1996-11

Approaching Visual Search in Photo-Realistic Scenes

https://hdl.handle.net/2144/2333

APPROACHING VISUAL SEARCH IN PHOTO-REALISTIC SCENES

Robert K. Cunningham, Jacob Beck, Ennio Mingolla

November 1996

Technical Report CAS/CNS-96-034

Boston University Center for Adaptive Systems and
Department of Cognitive and Neural Systems
677 Beacon Street
Boston, MA 02215

Approaching Visual Search in Photo-Realistic Scenes

Robert K. Cunningham[1,2], Jacob Beck[1], Ennio Mingolla[1]

[1]Center for Adaptive Systems
Dept. of Cognitive and Neural Systems
Boston University
677 Beacon Street
Boston, MA 02215

[2]Machine Intelligence Technology
MIT Lincoln Laboratory
244 Wood Street
Lexington, MA 02173

Ennio Mingolla is supported in part by the Office of Naval Research (ONR N00014-94-1-0597 and ONR N00014-95-1-0409).

Robert K. Cunningham can be reached at rkc@xn.ll.mit.edu.

Keywords: Visual search, shading, shadows, illumination direction, photo-realistic rendering, naturalistic scene

Running Head: Visual search in photo-realistic scenes

Portions of this work were presented to the annual meeting of the Association for Research in Vision and Opthalmology in Florida, May 1996.

Abstract

Visual search is extended from the domain of polygonal figures presented on a uniform background to scenes in which search is for a photo-realistic object in a dense, naturalistic background. Scene generation for these displays relies on a powerful solid modeling program to define the three dimensional forms, surface properties, relative positions, and illumination of the objects and a rendering program to produce an image. Search in the presented experiments is for a rock with specific properties among other, similar rocks, although the method described can be generalized to other situations. Using this technique we explore the effects of illumination and shadows in aiding search for a rock in front of and closer to the viewer than other rocks in the scene. For these scenes, shadows of two different contrast levels can significantly decrease reaction times for displays in which target rocks are similar to distractor rocks. However, when the target rock is itself easily distinguishable from distractors on the basis of form, the presence or absence of shadows has no discernible effect. To relate our findings to those for earlier polygonal displays, we simplified the non-shadow displays so that only boundary information remained. For these simpler displays, search slopes (the reaction time as a function of the number of distractors) were significantly faster, indicating that the more complex photo-realistic objects require more time to process for visual search. In contrast with several previous experiments involving polygonal figures, we found *no evidence* for an effect of illumination direction on search times.

1.0 Introduction: Why Photo-Realistic Displays?

Most laboratory visual search experiments require participants to find a discrete target item among a variable number of discrete distractor items. In almost all previous search experiments, target and distractor items are either letters, or are composed of line segments or polygonal elements (occasionally with polyhedral interpretations) scattered about a uniform background, usually with no occlusion. (See, for example, Figure 1.) Often the items are presented in a highly regular display, such as on a circle equidistant from the fixation point, or scattered about a rectangular grid. (For examples of circular displays, see Cohen and Ivry, 1991. For examples of rectangular displays as well as pointers to related experiments, see Treisman and Gelade, 1980; Wolfe, Cave and Franzel, 1989; Wolfe, 1994b; Grossberg, Mingolla and Ross, 1994.)

For regularly arranged, discrete stimuli, the definition of terms such as the number and the density of the distractors is relatively straightforward, and classification into fast/slow search can be done relatively easily. However, such displays only approximate the real-world task of visual search, where objects occlude one another, butt against one another, cast shadows on one another, and are irregularly placed throughout the visual field. Unfortunately, few published experiments reveal what makes visual search fast or slow in a real-world visual search task. In this paper we describe two sets of experiments that examine, for photo-realistic images, the effect of shadows and illumination direction on visual search.

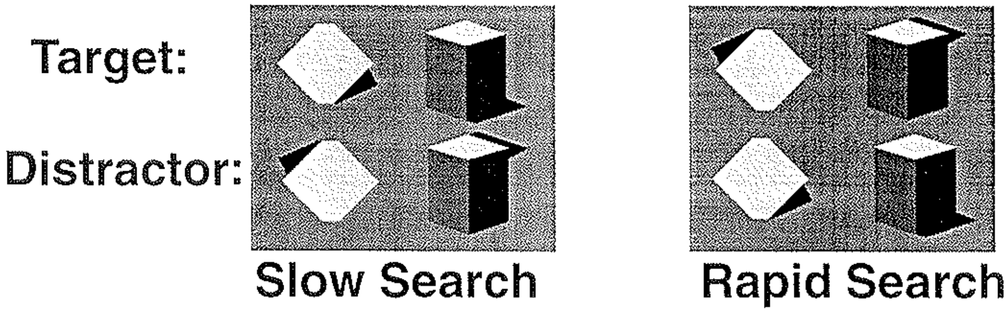


FIGURE 1. Shadows in polygonal visual search. "Shadows" that appear below the object that casts them among shadows appearing above require more time to locate than shadows appearing above the casting object among shadows appearing below. (After Rensink, R. A., and P. Cavanagh (1993).)

2.0 Towards Photo-Realistic Visual Search Scenes

A few experiments have attempted to examine visual search tasks in continuous, natural backgrounds. In the experiments of Biederman, Glass and Stacy (1973), participants searched for an object (e.g. a fire hydrant) among distractors (e.g. objects in a street scene) in photos that were cut into sixths and reassembled. In all of the resulting images the cut lines were visible, and since the objects are not exactly one sixth the size of the image, some contextual information remains. Thus, although the images started out as continuous, real-world scenes, in the displays the objects are separated from one another by unnatural dividing lines. More recently, Zelinsky, Hayhoe and Ballard employed digitized images of a few objects (hand tools) on a scene-appropriate background (an empty workbench) in a search paradigm, but the items were distributed so that no item abutted or overlapped any other object (Zerlinsky et al., 1995).

Another approach employed image tiles "painted" with a digital painting program representing overhead terrain views (Wolfe, 1994). The images were cleverly designed so that each tile seamlessly fit with its neighbor, giving the appearance of a continuous surface when fully assembled. Wolfe's image tiles contained objects of a variety of colors, orientations and sizes, but none of them would be confused with photographs of natural scenes.

