of the color of the intersecting regions.

### Method

**Subjects.** Fifteen individuals from the Pennsylvania State University subject pool were tested. None had participated in Experiment 1. All reported normal or corrected-to-normal visual acuity and color vision, and all were naive as to the purpose of the experiment before being tested. The subjects again received extra credit in a psychology course for participating in the experiment.

**Equipment.** The equipment was the same as that in Experiment 1.

**Stimuli.** Displays again consisted of one to eight pairs of green pluses ( $1.7^\circ \times 1.7^\circ$  diameter) and purple squares ( $1.38^\circ \times 1.38^\circ$ ). In this case, however, the two items overlapped each other so that the plus occluded part of the square or the square occluded part of the plus (see Figure 4). To provide an increased area of overlap, the lines were thicker than those used in Experiment 1 ( $0.22^\circ$ ). This resulted in an area of overlap, regardless of which item was in front, of four square regions, each  $0.22^\circ \times 0.22^\circ$  in size. The ends of the plus sign extended beyond the sides of the square by  $0.32^\circ$ . Each pair was centered on one of eight equidistant locations along an imaginary circle that had a diameter of  $12.30^\circ$ . Which position a given pair appeared in was determined randomly, with the exception that the target pair (defined in the Task section) appeared in each of the eight locations an equal number of times within a block of trials.

**Task.** The task was to search for a pair of items in which the plus sign was in front of the square among pairs in which the square was in front of the plus sign. The subjects again responded on the “z” and “/” keys on the keyboard, indicating “target present” with their dominant hand and “target absent” with their nondominant hand. As before, they were asked to make their responses as quickly as possible while maintaining an accuracy level of 95% or higher.

Table 1  
Slopes (in Milliseconds per Item) and  $R^2$  for Experiments 1–5

Experiment	Target Present		Target Absent	
	Slope	$R^2$	Slope	$R^2$
1A	44	.944	79	.981
1B	36	.970	62	.990
2	130	.994	234	.999
3	36	.998	102	.999
4	61	.991	148	.998
5A	18	.999	60	.977
5B	7	.835	9	.867

**Table 2**  
Percentage Error Rates for Experiments 1–5

Experiment	Set Size (Target Present)				Set Size (Target Absent)			
	1	2	4	8	1	2	4	8
1A	2.29	2.86	3.40	7.55	2.78	2.70	2.07	1.69
1B	2.35	2.69	2.58	6.19	2.86	1.99	0.98	1.29
2	1.99	4.96	7.88	8.93	1.39	1.46	1.25	1.38
3	4.59	2.40	3.08	4.76	3.61	4.08	1.81	0.87
4	–	3.48	4.46	8.40	–	2.08	1.52	0.93
5A	–	1.85	3.17	2.90	–	2.21	0.92	1.06
5B	–	1.87	2.50	3.18	–	3.34	1.71	0.92

**Design and Procedure.** The design and procedure were the same as in Experiment 1.

**Results and Discussion**

Depth relations defined by the single monocular cue of occlusion clearly failed to support efficient visual search. Figure 5 shows the mean correct RTs from Experiment 2 for the target-present and target-absent trials as a function of set size. Table 1 gives the slopes and *R*<sup>2</sup>s of the best-fitting lines for target-present and target-absent trials separately.

The RTs were submitted to a 2 (target: present, absent) × 4 (set size: 1, 2, 4, 8) within-subjects ANOVA. All sources of variance were significant [target, *F*(1,14) = 197.41, *MS*<sub>e</sub> = 12,153, *p* < .001; set size, *F*(3,42) = 822.82, *MS*<sub>e</sub> = 11,614, *p* < .001; target × display size interaction, *F*(3,42) = 212.18, *MS*<sub>e</sub> = 3,781, *p* < .001].

All of the same analyses were conducted on the arc-sine transformations of the ERs (given in Table 2). As in Experiment 1, more errors were made on target-present trials than on target-absent trials [*F*(1,14) = 60.57, *MS*<sub>e</sub> = 0.0081, *p* < .001], and errors increased more steeply as a function of set size for target-present trials than for target-absent trials [*F*(3,42) = 10.86, *MS*<sub>e</sub> = 0.0058, *p* < .001]. No other effects that were in a different direction from those of the RTs were revealed.

Post hoc analyses on the slopes across Experiments 1 and 2 confirmed that search performance was even less efficient in Experiment 2 than in Experiment 1, in which left–right and up–down relations were tested. In fact, the slopes were so steep in Experiment 2 that they are consistent with subjects’ having had to move their eyes and fixate individual pairs of items in order to classify them as targets or distractors.

**EXPERIMENT 3**

So far there is no evidence to support the hypothesis that depth relations are special and can be apprehended more efficiently than other types of spatial relations. In Experiment 2, however, occlusion was the only cue to depth. It is possible that this was not a sufficiently salient depth cue to support efficient search. In Experiment 3, occlusion and size were used together to cue the depth relations. As in Experiment 2, the target was a-plus-in-front-of-a-square pair, and distractors were a-square-in-

front-of-a-plus pairs. In this case, however, when an item was in front, it was larger than when it was in the back. This monocular cue simulates the change in image size on the retina as an object moves farther away from the observer. The predictions were the same as they were for Experiment 2.

