

# Influence of Scene-Based Properties on Visual Search

James T. Enns and Ronald A. Rensink

The task of visual search is to determine as rapidly as possible whether a target item is present or absent in a display. Rapidly detected items are thought to contain features that correspond to primitive elements in the human visual system. In previous theories, it has been assumed that visual search is based on simple two-dimensional features in the image. However, visual search also has access to another level of representation, one that describes properties in the corresponding three-dimensional scene. Among these properties are three dimensionality and the direction of lighting, but not viewing direction. These findings imply that the parallel processes of early vision are much more sophisticated than previously assumed.

It is easy to detect a vertical line placed among a group of horizontal lines. The vertical line “pops out”, drawing attention to itself regardless of how many horizontal lines are present. In contrast, searching for a T-shaped target among L-shaped distractors requires conscious effort, and search time increases linearly with the number of L-shaped distractors in the display. These two classes of search exemplify the visual search paradigm, a useful tool for determining the primitive elements of early human vision.

In theories of visual search, it is hypothesized that there are two subsystems (1-3). The first is a preattentive system capable of detecting simple features (e.g., oriented lines) in parallel across the image. Processes at this stage do not detect spatial relations between features (e.g., the relative locations of line segments). The spatial relations can only be determined by a second system that inspects each collection of features in a serial fashion.

When talking about features, however, one must distinguish between the world of objects in three-dimensional space (i.e., the *scene*) and its projection onto a two-dimensional array (i.e., the *image*). In a scene of objects illuminated by a distant point source, the array of image intensities is determined by: (i) direction of lighting, (ii) surface locations and orientations, (iii) surface reflectances, and (iv) viewing direction. These properties are captured by relations among image features.

The experiments in this report tested the sensitivity of preattentive vision to several of these relations. Target and distractor items were composed of polygons shaded with one of three intensities: white (all pixels lit), gray (alternate pixels lit), and black (no pixels lit) (Figs. 1 through 4). Some of the items corresponded to projections of simple blocks under various

orientation, viewing, and lighting conditions (Figs. 1A, 2A, and 3). Others could not be interpreted as three-dimensional objects (Fig. 1 B to D, and Fig. 2, B and C). We asked whether visual search could distinguish among these items, and if so, which scene properties were relevant.

A Macintosh computer was used to generate the displays, control the experiments, and collect the data (4). Observers searched for single target item among 1, 6, or 12 items (5). The target was present in half the trials and randomly distributed throughout the trial sequence (6). Observers were instructed to maintain fixation and to keep errors below 10% (7). Target presence or absence was reported by pressing one of two response keys. In each experiment ten observers completed four to six sets of 60 trials per condition (8).

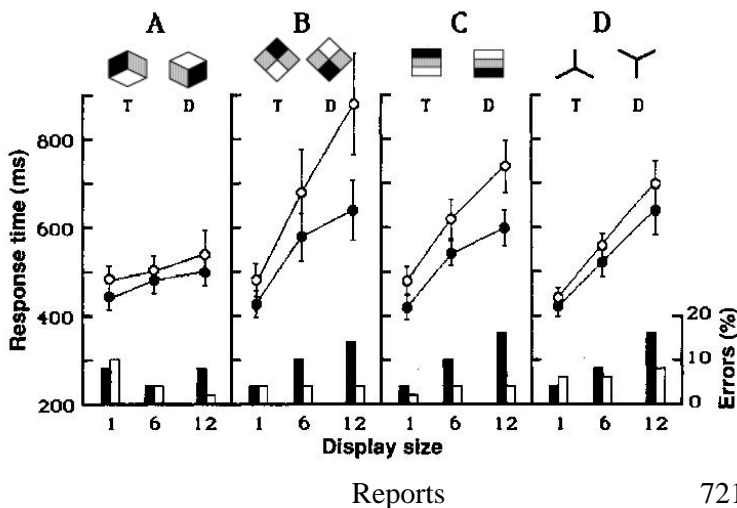
Experiment 1 demonstrated that certain relations among simple features can be detected preattentively. In Fig. 1A, items correspond to blocks differing in orientation and lighting. Regardless of how many items were in the display, observers were quick to report target presence or absence (6 ms/item for both conditions). In contrast, observers were much slower to find the target when items were two dimensional ( $p < .01$ ). For example, search for spatial relations among the polygons in flat items required 19 and 35 ms/item (Fig. 1B) and 15 and 23 ms/item (Fig. 1C). Search based only on the edges that distinguished the two blocks took 20 and 22 ms/item (Fig. 1D).