Recent advances in the computer processing available to the average user permits a significant improvement: by using a solid modeler coupled with a ray tracing program, a model can be built of a real-world scene and the properties of that scene can be directly manipulated. Unfortunately, micro-computer processing is not yet fast enough to permit rendering full screen scenes during a visual search experiment, so an alternate approach must be employed. After examining trade-offs between computational capacity, persistent storage needs and experimental needs, we elected to construct scenes by sampling from tiles that were independently rendered prior to initiation of the experiment. (See Appendix 1.) The selected tiles were then butted together to construct a visual scene. In order to make the seams between tiles invisible, care was taken to create tiles in which top and bottom tile edges (and left and right edges) formed a continuous image when abutted. This was accomplished by carefully controlling illumination as well as placement of objects which crossed the tile edges.

3.0 Visual Search with Shading, Cast Shadows and Occlusion

Visual search with three dimensional naturalistic objects permits experiments not possible with two dimensional polygonal drawings. Perhaps the most natural extension to the previous 2D work involves search for an object at a different relative depth than that of other objects. Cues for relative position in static scenes include *occlusion*:, objects in front occlude those behind; *size*, larger objects tend to be nearer objects; *cast shadows*, objects closer to an illumination source can shadow objects farther away from that source; *shading*, local brightness is determined by the illumination source, the surface normal and reflectance off of neighboring objects; and *relative position*, nearby objects resting on a sloping ground plane are projected to the bottom of the scene. Others have examined the effect of size gradients and contextual ground plane cues (Sun and Perona, 1996), and cast shadows (Rensink and Cavanagh, 1993). We will examine the effects of occlusion, shading and cast shadows in scenes where the target object is in front of other objects in the scene.

In earlier work (Cavanagh and Leclerc, 1989), the necessary conditions for perceiving depth from shadows were studied. The authors varied the contrast of texture components in both the shadow and non-shadow regions. Participants then modulated the contrast between shadow and non-shadow regions until they perceived a three dimensional scene. In order to perceive depth from shadows the shadow regions had to be darker than the surrounding, non-shadow regions. In addition, contrast polarity had to be consistent at both macro and micro scales at the shadow border (i.e., at shadow-non-shadow boundaries, texture components in the darker region had to be darker than the abutting texture components in the lighter region). In Cavanagh and Leclerc's displays dark shadows and lighter shadows differed only in intensity, but in real scenes very dark shadows obscure the background in the scene, while less dark shadows do not completely obscure the background. In some cases, the darkest shadows appear as stains in a scene, rather than as shadows (Beck, 1972).

In one visual search task involving shadows (Rensink and Cavanagh, 1993), participants were required to detect shadows rather than to use the shadow to aid in relative depth discrimination. The authors found that visual search is slower when detecting dark regions attached *below* a polygon among dark regions attached *above* a polygon than vice-versa. (See Figure 1.) If illumination is from above, if the polygon is interpreted as a polyhedron and if it rests on a planar surface, then the dark regions below the polygon can be interpreted as shadow. If similar assumptions are made with illumination from below, then the dark regions above the polygon can likewise be interpreted as shadows. This asymmetry in visual search reaction time is similar to an asymmetry found for search among shaded polygons which resemble differently illuminated polyhedra. One example of this appears in Figure 2.

4.0 Illumination Direction

Many psychological textbooks contain a visual illusion which purports to demonstrate that, in the absence of other information, the human visual system is biased toward assuming that illumination is from overhead. Although the demonstration originated more than a century and a half ago (Brewster, 1826), the commonly presented modern image is of a moon crater that appears to be a circular elevation when inverted. Early in this century there was a great debate as to whether this bias was innate or learned (Metzger, 1936; von Fieandt, 1938). Early work (von Fieandt, 1938) showed that children from 4-7 years old exhibit the same biases that adults do, suggesting that if this bias is learned, then it is learned quite young. Later work with 3-, 5-, and 7-year old children confirmed this result (Benson and Yonas, 1973), but added the caveat that children must learn how to interpret illumination sources in pictures–when pictures of objects with alternate interpretations that depended on knowledge of the position of illumination were oriented horizontally,

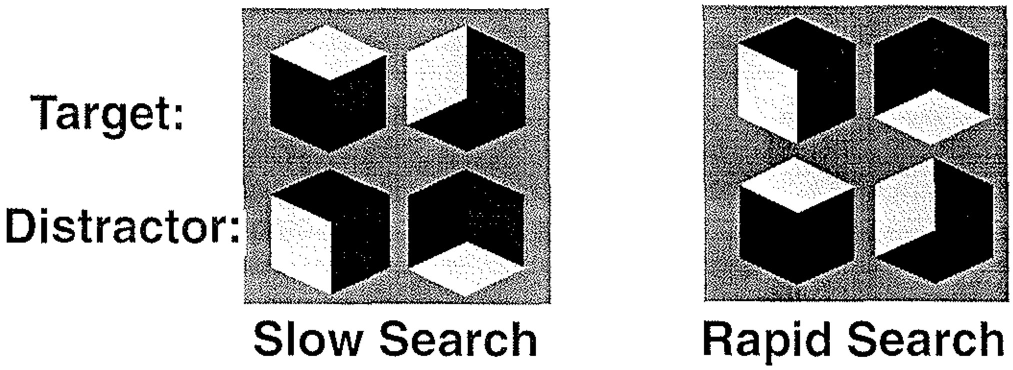


FIGURE 2. Shaded parallelograms which give rise to a 3D box percept. Search is rapid when the target object is illuminated from below, but slow when the target object is illuminated from above. (After Enns, J. T., and Rensink, R. A., 1990.)

children did not make the adult assumption that illumination comes from the top of the photo, but when the pictures were oriented vertically, both groups made the assumption that light comes from overhead. Research with chickens reared in an environment illuminated from below also indicates that the assumption of overhead illumination was innate (Hershberger, 1970), although earlier work had suggested otherwise (Hess, 1947; Hess, 1950). Later work with humans indicated that the reference frame is determined with respect to the head; illumination "from above" is determined in retinal coordinates (Kleffner and Ramachandran, 1992). It is this last work that began to look at illumination direction in the context of visual search, although the targets and distractors were computer-shaded disks that can be perceived as illuminated spheres, although they are drawn images rather than photos of physical phenomena or photo-realistically rendered scenes.