**Method**

**Subjects.** Sixteen individuals from the Pennsylvania State University subject pool were tested. None had participated in Experiments 1 or 2. All reported normal or corrected-to-normal visual acuity and color vision, and all were naive as to the purpose of the experiment before being tested. The subjects received extra credit in a psychology course for their participation in the experiment.

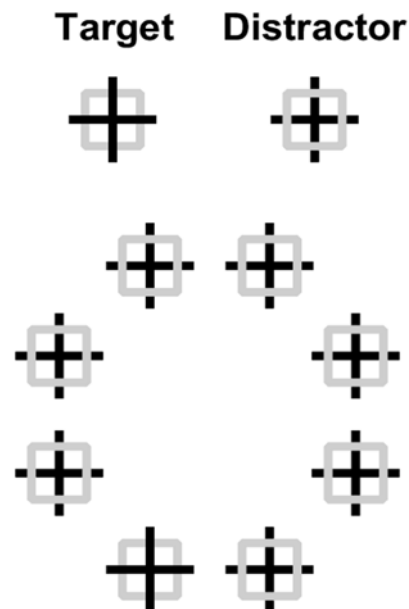
**Equipment.** The equipment was the same as that in Experiment 1.

**Stimuli.** The stimuli were the same as in Experiment 2, except that two different size plus signs (1.97° × 1.97° and 2.18° × 2.18°) and two different size squares (1.56° × 1.56° and 1.87° × 1.87°) were used (see Figure 6). The smaller version was used when it was the back item in a pair. The larger version was used when it was the front item in a pair.

**Task, Design, and Procedure.** The task, design, and procedure were all identical to those in Experiment 2.

**Results and Discussion**

Adding the depth cue of size did improve search efficiency in comparison with occlusion alone, but not to the levels that would be considered suggestive of preattentive analysis of depth relations. Figure 7 shows the mean correct RTs for the target-present and target-absent trials as a function of set size for Experiment 3. Table 1 gives the slopes and *R*<sup>2</sup>s of the best-fitting lines for target-present and target-absent trials separately.



**Figure 4.** Illustrations of displays used in Experiment 2. (See text for details.)

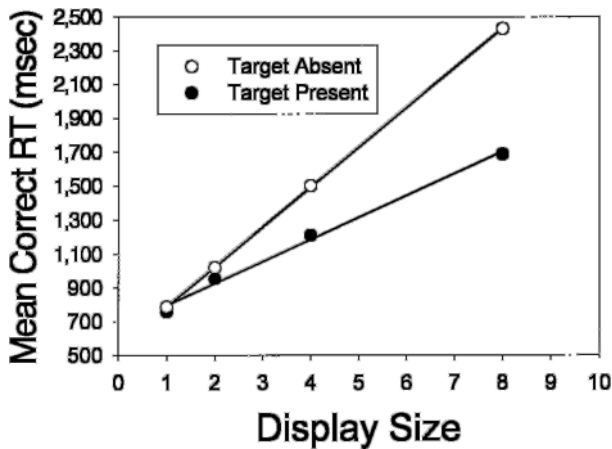


Figure 5. Mean correct RTs for Experiment 2. The best-fitting lines to the target-present and target-absent data. Slopes and  $R^2$ s are given in Table 1. Note increased RT range.

The RTs were submitted to a 2 (target: present, absent)  $\times$  4 (set size: 1, 2, 4, 8) within-subjects ANOVA. All sources of variance were significant [target,  $F(1,15) = 63.07$ ,  $MS_e = 25,352$ ,  $p < .001$ ; set size,  $F(3,45) = 61.57$ ,  $MS_e = 23,965$ ,  $p < .001$ ; target  $\times$  display size interaction,  $F(3,45) = 29.00$ ,  $MS_e = 11,518$ ,  $p < .001$ ].

All of the same analyses were conducted on the arcsine transformations of the ERs (given in Table 2). As before, more errors were made on target-present trials than on target-absent trials [ $F(1,15) = 10.92$ ,  $MS_e = 0.0055$ ,  $p < .01$ ], and errors increased as a function of set size more steeply for target-present trials than for target-absent trials [ $F(3,45) = 7.67$ ,  $MS_e = 0.0081$ ,  $p < .001$ ]. No other effects that were in a different direction from those of the RTs were revealed, indicating that the RT results were not obscured by a speed-accuracy tradeoff.

Although post hoc analyses across Experiments 2 and 3 confirmed that adding a size cue aided performance in comparison with having just the occlusion cue, similar comparisons across Experiments 1 and 3 revealed that, if anything, search in Experiment 3 was slightly less efficient than the right-left and up-down versions reported in Experiment 1.

Regardless, the pattern in Experiment 3 is certainly not one of efficient visual search. Thus, so far, there was still no evidence to support the hypothesis that depth relations are different from other types of spatial relations with regard to their ability to support efficient visual search. Experiment 4 provided a final attempt at seeking efficient search performance for a target defined by depth relations.

#### EXPERIMENT 4

There is still no evidence, then, to support the hypothesis that depth can be apprehended more efficiently than other types of spatial relations. In a third and final attempt to find evidence for such, in Experiment 4, depth was defined by binocular disparity, which provides a

very salient and phenomenally compelling cue to depth (e.g., Julesz, 1971; Wheatstone, 1838). Moreover, binocular disparity is the cue that Nakayama and Silverman (1986) used in their study, in which they found efficient search for a depth  $\times$  color conjunction target. Therefore, if any depth cue is going to allow for efficient visual search, then targets defined by binocular disparity ought to do so.