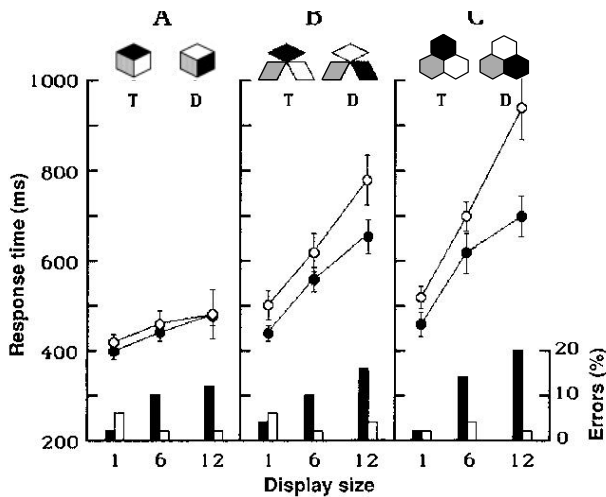
Because visual search is sensitive to spatial relations that capture three dimensionality, we asked whether it has a similar sensitivity to intensity relations. In experiment 2 we examined this question by using items that differed only in the intensities assigned to the polygons. Once again, search was rapid for items corresponding to three-dimensional blocks [8 and 6 ms/item (Fig. 2A)] and slow for

**Fig. 1.** Experiment 1. The target (T) and distractor (D) items in the 4 conditions (A to D). Filled circles and bars represent data from target-present trials; open circles and bars represent target-absent trials. (A) Search is rapid when the items correspond to 3-dimensional blocks of different orientations and lighting. (B to D). Search is slow when items are 2-dimensional. Values are mean  $\pm$  SEM. Display size indicates number of item present in a trial.

J.T. Enns, Dept of Psychology, University of British Columbia, Vancouver, Canada, V6T 1Z4.

R.A. Rensink, Dept of Computer Science, University of British Columbia, Vancouver, Canada, V6T 1Z4.





**Fig. 2.** Experiment 2. Search is rapid for 3-dimensional blocks that differ only in lighting (A) but not for 2-dimensional items that have similar intensity relations (B and C). Values are mean  $\pm$ SEM. Display size indicates number of item present in a trial. Filled circles and bars represent data from target-present trials; open circles and bars represent target-absent trials.

items that could not be given such an interpretations [19 and 25ms/item;  $p < .01$  (Fig.2B) and 20 and 36 ms/item (Fig. 1C)]. Together, experiments 1 and 2 show that rapid search is possible only when items can be interpreted as three-dimensional objects.

Objects are more easily apprehended when they are below the line of sight (9) and more readily grouped when they are lit from the same direction (10). In experiment 3 we asked if these factors influence visual search (Fig. 3). Note that lighting direction for each item is determined by its pattern of intensities (Fig.4). Items with white tops can be interpreted as blocks lit from above; items with black tops as blocks lit from below. Conditions in Fig. 4A and 4B, compared to Fig. 4C and 4D, tested for the influence of lighting direction

Viewing from above resulted in somewhat faster base-line responses than viewing from below ( $p < .06$ ), but viewpoint had no significant effect on search slopes ( $p > .2$ ). In contrast, direction of lighting had a large effect on search rate ( $p < .01$ ). When the target had a black top, search was relatively easy (6 and 5 ms/item for Fig. 4A; 10 and 11 ms/item for Fig 4C). Search for a white-topped target was much slower (21 and 23 ms/item for Fig. 4B; 18 and 23 ms/item for Fig. 4D).

When a switch between target and distractor items leads to slower search, it indicates that the easily found items contain a primitive feature not present in the other (11). Thus, the results of experiment 3 show that the scene-based property captured by these items behaves like other preattentive features (1-3). This preattentive feature seems to be the deviation from the standard direction of lighting, i.e., lighting from above.