Additional insight into assumptions of scene analysis is gained from the shape-from-shading literature (Reichel and Todd, 1990). Results from experiments in which participants are asked to determine relative location in depth support the view that the visual system is biased towards viewing objects from above rather than viewing objects from below. Results from experiments in which participants are asked to search for an object at a particular relative depth yield the same conclusion. Reichel and Todd suggest that much of the bias for illumination from above described by others can be accounted for by a bias towards viewing objects from above.

An alternate interpretation of the shaded disk results of Kleffner and Ramachandran (1992), consistent with other visual search work (Enns and Rensink, 1990), is that all bistable objects are seen (on the short timescale of a visual search experiment) as convexities, and that the convexity in conjunction with illumination from below enables rapid search. Instead of shaded disks which give rise to the interpretation of an illuminated sphere, Enns and Rensink (1990) use shaded parallelograms which give rise to the interpretation of illuminated boxes. (See Figure 2.) As in the spherical convexity/concavity experiments, a solitary shaded box illuminated from below is easy to find (i.e., search is very rapid) among shaded boxes illuminated from above, whereas a shaded box illuminated from above is hard to find among those illuminated from below. This is true regardless of whether the boxes are seen from below or above, although the baseline for objects seen from below is slightly higher. This bias may be interpreted in light of results from the shape-from-shading literature (Reichel and Todd, 1990), in which it is reported that there is a bias towards seeing objects from above.

Other recent work examined the effect of "room context", in which the line where walls met the floor or where walls met the ceiling are shown, and perspective projection is used to vary the apparent position of the illuminated cubes (Sun and Perona, 1996). The results indicate that the cubes sitting on the floor sped up visual search (although non-monotonically as a function of the number of distractors), and the cubes hanging from the ceiling slowed down visual search (also non-monotonically with the number of distractors). Although the work of Sun and Perona extends visual search so that the background shows a more realistic connection to the target objects, the objects in the display appear to be floating in the space near the ceiling or floor without being attached to either.

With this work in mind, we extend the visual search paradigm by performing a new series of experiments with the following properties: (1) objects are rendered with accurate models of complex surfaces and illumination; (2) backgrounds are constructed in such a way as to ensure that there is consistency between a target object and the background on which it is displayed; (3) scenes are rendered on fronto-parallel planes, to avoid preferences for viewing objects from below or above. Using these techniques we re-examine the effects of illumination direction on visual search scenes. In many of our experiments, we also consider backgrounds which contain shading and shadow patterns that are consistent with illumination covering a complete image or scene.

By examining shadows and shading for dense, photo-realistic scenes, we hope to answer several questions: (1) Do shadows aid search for an object in depth? (2) Does shadow intensity affect the search rate? (3) Do these complex displays speed search (because the rich shading patterns provide an immediate sense of 3D) or slow search (because the rich shading patterns take longer to process) when compared with images which only include the rock borders?

4.1 General Methods

4.1.1 Constructing Naturalistic Stimuli: Tools and Techniques

Stimuli were created on a Power Macintosh 6100/60 using Strata Vision 3D (Woodland et al., 1994). Each 3D object was first modeled by explicitly modifying wireframe representations of fieldstones until a craggy, realistic looking rock shape was achieved. A granite texture with fine grain was then mapped onto each fieldstone. Once a sufficient collection of fieldstones was assembled, piles of fieldstones were assembled by packing each rock model onto a ledge so that it didn't overlap with another rock or the ledge--in effect, laying the models down much as one would lay down real 3D rocks. (See Figure 3.) Two advantages are gained by using computer rendering tools rather than real rocks: first, it is possible to accurately control many aspects of illumination (in some cases enabling comparisons that would be impossible to make using real objects and real lights), and second, can easily construct scenes that reuse the same boundary, thereby allowing the creation of a seamless tiled scene. Notice how the left and right boundaries and the top and bottom boundaries of the rock pile in Figure 3 are made out of identical rocks, thereby allowing a tile cut from this pile to seamlessly butt up against a second tile. By reusing the boundary region for every pile constructed for a given experiment, one can be certain that each tile will seamlessly connect up to its neighbor. Recall that our tiling scheme is dictated by capacity limitations for storing complex images. (See Appendix 1.)

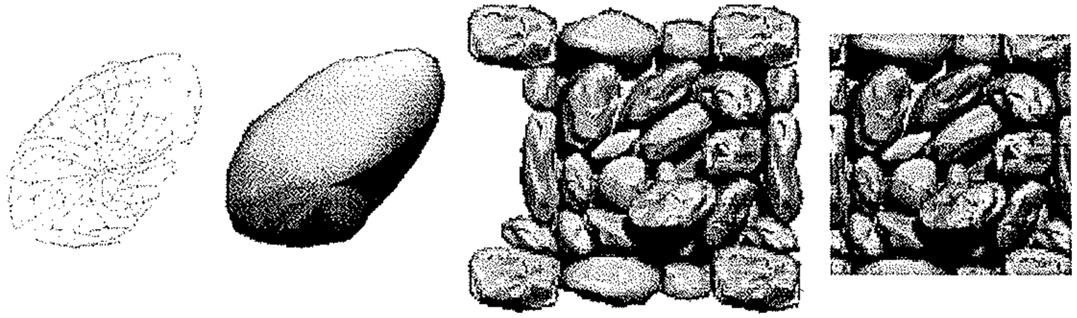


FIGURE 3. Tile construction for continuous, naturalistic stimuli. First a wireframe model for a rock is constructed. A granite texture is then wrapped onto the wireframe. Many rocks are then piled together onto a ledge, an illumination model is defined, and an image is rendered. (Note how the horizontal and vertical sides are identical.) A tile is then cut from the rock pile.