Because we were interested in whether search for depth relations, in particular, can be efficient, we presented the pairs of items randomly across three depth planes instead of just two. If only two depth planes were used (e.g., presenting all of the front items in one plane and all of the back items in another plane), the subjects might adopt the strategy of limiting their search to a single depth plane and look for an odd-item-out. For example, the subjects might search within the front depth plane for a purple plus among green squares. Given Nakayama and Silverman's (1986) results, this situation should certainly yield efficient search (which we confirmed in Experiment 5). Such a finding, however, would not indicate anything about search for a particular spatial relationship between items within a pair. To avoid this situation, displays were created with three depth planes. Any given pair appeared either with one item in the front depth plane and one in the middle depth planes, or alternatively, with one item in the middle depth plane and one in the back. Which of these two alternatives applied to a given pair was determined randomly. These displays discouraged the strategy of simply searching within a given depth plane.

#### Method

**Subjects.** Fifteen individuals from the Pennsylvania State University subject pool were tested. None had participated in any of

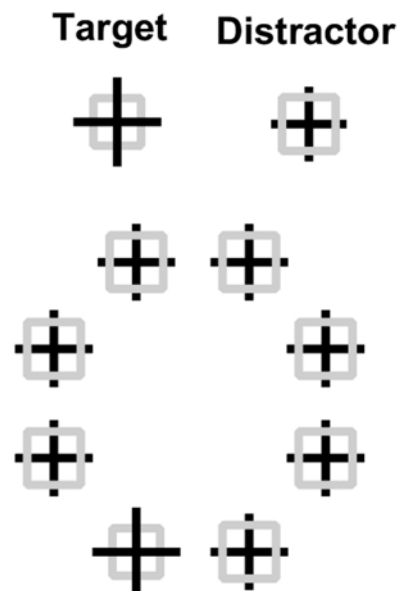


Figure 6. Illustrations of displays used in Experiment 3. (See text for details.)

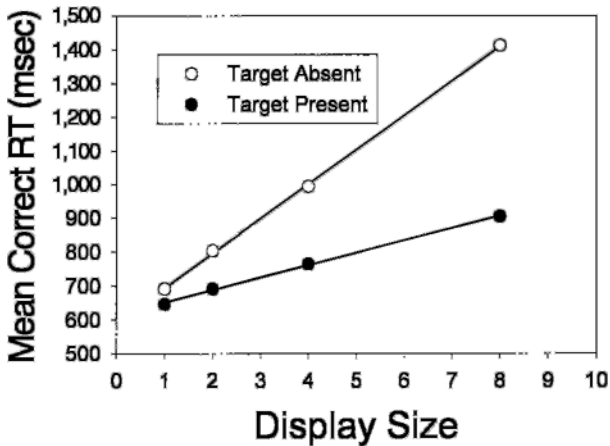


Figure 7. Mean correct RTs for Experiment 3. The best-fitting lines to the target-present and target-absent data. Slopes and  $R^2$ s are given in Table 1.

the preceding experiments. All reported normal or corrected-to-normal visual acuity and color vision, and all were naive as to the purpose of the experiment before being tested. The subjects received extra credit in a psychology course for participating in the experiment.

**Equipment.** Stimuli were presented on a different 21-in. Nanao FlexScan F2-21 color monitor that was controlled by a Number 9 Verge 3D PCI graphics board in a pentium-based computer. Viewing distance was approximately 70 cm. The room was dimly lit with indirect incandescent lighting. Binocular disparity was introduced using a Stereographics CrystalEyes II stereo display system. This system presented separate left- and right-eye images using a set of liquid crystal shutter glasses that were synchronized with the monitor. The frame rate was 60 Hz per eye.

**Stimuli.** The stimuli were the same as in Experiment 3, except that the perception of depth was induced through binocular disparity. In particular, the pairs of items were distributed across three different depth planes—one at the horopter (0 disparity), one in back of the horopter (a virtual distance of 2.5 cm), and one in front of the horopter (a virtual distance of 2.2 cm). The items within a pair were presented at two different depth planes so that one was at the horopter and one was in either the front or the back depth plane. This resulted in a given pair's never being separated by more than one depth plane. Which pair of depth planes a given pair appeared in was determined randomly for each pair of items. The resulting displays consisted of pairs of items—mostly distractors (squares in front of plus signs)—distributed sporadically across the three depth planes (see Figure 8).

**Task.** The task was the same as in Experiment 3.

**Design.** A 2 (target: present, absent)  $\times$  3 (set size: 2, 4, 8) within-subjects design was used. This was the same as in Experiments 1–3, except that three instead of four set sizes were used. The subjects completed eight blocks of 48 trials each, resulting in 64 observations in each of the six conditions.

**Procedure.** The procedure for this experiment was the same as that in Experiment 3, except that the instructions were given orally, rather than appearing in writing on the computer screen. In addition, all subjects were asked during practice whether they perceived the depth in the displays. All reported that they did.