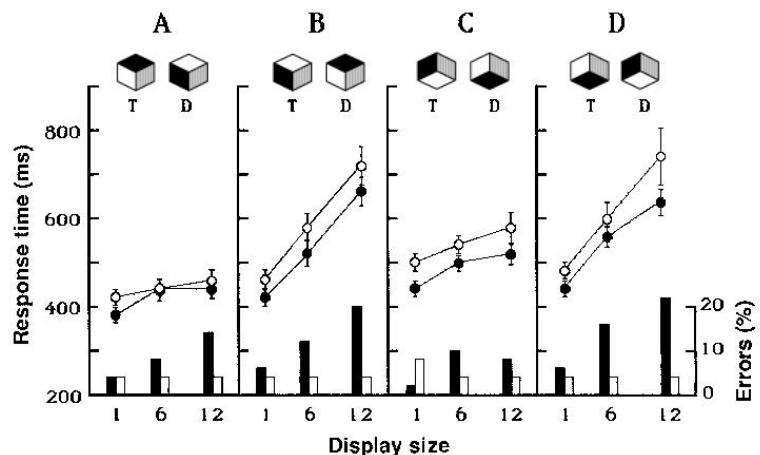
The findings of experiments 1 to 3 are akin to the discovery that rapid search can be based on conjunctions of features such as binocular disparity and color (12) and of motion and form (13). However, the features presented in these experiments are more complex, describing scene-based properties derived from spatial relations in single static images.

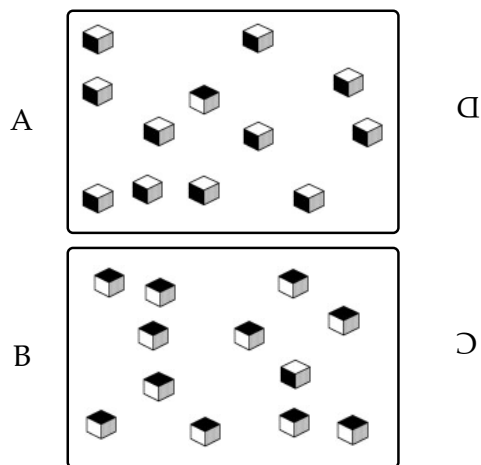
Which properties of the scene might these be? First, experiments 1 and 2 showed that early vision is sensitive to spatial and intensity relations that convey three-dimensionality (14). Moreover, experiment 1 (Fig. 1D) and 2 (Fig. 2C) showed that local image relations consistent with three-dimensional corners were not sufficient to permit rapid search. That is, the underlying processes seem to test for a consistent interpretation of the entire item. Although in machine vision this can be done by a constraints satisfaction algorithm such as line-labeling (15), further tests will be needed to determine whether this is the process used in human vision.

Second, not every relevant scene property had an influence on search. Experiments 2 and 3 showed that the direction of viewing had no effect on how easily a target could be found. We have run additional tests that generalized this result for blocks rotated 60° and 90° from those in experiments 2 and 3.

Third, our experiments showed that visual search can be influenced by the directions of lighting in the items, although other scene properties may also be involved (16). As such, our results are consistent with reports (10,17) that viewers are able to assign the correct direction of lighting to a scene only on the basis of intensity gradients in an image. However, our results support two stronger claims: (i) that preattentive processes determine lighting direction

**Fig. 3.** Experiment 3. Search is rapid for a bottom-lit block among top-lit blocks (A and C), but slow for a top-lit block among bottom-lit blocks (B and D). Viewing from above (A) and (B) versus viewing from below (C) and (D) does not affect search rate significantly, although viewing from below (C) does slow baseline response time somewhat. Values are mean  $\pm$ SEM. Display size indicates number of item present in a trial. Filled circles and bars represent data from target-present trials; open circles and bars represent target-absent trials.





**Fig. 4.** Examples of 12-item target-present displays in experiment 3. When viewed from above, the item with the black top stands out from white-topped items (A) more readily than the reverse arrangement (B). Turning the page upside down reverses the relative difficulty of the displays: (C) (black top) is now easier than (D) (white top). What remains constant, however, is that search for a black-topped item is still easier than search for a white-topped item.

for objects in parallel over the image and (ii) that it is the deviation from the standard direction that is detected most readily. We also note that these effects did not require intensities to be varied smoothly (10,17)—three intensities were sufficient. Perhaps the underlying processes make use of the fact that directions of lighting can be calculated by using only the orientations of the lines and the intensities of the three regions at each vertex in the image (18).

Taken together, these experiments imply that visual search has access to a level of representation that describes several properties of the three-dimensional scene. Therefore, search cannot be based entirely on the simple properties thought to be encoded at the earliest stages of cortical processing (e.g., two-dimensional orientation, contrast, and motion registered by neurons in area 17). Either these cells are also sensitive to scene-based properties, or else visual search must access areas higher in the cortical hierarchy. In addition, these findings suggest that computational studies of vision should examine the extent to which scene properties can be computed in parallel early in the visual stream (19).