The rock pile is then rendered in greyscale with 3x3 anti-aliasing, fine texture detail and orthographic projection using the StrataVision 3D ray-tracer. Once a tile is constructed and rendered, the central portion is cut and imported into VSearch (Enns et al., 1990) which then randomly selects from among eight tiles: four non-target tiles and four target tiles. Images are gamma corrected (using the "Mac Standard Gamma") and displayed on an Apple color monitor.

After each experiment was run, participants were asked if they realized that the scenes were constructed out of tiles, and if they used a search strategy that exploited that knowledge. Fewer than half of the naive participants realized the construction method. Of those who did realize the scene was constructed out of tiles, none claimed that they ignored the tile boundaries (where target rocks could not reside because of the way a scene is assembled from tiles).

4.1.2 Participants

All participants were graduate students in the Cognitive and Neural Systems program at Boston University and had normal or corrected to normal vision. The authors took part in pilot experiments, but their results were not included in any of the reported experiments. For experiment one, all five participants were naive participants. For convenience, all five participants ran experiments 2, 3 and 4 in one sitting. For these experiments, one participant was experienced with visual search experiments, another had taken part in other similar experiments, and three others were naive. For experiment 5, one participant participated in other similar experiments, and four others were naive.

4.1.3 Common Procedure

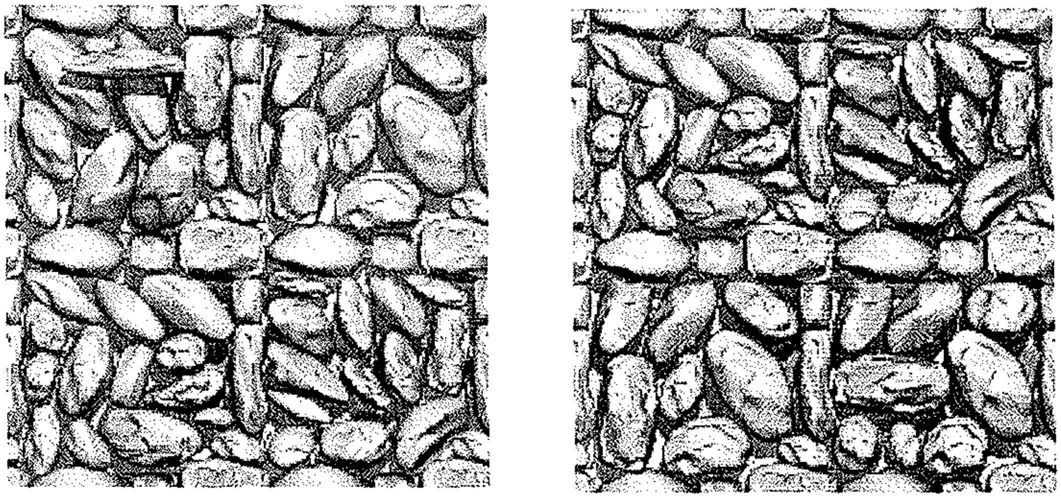
Participants were seated with their chins in a chin rest located 60 cm from the computer monitor. For each trial a participant reported "present" if an object in the scene is located in front of (i.e., closer to the participants than) other rocks in the display. Participants were instructed to respond "as rapidly as possible without making errors." Trials were ordered into blocks in which the main experiment conditions were kept constant. Within each block, the number of tiles presented varied so that either one, four and nine or two, four and nine tiles (depending on experiment) were presented in each block. The number of trials in each block varied from experiment to experiment. (See the description for each experiment.) Prior to the actual experiment, each participant performed a practice run in which the entire set of experiment conditions was presented; however, each block in the practice run consisted of a small number of trials (roughly 1/3 of the number in the actual experiment). Since all participants ran all conditions, all ANOVA analysis used within-participant tests.

5.0 The Value of Shadows

5.1 Experiment 1: Target Object Shape

The first experiment examined the conditions under which shadows and shading would affect search times for these displays.

Methods. To reduce the chance that participants were identifying the target rock by features distinct from those of other rocks, two different target rocks were used: one narrow and the other wide. Each was presented resting in approximately the same position on top of rocks in one of two different target background configurations, only one of which was displayed in any single target present trial. The two positions were at opposite sides of the tile, so viewers could not learn to ignore portions of the tile. One of the factors examined was the effect of the presence of soft shadows compared to no shadows. (See Figure 4 for an example of the narrow target with soft shadows.) In the shadow conditions the foreground and the background both contained shadows of equal intensity, so detection of the target could not be accomplished merely by detecting a shadow; participants had to determine the position of the target rock in the scene. Illumination arose from two sources: an ambient source (set to 61% of maximum), and a parallel light ray source (set to 85% of maximum) that shone from 45 degrees above the viewing direction and 15 degrees to the right. (Parallel light rays simulate illumination from a distant source such as the sun or moon.)



Target Present **Target Absent**

FIGURE 4. Four tile visual search scene. Left: target present. Right: target absent. This example is from Experiment 1, narrow target with shadow. Notice how seamless the image is, even though it is constructed from four tiles.

Shadows only provide depth information for illumination originating in the hemi-sphere from which the rock pile is observed, because illumination from the other hemi-sphere cast shadows that fall on the side of the rocks away from the observer. The technique for building scenes from tiles further limits the usable cone of illumination origins to about ±45 degrees, because beyond that shadows become so long as to extend beyond the tile boundaries. Flatter, more tightly packed rock piles could expand this cone, but pilot renderings examining illumination from oblique angles revealed that the long shadows associated with these angles were more difficult to rapidly interpret.

Shadow and no shadow conditions differed only by the presence or absence of a shadow (a feature that can be turned on or off in rendering); all shading in the scene, other than in the area of the shadow itself, was identical. Such nearly identical conditions would be impossible to recreate using real stones and lights.