**Results and Discussion**

Even when using binocular disparity as a cue to depth, no evidence was found to suggest that depth relations

allow for efficient visual search. Figure 9 shows the mean correct RTs for the target-present and target-absent trials as a function of set size for Experiment 4. Table 1 gives the slopes and  $R^2$ s of the best-fitting lines for target-present and target-absent trials separately.

The RTs were submitted to a 2 (target, present, absent)  $\times$  3 (set size, 2, 4, 8) within-subjects ANOVA. All sources of variance were significant [target,  $F(1,14) = 74.21$ ,  $MS_e = 26,068$ ,  $p < .001$ ; set size,  $F(2,28) = 215.45$ ,  $MS_e = 56,686$ ,  $p < .001$ ; target  $\times$  display size interaction,  $F(2,28) = 59.47$ ,  $MS_e = 9,615$ ,  $p < .001$ ].

All of the same analyses were conducted on the arcsine transformations of the ERs (given in Table 2). Again, more errors were made on target-present trials than on target-absent trials [ $F(1,14) = 47.08$ ,  $MS_e = 0.0053$ ,  $p < .01$ ], and errors increased more steeply as a function of set size for target-present trials than for target-absent trials [ $F(2,28) = 11.98$ ,  $MS_e = 0.0048$ ,  $p < .001$ ]. No other effects that were in a different direction from those of the RTs were revealed.

Given the salience of binocular disparity as a depth cue, there was mounting evidence that depth relations really are incapable of supporting efficient visual search.

**EXPERIMENT 5**

In the final experiment, we assessed whether the binocular disparity that was used in Experiment 4 was suffi-

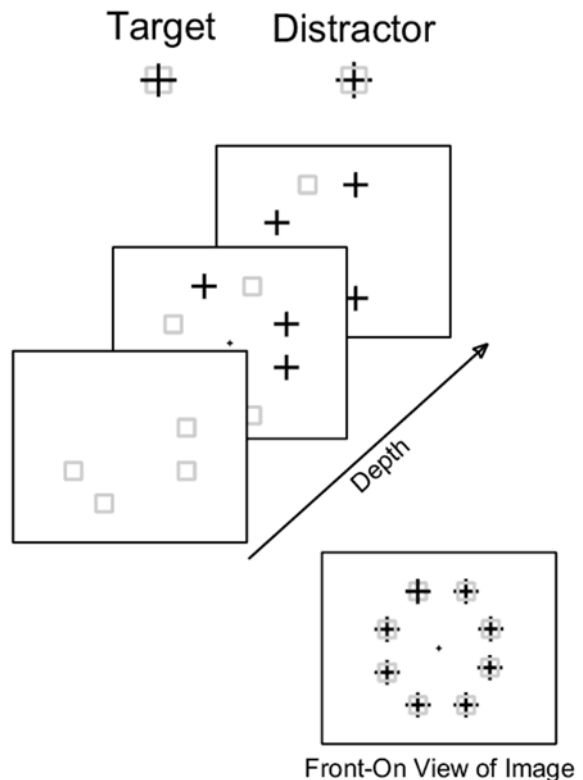


Figure 8. Illustrations of displays used in Experiment 4. (See text for details.)

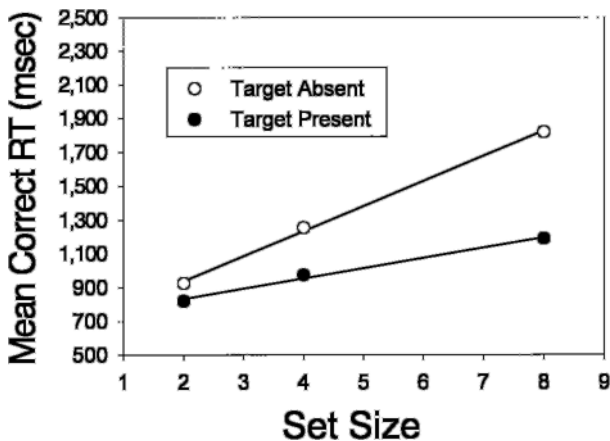


Figure 9. Mean correct RTs for Experiment 4. The best-fitting lines to the target-present and target-absent data. Slopes and  $R^2$ s are given in Table 1. Note the increased RT range.

ciently salient to support results analogous to those of Nakayama and Silverman (1986). This assessment was important, because, as was noted in the introduction to Experiment 4, their experiment suggested the possibility that depth relations might be special; if the binocular disparity used here could not support their findings, it would still be possible that more salient depth relations could support efficient search.

Experiment 5 was similar to Experiment 4, except that instead of shuffling the pairs of items across three different depth planes, they were presented at two depth planes. This configuration allowed for the strategy of limiting attention to one surface defined at a given depth plane and searching for the odd item out within that surface. So, for example, the task could be recast as being a search for a near plus, rather than for a particular depth relation. The logic was that, if the failure to obtain efficient search performance in Experiment 4 was due to the depth's being insufficiently salient, then we should still obtain inefficient search performance in Experiment 5. If, however, the depth was sufficiently strong, and the inefficient search in Experiment 4 was due to something else, then the subjects should be able to limit their search to the front (or back) surface and conduct an efficient search.