#### REFERENCES AND NOTES

1. J. Beck, in *Organization and Representation in Perception*, J. Beck, Ed. (Erlbaum, Hillsdale, NJ, 1982), pp. 285-317.
2. B. Julesz, *Trends Neurosci.*, **7**, 41 (1984).
3. A. Treisman, *Sci.Am.* **255**, 114B (November 1986).
4. J.T. Enns, E.P. Ochs, R.A. Rensink. *Vsearch: Macintosh Software for Experiments in Visual Search*. 1989.
5. Each item subtended less than  $1.5^\circ$ . Items were placed randomly on an imaginary  $4 \times 6$  grid subtending  $10^\circ \times 15^\circ$  arc, and were randomly jittered by  $0.5^\circ$  to prevent influences of item collinearity.
6. Each trial began with a fixation symbol for 750 ms, followed by the display, which remained visible until the observer responded. The response was followed by accuracy feedback (a plus or minus sign), which served as the fixation point for the next trial.

7. Although each observer maintained an overall error rate of less than 10% in each condition, there were systematic differences in accuracy (Figs. 1 to 3). In particular, target-present trials led to more errors than target-absent trials, as is commonly noted [e.g., R.Klein and M.Farrell, *Percept & Psychophys*, **46**, 476 (1989); GW Humphreys et al. *J Exp Psychol. Gen.* **118**, 258 (1989).] Most important for present purposes, however, was the observation that errors increased with response time, indicating that observers were not simply trading accuracy for speed.
8. The response time data were analyzed as follows: First, simple regression lines were fit to the target-present and -absent data for each observer (average fit of these lines ranged from  $r=0.87$  to 1.00 among conditions and experiments). Second, the estimated slope and intercept parameters were submitted to analyses of variance. Finally, Fisher's least significant difference (LSD) tests were used to determine the reliability of simple effects in the context of significant main effects and interactions. The reported  $p$  values, therefore, refer to LSD tests and (by implication) to the higher-order effects of which they are a part.
9. L. Kaufman, *Sight and Mind* (Oxford, New York, 1974).
10. V.S. Ramachandran, *Sci Am* **259**, 76 (August 1988).
11. A. Treisman and S. Gormican, *Psychol. Rev.*, **95**, 15. (1988).
12. K. Nakayama and G.H. Silverman, *Nature*, **320**, 264. (1986).
13. P. McLeod, J. Driver, J. Crisp, *ibid.* **332**, 154. (1988).
14. This finding is similar to the observation that briefly presented lines are detected more readily when the surrounding lines can be interpreted as a three-dimensional object [N. Weisstein and C.S. Harris, *Science* **186**, 752 (1974); J.T. Enns and W. Prinzmetal, *Percept & Psychophys* **35**, 22 (1984).
15. A.K. Mackworth, *Artif. Intell.* **4**, 121 (1973); *Perception* **5**, 349 (1976).
16. Strictly speaking, experiment 3 shows that search is influenced by the pattern of intensities assigned to the faces of the same three-dimensional block. Since these intensities are a joint function of the direction of lighting, surface orientation, and surface reflectance, at least one of these factors must be represented preattentively. We discuss the direction of lighting account in the text, but the other two factors may also be relevant. Consider first an account based on surface orientation. If we assume a constant direction of lighting, items can be interpreted as blocks with black, gray, and white faces. If we further assume that a three-dimensional orientation is assigned to the same-color face in each block (for example, the black face), then the face with the incongruent orientation should stand out. Alternatively, search may be governed by surface reflectance. If observers are able to group the blocks preattentively on the basis of the orientation of one of the faces, then the face with the incongruent color should stand out. We have emphasized the direction of lighting because the other two require arbitrary associations to be made between the orientations and colors of the faces of the blocks. We have neither empirical nor computational grounds to show that such associations are the basis for search.
17. J.T. Todd and E. Mingolla, *J. Exp Psychol. Human Percept Perform.* **9**, 583 (1983).
18. B.K.P. Horn. *Artif. Intell.* **8**, 201. (1977).
19. Most parallel algorithms in computational vision are based entirely on local quantities [W.E.L. Grimson, *Comput. Vis Graph Image Process* **24**, 28 (1983); R.J. Woodham, *Artif. Intell.* **17**, 117 (1981)]. However, our results (Fig. 1D and 2C) show that rapid search cannot be based only on local configurations—a consistent interpretation of the whole item is required. To determine such consistency, algorithms must use information outside the immediate location of any single location.
20. Supported by Natural Sciences and Engineering Research Council (J.E. and R.R. through R.J. Woodham) and UBC Centre for Integrated Computer Systems Research (R.R.). We thank E. Ochs for programming assistance, A. MacQuistan for collecting data, and A.K. Mackworth and E. Bandari for helpful comments on earlier drafts.