Finally, the background in the immediate vicinity of the target was varied so as to change the amount or type of occlusion. We call the three configurations "Good Continuation", because one of the occluded rocks passed entirely behind the target; "No Continuation", where one end of the rock that passed behind the target was moved off at an angle so as to make the resulting pieces appear to be two pieces; and "Near Maximum Occlusion", because almost all of the bounding contour of the target rock was in front of another rock. Thus there were six separate conditions in the experiment: Good Continuation, No Continuation and Near Maximum Occlusion, all with and without shadow. (See Figure 5.) Blocks in the practice run consisted of 15 trials, while blocks in the two recorded runs had 50 trials, for 600 total recorded trials.

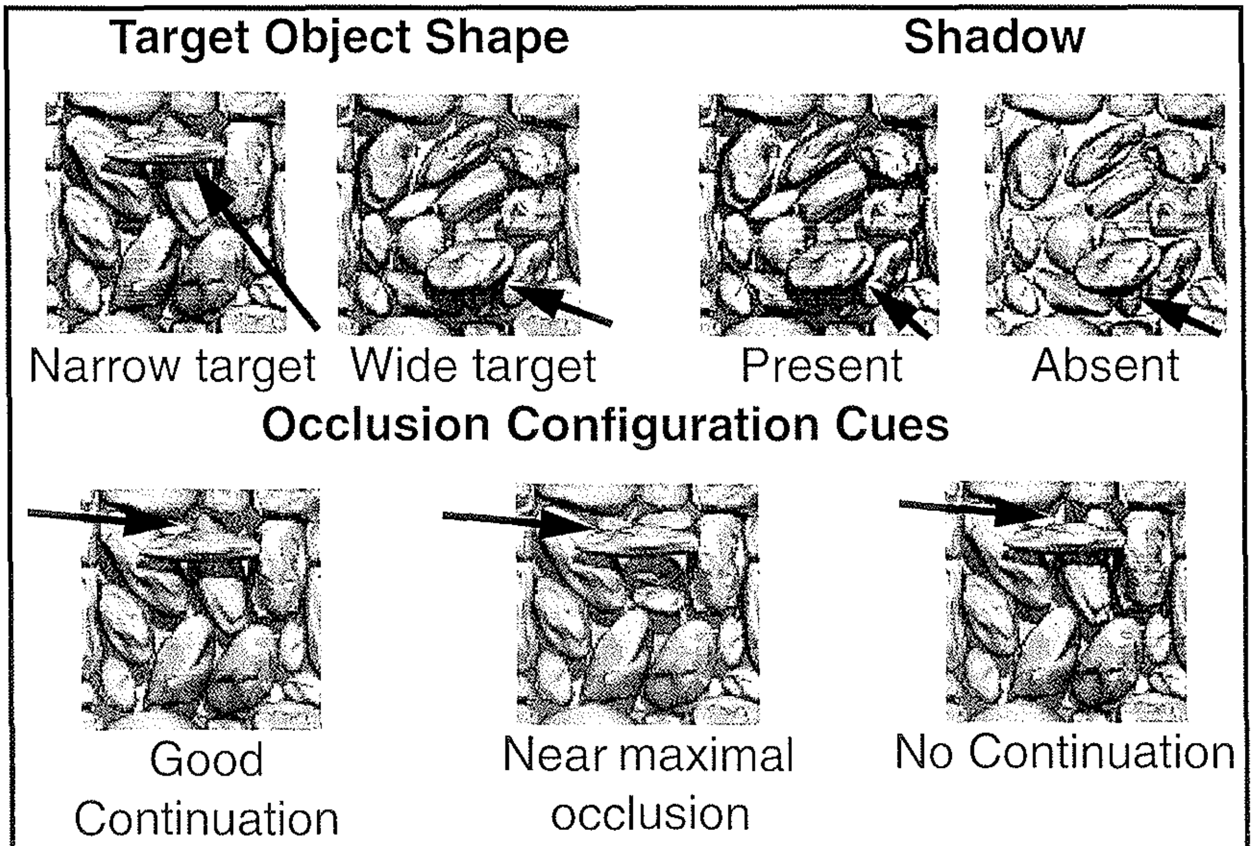


FIGURE 5. Conditions of Experiment 1.

Results. In order to report search rates, a definition of "distractor" must first be made. As noted before, it is unclear if visual search is object or area based, and recent work with visual search tasks show effects of grouping on search times (Treisman, 1982; Rensink and Enns, 1995). For this work both the number of fieldstones and the number of tiles in a display are easily defined. For reporting search rates, we will report time per fieldstone, because this measure yields search times that are comparable with the geometric visual search tasks.

Overall (within-subject) analysis for target present conditions reveals a significant effect of shadow, F(1,4)=65.58, $p<0.05$; a significant effect of target type, F(1,4)=19.65, $p<0.05$; a significant effect of the number of tiles in the scene, F(2,8)=27.23, $p<0.05$. It also reveals interactions of configuration x shadow, F(2,8)=10.28, $p<0.05$, shadow x target, F(1,4)=31.79, $p<0.05$; and shadow x number of tiles, F(2,8)=27.37, $p<0.05$. (See Figure 6.)

The interactions of target x shadow and shadow x number of tiles occurs because search for the narrow target is consistently rapid, regardless of the presence or absence of shadow, at approximately 4-5 msec/distractor fieldstone. For the wide target the presence of the shadows speeds the search; without the shadows the search rate is approximately 13 msec/distractor fieldstone, whereas it is approximately 4 msec/distractor fieldstone with the shadow present. The origin of the interaction of configuration x shadow is revealed by a Bonferroni test of the 9 tile results for the three configurations: there is a significant difference between reaction times for near-maximum-occlusion when compared with both good continuation and no continuation, df=24, $p<0.05$, but when the good continuation and no continuation configurations are compared, no significant difference is found, df=24, $p>.05$.

Discussion. Why is it that the narrow target produces rapid search times regardless of shadow presence or absence, whereas the wide target requires the presence of shadows? One hypothesis is that the narrow target is sufficiently elongated so as to excite the orientation tuned cells that are thought to speed search in the geometrical displays. (e.g., (Wolfe et al., 1989; Grossberg et al., 1994)). An alternative is that the shape of the narrow target is so different than that of the other rocks in the displays that the difference in forms causes visual pop-out to occur. Since the shapes are nearly ovoid, we can crudely represent their projected shape via a single number, the elongation, which is equal to the length of the longest projected axis divided by the length of its perpendicular axis. The elongation of a circle is one; all other ellipses have an elongation greater than one. Figure 7 displays the histograms for typical narrow and wide target present cases when nine tiles are present in the scene.