We report two versions of Experiment 5 (A and B), because the initial version yielded intermediate results. In particular, although search slopes were substantially reduced in Experiment 5A in comparison with those in Experiment 4 and the earlier experiments, they were not reduced by as much as we expected them to be, given the very efficient searches observed by Nakayama and Silverman (1986). In Experiment 5B, we separated the pairs of items (squares and pluses) so that they were no longer occluding each other, forming tightly grouped pairs. This made the search displays more similar to those of Nakayama and Silverman, in which there was no occlusion among search items. Search was indeed much more

efficient in the second version. Some possible reasons for the difference in performance across Experiments 5A and 5B are discussed below.

## Method

**Subjects.** Thirty individuals from the Pennsylvania State University subject-pool were tested—15 in Experiment 5A and 15 in Experiment 5B.<sup>2</sup> All reported normal or corrected-to-normal visual acuity and color vision, and all were naive as to the purpose of the experiment before being tested. The subjects again received extra credit in a psychology course for participating in the experiment.

**Equipment.** The equipment was the same as in Experiment 4.

**Stimuli.** The stimuli were the same as in Experiment 4, except that the front item in each pair appeared in a plane in front of the horopter at a virtual distance of 2.2 cm, and the back item in each pair appeared at the horopter. In Experiment 5B, the stimuli making up the pairs were also separated within the depth plane by a  $0.57^\circ$  (edge to edge) so that no item occluded any other item. Which item was to the left and which was to the right was randomly determined for each pair.

**Task, Design, and Procedure.** The task, design, and procedure were the same as in Experiment 4.

## Results and Discussion

Binocular disparity was a sufficiently salient cue to depth for our purposes. The results provide a conceptual replication of Nakayama and Silverman's (1986) finding of efficient search performance for a depth  $\times$  color conjunction target. Even though it was not their explicit task, the subjects presumably searched for and found the odd item out in the front (or back) depth plane, thus accounting for the relatively efficient search performance. Figures 10A and 10B show the mean correct RTs for the target-present and target-absent trials as a function of set size for Experiments 5A and 5B, respectively. Table 1 gives the slopes and  $R^2$ s of the best-fitting lines for target-present and target-absent trials, separately.

The RTs for Experiments 5A and 5B were submitted to separate 2 (target: present, absent)  $\times$  3 (set size: 2, 4, 8) within-subjects ANOVAs. For Experiment 5A, all sources of variance were significant [target,  $F(1,14) = 17.6301$ ,  $MS_e = 86,917$ ,  $p < .001$ ; set size,  $F(2,28) = 18.315$ ,  $MS_e = 23,290$ ,  $p < .001$ ; target  $\times$  set size interaction,  $F(2,28) = 25.20$ ,  $MS_e = 5,105$ ,  $p < .001$ ]. For Experiment 5B, both main effects were significant [target,  $F(1,14) = 10.01$ ,  $MS_e = 20,763$ ,  $p < .01$ ; set size,  $F(2,28) = 6.55$ ,  $MS_e = 2,865$ ,  $p < .01$ ]. The interaction, however, was not significant [ $F(2,28) = 2.78$ ,  $MS_e = 1,428$ , n.s.].

All of the same analyses were conducted on the arcsine transformations of the ERs (given in Table 2). There were no reliable effects that were in a different direction from those of the RTs.

In order to compare performance across Experiments 4, 5A, and 5B, a follow-up 2 (target: present, absent)  $\times$  3 (experiment: 4, 5A, 5B) mixed-design ANOVA was conducted on RT  $\times$  set size slopes, with experiment as the between-subjects variable. All sources of variance were significant [target,  $F(1,42) = 82.7$ ,  $MS_e = 642$ ,  $p < .001$ ; experiment,  $F(2,42) = 31.8$ ,  $MS_e = 2,520$ ,  $p < .001$ ; target  $\times$  experiment interaction,  $F(2,42) = 18.3$ ,  $MS_e = 642$ ,



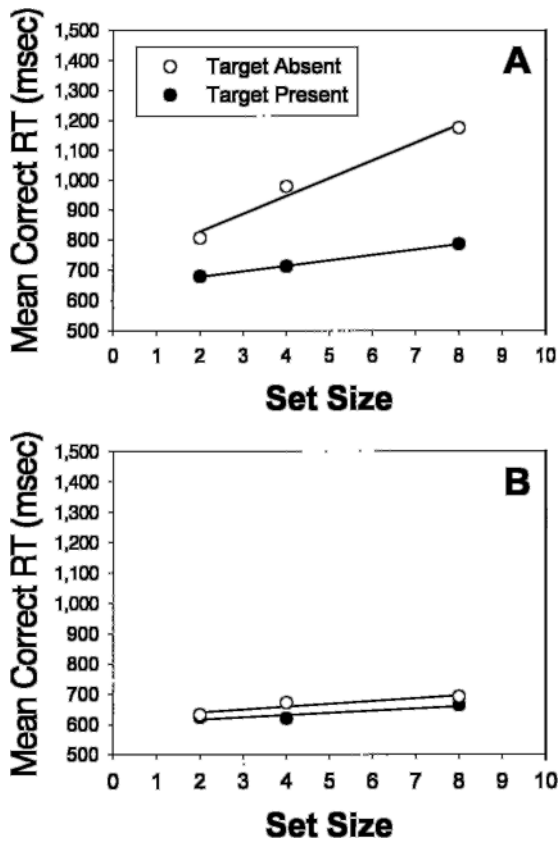


Figure 10. Mean correct RTs for Experiments 5A (A) and 5B (B). The best-fitting lines to the target-present and target-absent data. Slopes and  $R^2$ s are given in Table 1.