The fact that target type, shadow presence, and local context all significantly affect the participants' reaction times give support for the hypothesis that the participants were using depth cues to pick out the target rocks. Since search for the narrow target was universally fast, participants probably used the shape of the target itself as the primary cue rather than the cues of illumination and occlusion. It is unlikely that they used the wide target as the primary cue, since (1) the same fieldstone appears in one of the distractor displays, albeit at a different orientation, and (2) the response rate for the scene is influenced by the depth cues present in the scene. Therefore, in the remaining experiments only the wide target was presented.

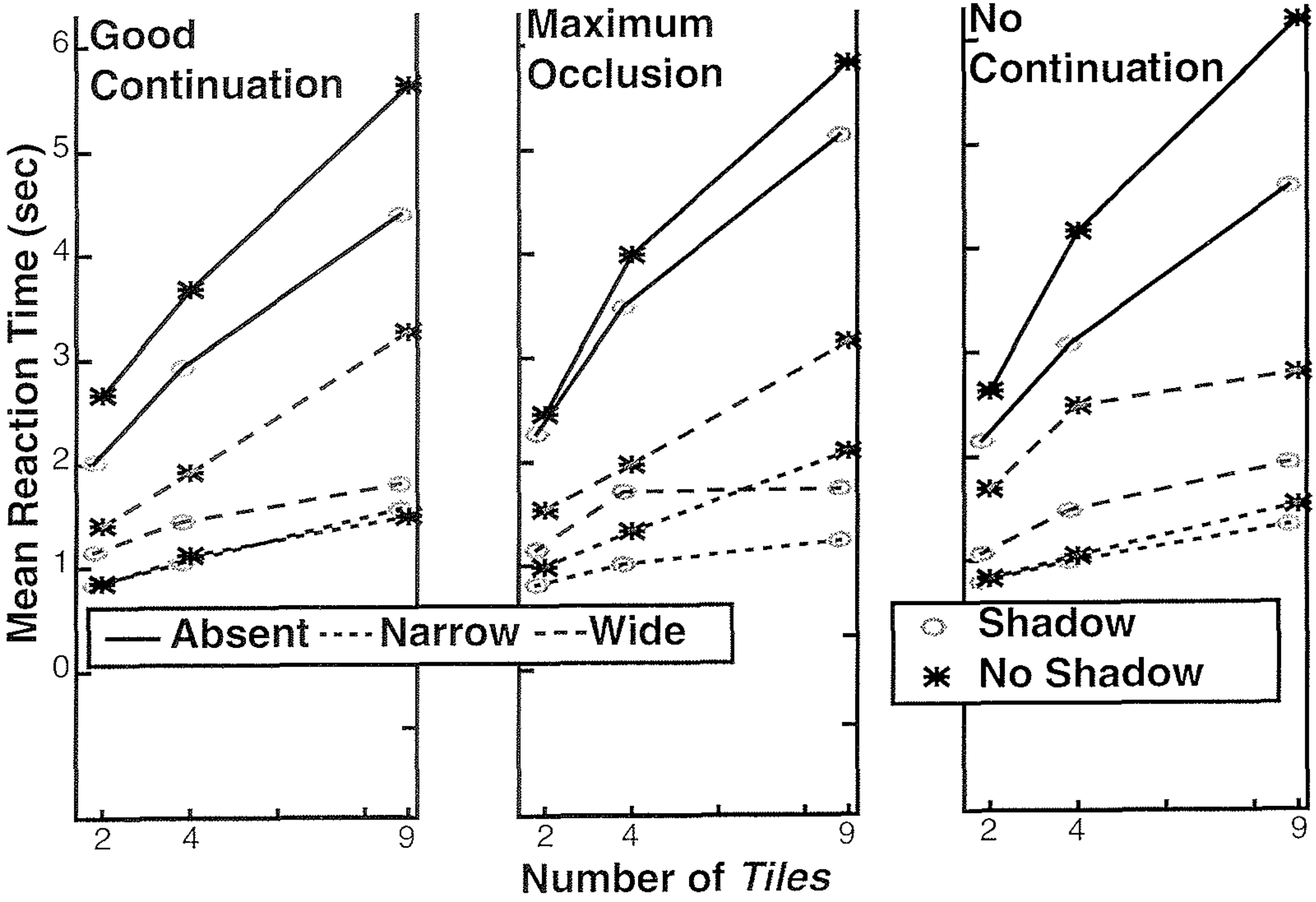


FIGURE 6. Results for Experiment 1. Results for each of the three configurations appear on separate graphs. Search for the narrow target is consistently rapid. The presence of shadows in the wide target scenes consistently increased search rate.