$p < .001$ ]. Another follow-up 2 (target: present, absent)  $\times$  2 (experiment: 4, 5A) mixed-design ANOVA comparing Experiment 5A with Experiment 4 confirmed that Experiment 5 yielded less steep slopes and a smaller difference between target-present and target-absent slopes than did Experiment 4 [target,  $F(1,28) = 80.7$ ,  $MS_e = 879$ ,  $p < .001$ ; experiment,  $F(1,28) = 18.4$ ,  $MS_e = 3,680$ ,  $p < .001$ ; target  $\times$  experiment interaction,  $F(1,28) = 5.91$ ,  $MS_e = 487$ ,  $p < .05$ ]. A final follow-up 2 (target: present, absent)  $\times$  2 (experiment: 5A, 5B) mixed-design ANOVA confirmed that Experiment 5B yielded still less steep slopes and a reduced difference between target-present and target-absent slopes than did Experiment 5A [target,  $F(1,28) = 26.30$ ,  $MS_e = 487$ ,  $p < .001$ ; experiment,  $F(1,28) = 9.79$ ,  $MS_e = 1,810$ ,  $p < .004$ ; target  $\times$  experiment interaction,  $F(1,28) = 13.5$ ,  $MS_e = 487$ ,  $p < .001$ ].

A question arises at this point as to why the subjects apparently failed to adopt the strategy that they used here in the previous experiments. In all of the previous experiments, the subjects could, in principle, have limited their attention to the front (or back) surface and looked for the odd item out. A possible explanation is suggested by contrasting Experiments 5A and 5B, which differed in the proximity and alignment of items within pairs relative to

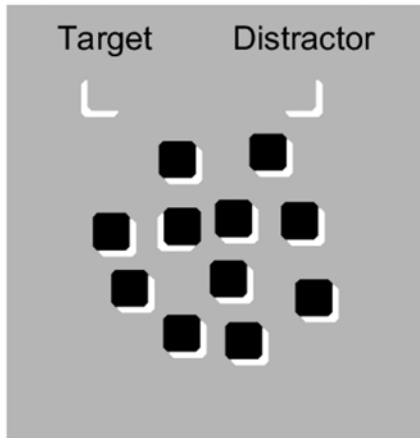
that between pairs. In particular, the extreme proximity and alignment of items within a pair (i.e., "right on top of each other") and the relatively large separation between pairs may have forced the subjects in Experiments 2, 3, and 5A to perceive items within a pair as strongly grouped with each other, rather than with other items on the same surface in depth. In other words, strong pairwise grouping of this sort could have prevented a search-within-a-surface strategy by overriding the alternative surface-based grouping. This situation may have been particularly likely in Experiments 2 and 3, in which depth was depicted only through monocular cues that were probably less effective in supporting strongly separated surface representations than was binocular disparity.

In addition to preventing the more efficient search-within-a-surface strategy, strong pairwise grouping within these search displays may have had another consequence. Results reported by Wolfe and Bennett (1997) suggest that access to a within-group or within-object structure requires the allocation of attention. If the pairs of items were forced into a grouped representation, accessing the spatial relationship between items (group structure) may have required the allocation of attention, thereby leading to the observed inefficient search performance. Such an account is consistent with Logan and colleagues' (Logan, 1995; Logan & Compton, 1996; Logan & Sadler, 1996) explanation for why the apprehension of spatial relations, in general, requires the allocation of attention.

## GENERAL DISCUSSION

Despite the use of what should have been sufficiently salient cues to depth in the present experiments, no evidence was found to support the hypothesis that depth relations are available for the early guidance of efficient visual search. Search time for targets defined by depth relations consistently increased with the increasing numbers of distractor pairs in the displays. How do these findings fit within the context of previous results concerning the preattentive processing of 3-D characteristics of the visual scene?

As described earlier, Nakayama and Silverman (1986) provided results that suggested that search can be limited to one of several surfaces defined in depth by binocular disparity, thereby leading to efficient search for a conjunction target defined by depth and color (see also Nakayama & He, 1995). That this occurred suggests that the representations of the surfaces in those 3-D scenes were established early, perhaps without recourse to the selective allocation of attention within the scene. That these surfaces were defined on the basis of depth in turn suggests that depth information was represented in a similar early, efficient, and possibly preattentive way. What this study does not speak to, however, is whether *relative* depth information was represented in this early and efficiently. As we have noted, the subjects in Nakayama and Silverman's task could simply segregate the scene into distinct surfaces and then limit their search to one of the surfaces. They



**Figure 11.** Illustration of displays used by He and Nakayama (1992). When the black squares were presented through binocular disparity in front of the white Ls, the Ls were perceptually completed into squares, and search was highly inefficient. In contrast, when the Ls were presented in front, they were not completed, and search was efficient.

therefore need not have had any representation of which surface was in front of which other surface.