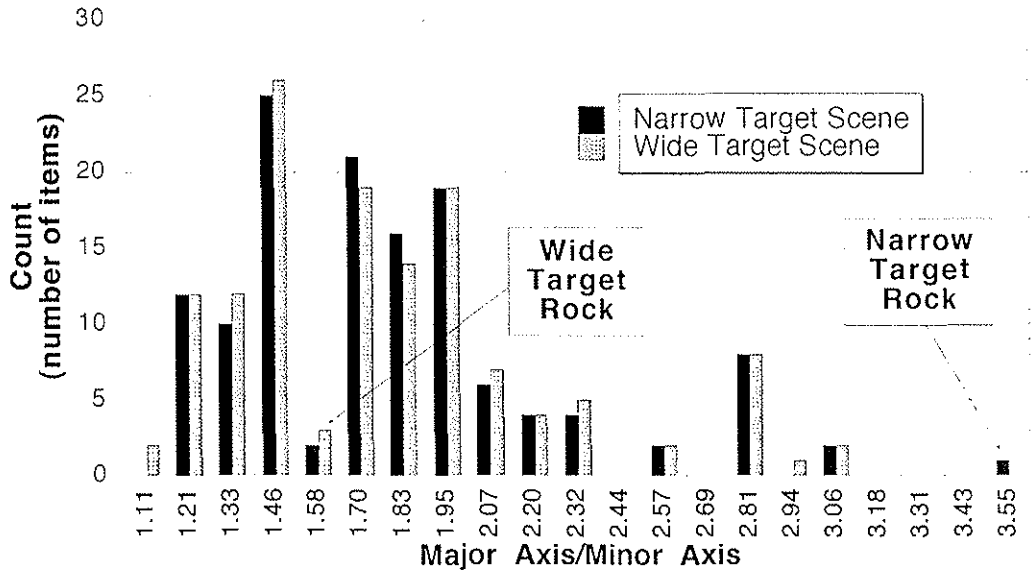


FIGURE 7. Histogram of the elongation of projected rocks for two typical nine tile, target present, search scenes. Elongation is defined as the length of the longest projected axis of a rock divided by the length of the perpendicular axis. The narrow tile is elongated when compared to the other rocks in the scene. The slight differences between the histograms for the wide target and the narrow targets arise because the tile in which the target is present differs slightly.

5.2 Experiment 2: Shaded and Shadowed Rock Displays

In this next experiment we examine the effect of the intensity of the shadow. The darkest shadows obscure occlusion information but provide greater contrast than the paler, softer shadows do. By varying the darkness of the shadow we examined if search is accelerated by the presence of a shadow for a range of shadow intensities.

Methods. Production of the shadows was by a different means than in Experiment 1, in order to more realistically model a series of possible illumination conditions for a real pile of rocks. Recall that in the first experiment the shading for the two scenes with and without shadows was identical (outside of the shadows), but that the shading pattern for the no shadow case was impossible for a real scene. (Recall that shadows were "turned off" by a option in the rendering program. This option effectively made an object transparent to those light rays that would normally be blocked and cause a shadow to be cast on a neighboring surface.)

In Experiment 2 the shading pattern for all cases was realistically rendered, but there were subtle differences between the shading due to changing illumination conditions. Illumination sources were set so that the mean intensity over a non-shadowed region was nearly identical for all three conditions, and the mean intensity of the soft shadow condition was equal to the mean intensity of the soft shadow condition in Experiment 1. (Compare the three conditions in Figure 8.) To ensure that these means were nearly identical, a single region of one tile was selected, cut from the tile and its mean computed and illumination parameters for the new scene were adjusted until the mean intensities of the same region matched. When the mean intensities matched, all tiles in the two experiments were visually inspected verify that the shading patterns were similar. In this experiment three light sources were present: the ambient light source, which was set to 25% of maximum, a parallel light ray source placed along the line of sight, and a second parallel light ray sourced placed at about 45 degrees overhead and 15 degrees to the right of center. For the no shadow case, all illumination came from along the line of sight[1], for the soft shadow case, some illumination came from above and some from along the line of sight, and for the hard shadow case, most illumination came from above. Four targets and four distractors were displayed in a series of unbalanced blocks. Participants observed all conditions in practice blocks of 16, then viewed all conditions again in random order in blocks of 64, for a total of 192 recorded trials.

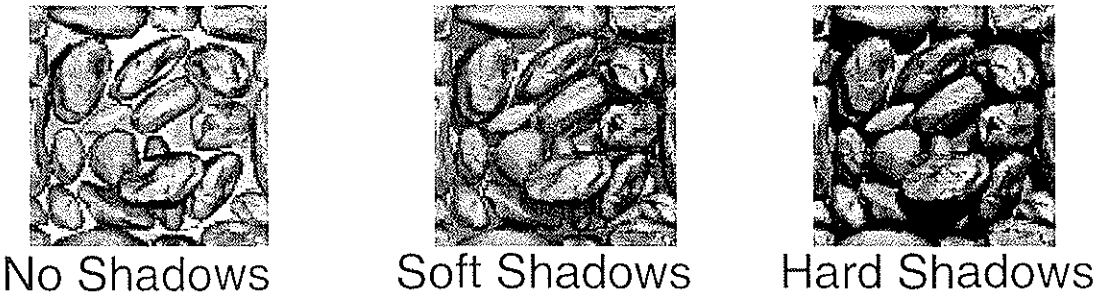

No Shadows Soft Shadows Hard Shadows

FIGURE 8. Conditions of Experiment 2. The no shadows condition provides relative depth cues via occlusion cues. The soft shadows condition provides relative depth cues via occlusion cues and the cast shadow. The hard shadows condition provides relative depth cues via occlusion cues everywhere but in the area of the high contrast shadow.

Results. Reaction time results appear in Figure 9. The findings agreed with those of Experiment 1: an overall (within-subject) analysis of variance reveals an effect of shadow, $F(2,8)=50.77$, $p<0.05$; of the number of tiles, $F(2,8)=40.83$, $p<0.05$; and an interaction between illumination and the number of tiles, $F(4,16)=45.07$, $p<0.05$. Further analysis with the Tukey test examining the three conditions (no shadow, soft shadow, hard shadow) of the nine tile trials reveals that reaction

---

[1] For this experiment the "no shadow" case is really misnamed; shadows exist but occur behind the objects in the scene, so they cannot be seen by the participants.

times for *both* hard and soft shadow are significantly different from that for no shadow (Tukey, q(3,12); $p < 0.05$). It also reveals that the differences between the hard and soft shadow reaction times for nine tiles are not significant (Tukey, q(3,12); $p > 0.05$).