Another study, however, suggests that even relative depth information may be represented preattentively. Using binocular disparity cues to depth, He and Nakayama (1992) showed that an L-shaped target was very difficult to find among mirror L distractors when a square that abutted the L appeared to be in front of it. In contrast, when the square appeared behind the L, the L popped out of the mirror L distractors (see Figure 11). Presumably this was because when the square appeared in front, the L-shaped regions were perceptually completed to form squares, and the target and distractor stimuli differed only in the relative placement of the front square on top of the black square. This is a left–right/up–down spatial relation, which Logan (1994) has shown to yield inefficient search. In contrast, when the squares appeared behind the Ls, the Ls were not completed, and the L to mirror L contrast could yield a fairly efficient search. The fact that the L-shaped regions were perceptually completed (i.e., extended beyond their explicit contours) when the black square was in front suggests that the two regions were assigned relative depths. In particular, it suggests that the black regions were, at some point in processing, tagged as *occluding* (i.e., in front of), and the white regions were, at some point in processing, tagged as *occluded* (i.e., behind). That the L-shaped regions were not perceptually completed when the black squares were behind similarly suggests that the two regions were assigned relative depths. In this case, the white regions were, at some point in processing, tagged as *occluding* and the black regions as *occluded*. Finally, that the latter condition yielded efficient visual search and the former yielded inefficient search further suggests that the relative depth

assignments occurred sufficiently early to support efficient search.<sup>3</sup>

Even if depth relations can be assigned early, however, this would not require that representations of those relations be accessible to search processes. One reason for this is that later representations of the scene can override earlier representations of the scene, and visual search often appears to have access only to the later representations (e.g., He & Nakayama, 1992; Rensink & Enns, 1995, 1998). In fact, a major point of He and Nakayama's study was to demonstrate this, not with regard to depth relations, but with regard to the L shapes in their displays. The L shapes in the black-squares-behind condition were clearly sufficient to support efficient visual search. The same L shapes in the black-squares-in-front condition, however, were insufficient to support efficient visual search. Presumably this is because in the latter condition, the representations of the white regions as completed white squares hindered access to the uncompleted L shape that was present in the image. Rensink and Enns (1995) argue that visual search processes can access early representations of the scene (e.g., contours, hues, etc.) only if those representations have not been “preempted” by later interpretations of the scene. Thus, they refer to the *pre-emption* of earlier representations by later representations with regard to access to visual search processes.

In light of this analysis, it is possible that depth relations were extracted early, but that those representations were preempted by later representations of the display, thereby rendering them inaccessible to the visual search system. As we have considered in the Discussion section following Experiment 5 of the present study, for example, it is possible that strongly grouped representations of pairs of items preempted representations of individual items, thereby rendering the structure within the pairs inaccessible to early search processes (Wolfe & Bennett, 1997).

Another possibility regarding the representations of depth relations in the present study, as well as in He and Nakayama's (1992) study, is that they never were established early and, for this reason, were unavailable to support efficient search. Rensink and Enns (1998) reported a study similar to that of He and Nakayama. In this case, monocular cues to depth were used so that instead of manipulating which region was occluded and which was occluding by manipulating the depth of the regions via binocular disparity, they manipulated whether the two regions abutted or not. In the abutted case, the interpretation of a completed figure behind an occluding figure was supported, whereas in the nonabutted case no completion was supported. The results from their first experiment were analogous to those of He and Nakayama's. Despite finding evidence of early completion, however, Rensink and Enns (1998) found no evidence that “surface stuff” or boundaries were represented behind the occluding regions. Thus, it is possible that unlike our interpretation of He and Nakayama's results, the early completion of occluded figures does not depend on the assignment of

relative depths, but rather on image-based characteristics of the display, such as T junctions (for a related discussion, see Enns, 1992). Indeed, under this interpretation, rather than representations of depth relations giving rise to perceptual completion, it is possible that perceptual completion gives rise to depth relations. This interpretation seems more difficult to apply to the He and Nakayama displays, because the images of the two different versions of the displays—one in which completion occurred and one in which completion did not occur—were nearly identical. On the other hand, comparison between the two studies is complicated by the fact that overall response time (i.e., the *y*-intercepts of the search functions) was much greater in the He and Nakayama study than in the Rensink and Enns (1998) study.

A final possibility for why depth relations may have been unavailable for the early guidance of visual search in the present study, even if they were represented early, concerns the fact that different processes might require different types of representation of the same environmental information. The type of representation that is sufficient for purposes of perceptual completion behind occluding surfaces, for example, may be different from the type of representation that is needed for guidance and explicit report within a visual search task. Thus, representations of depth relations (or indeed other types of spatial relations) may be established early and in parallel across the visual field, and that information may be available to early perceptual organization processes, such as completion. Despite being available to these perceptual organization processes, however, the nature of these early representations may not be of the type that is needed for the processes that support the guidance of attention during search or for the explicit recognition of the relations.

An explanation of this sort can account for the pattern of results across the studies described here. It also fits well with Logan's (1994) interpretation of his data and with the models that he and his colleagues developed from them (Logan, 1995; Logan & Compton, 1996; Logan & Sadler, 1996). Specifically, they sought to draw a link between the visual apprehension of spatial relations and the explicit linguistic constructs involved in describing spatial relations between items. This link would be required when the explicit report of those relations was needed, but not necessarily when it was not. Finally, a different-types-of-representation explanation of this sort also fits well with recent findings that concern the role of attention in early perceptual organization processes. Specifically, Mack, Tang, Tuma, Kahn, and Rock (1992) reported results showing that patterns formed by Gestalt grouping by similarity are unavailable for explicit report when the to-be-grouped stimuli fall outside of the focus of attention. Using an implicit measure, however, Moore and Egeth (1997) found that the grouping patterns even in unattended regions of the display were established sufficiently to feed into and support other perceptual processes. Thus, without attention, representations of the stimuli were available for

processes of perceptual organization, despite being unavailable for explicit report.

## Conclusions

The main finding of this study is that, despite several different attempts, no evidence was obtained to support the hypothesis that targets defined by depth relations can support efficient visual search. These findings place depth relations in the same category with other spatial relations, such as up-down and left-right relations (e.g., Logan, 1994; Palmer, 1994). Whether representations of depth relations are nonetheless established and available for other early perceptual organization processes, despite being unavailable to the visual search system, remains unclear.

## REFERENCES

- ENNS, J. T. (1992). Sensitivity of early human vision to 3-D orientation in line-drawings. *Canadian Journal of Psychology*, **46**, 143-169.
- HE, Z. J., & NAKAYAMA, K. (1992). Surfaces versus features in visual search. *Nature*, **359**, 231-233.
- JULESZ, B. (1971). *Foundations of cyclopean perception*. Chicago: University of Chicago Press.
- LOGAN, G. D. (1994). Spatial attention and the apprehension of spatial relations. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 1015-1036.
- LOGAN, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology*, **28**, 103-174.
- LOGAN, G. D., & COMPTON, B. J. (1996). Distance and distraction effects in the apprehension of spatial relations. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 159-172.
- LOGAN, G. D., & SADLER, D. D. (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom, M. A. Peterson, L. Nadel, & M. Garritt (Eds.), *Language and space* (pp. 493-529). Cambridge, MA: MIT Press.
- MACK, A., TANG, B., TUMA, R., KAHN, S., & ROCK, I. (1992). Perceptual organization and attention. *Cognitive Psychology*, **24**, 475-501.
- MOORE, C. M., & EGETH, H. (1997). Perception without attention: Evidence of grouping under conditions of inattention. *Journal of Experimental Psychology: Human Perception & Performance*, **23**, 339-352.
- NAKAYAMA, K., & HE, Z. J. (1995). Attention to surfaces: Beyond a Cartesian understanding of focal attention. In T. G. Papathomas (Ed.), *Early vision and beyond* (pp. 181-188). Cambridge, MA: MIT Press.
- NAKAYAMA, K., HE, Z. J., & SHIMOJO, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In S. Kosslyn & D. N. Osherson (Eds.), *An invitation to cognitive science: Visual cognition* (2nd ed., pp. 1-70). Cambridge, MA: MIT Press.
- NAKAYAMA, K., & SILVERMAN, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, **320**, 264-265.
- O'CONNELL, K. M., & TREISMAN, A. A. (1990, May). *Is all orientation created equal?* Paper presented at the meeting of the Association for Research in Vision and Ophthalmology, Sarasota, FL.
- PALMER, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research*, **34**, 1703-1721.
- PODER, E. (1999). Search for feature and for relative position: Measurement of capacity limitations. *Vision Research*, **39**, 1321-1327.
- RENSINK, R. A., & ENNS, J. T. (1995). Preemption effects in visual search: Evidence for low-level grouping. *Psychological Review*, **102**, 101-130.
- RENSINK, R. A., & ENNS, J. T. (1998). Early completion of occluded objects. *Vision Research*, **38**, 2489-2505.
- STEINMAN, S. B. (1987). Serial and parallel search in pattern vision. *Perception*, **16**, 389-398.

- WHEATSTONE, C. (1838). On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London: Series B*, **128**, 371.
- WOLFE, J. M. (1994). Guided Search 2.0. *Psychonomic Bulletin & Review*, **1**, 202-238.
- WOLFE, J. M. (1998). What can 1 million search trials tell us about visual search. *Psychological Science*, **9**, 33-39.
- WOLFE, J. M., & BENNETT, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research*, **37**, 24-43.
- WOLFE, J. M., CAVE, K., & FRANZEL, S. L. (1989). Guided Search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 419-433.

#### NOTES

1. Because RT was the main dependent measure, and subjects were instructed to make their responses as quickly as possible, the purpose of the ER ANOVAs was to provide assurance against drawing conclusions from patterns in the RT data that could have been due to a

speed-accuracy tradeoff. We therefore present  $F$  and  $t$  values for the ER analyses only when there is a significant effect in the opposite direction of that in the RTs.

2. Owing to equipment failure while running these experiments, we were forced to throw out nearly as many subjects as were kept, because we could not be absolutely certain that they were experiencing depth, though they may have been. To be safe, however, we replaced the data for any dubious case. There was no pattern to the elimination (i.e., no more subjects were replaced for Experiment 5A than for Experiment 5B).

3. We wish to emphasize that this interpretation with regard to depth relations is not something that He and Nakayama (1992) concluded, but rather is something that we extracted as a further suggestion from their results. They were concerned with whether the establishment of surface representations, as reflected through evidence of perceptual completion, occurred early and in parallel.

(Manuscript received April 29, 1999;  
revision accepted for publication July 14, 2000.)