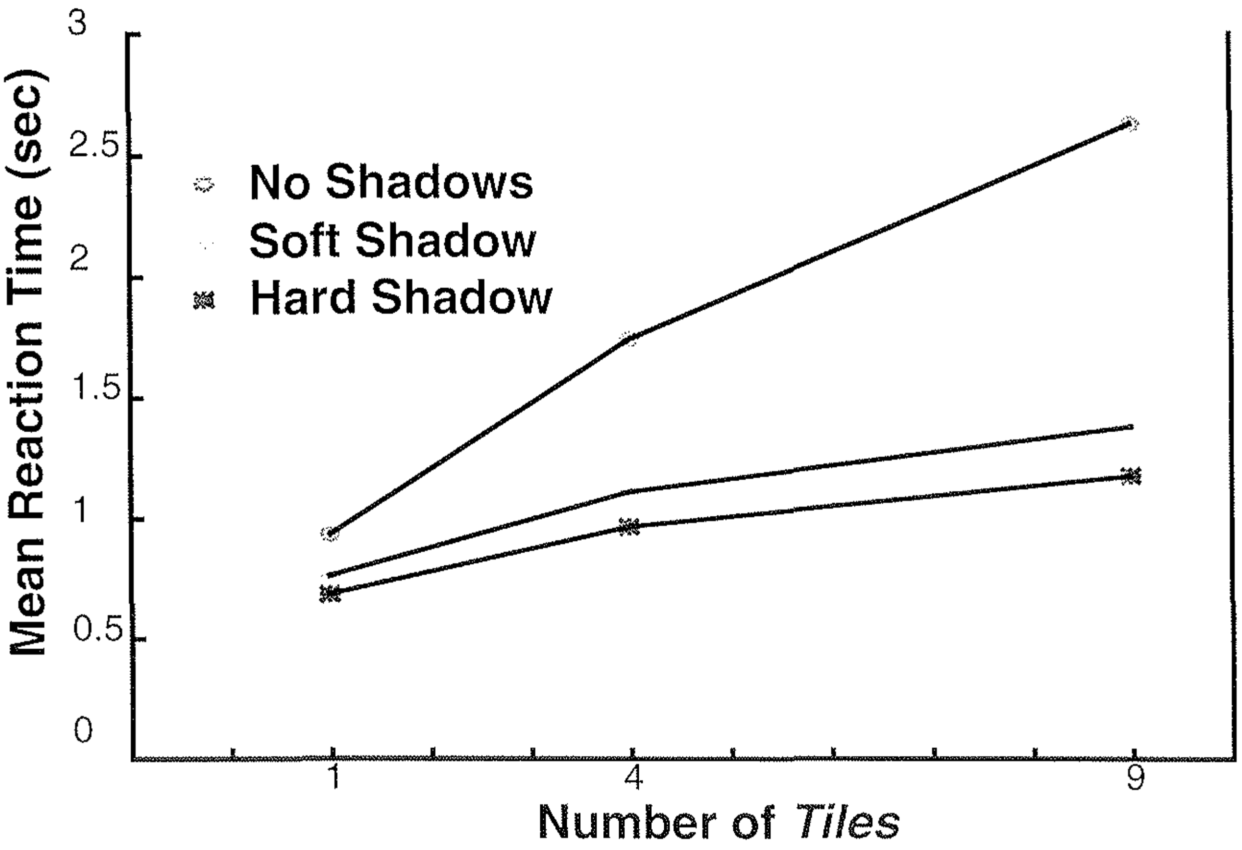


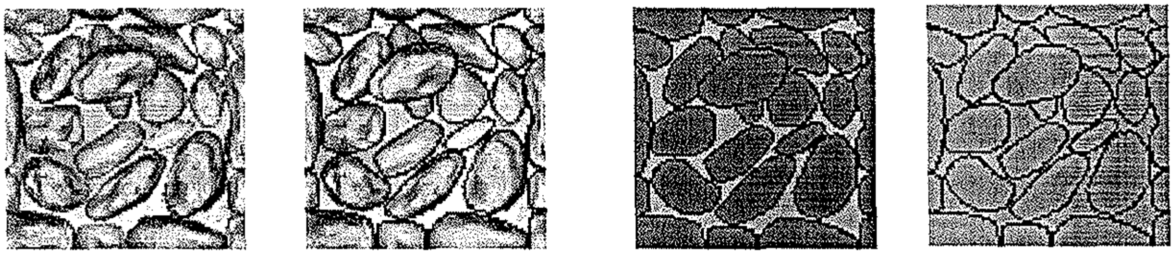
FIGURE 9. Results for Experiment 2, target present conditions. Both soft and hard shadow conditions are faster than the no shadow conditions, but they are not significantly different. (Target absent trial results appear in Figure 13.)

Discussion. Although the method for constructing shadow and no shadow conditions changed, the result did not: the presence of shadows in the scene significantly speeds the visual search task. In these displays as in the previous displays, shadows appear for all objects in the scene, including the background distractor rocks. Nevertheless, the larger shadow cast by the target rock facilitates visual search. Furthermore, the soft shadow, which did not hide occlusion, was not significantly faster than the hard shadow, which contrast more with the background but revealed no occlusion cues.

5.3 Experiment 3. Sources of Occlusion Cues

We next examined occlusion cues and compared results with our natural, photo-realistic displays to those using polygonal displays.

Methods. For this experiment the no shadow display of Experiment 2 is simplified to form a cartoon with the same occlusion cues but with none of the rich shading cues that make the scene seem so realistic. To form these cartoons, the boundary of each rock is traced in black. Once a boundary map is available, the subtle shading of each rock is reduced to a two-color drawing, with each rock shaded a uniform dark grey and the background colored a uniform light grey. The fieldstone displays are then further simplified by shading both background and fieldstones the same light grey. (See Figure 10.) As with experiment 2, there were 16 trials were present in the practice blocks, and 64 were present in the recorded blocks, for a total of 256 recorded trials.



Shaded fieldstones

Boundaries plus shaded fieldstones

Boundaries plus uniform colored fieldstones

Boundaries alone

FIGURE 10. Conditions of Experiment 3.

Results. A within-subject analysis of variance for the target present conditions reveals primary effects of the number of tiles, F(2,8)=161.09, $p<.05$; the condition, F(3,12)=4.37, $p<.05$; and interactions between the condition and the number of tiles, F(6,24)=2.69, $p<.05$. The condition x number of tiles interaction indicates that the different conditions have different slopes, a fact indicated by examination of the results presented in Figure 11. A comparison of slopes of multiple regression lines (Hald, 1952) reveals that the visual search rate for boundaries alone is significantly faster ($p<.05$) than the search rate for both shading alone and shading with boundaries, and that the search rate among the uniformly colored fieldstone shapes are faster ($p<.05$) than search rate for shading with boundaries.

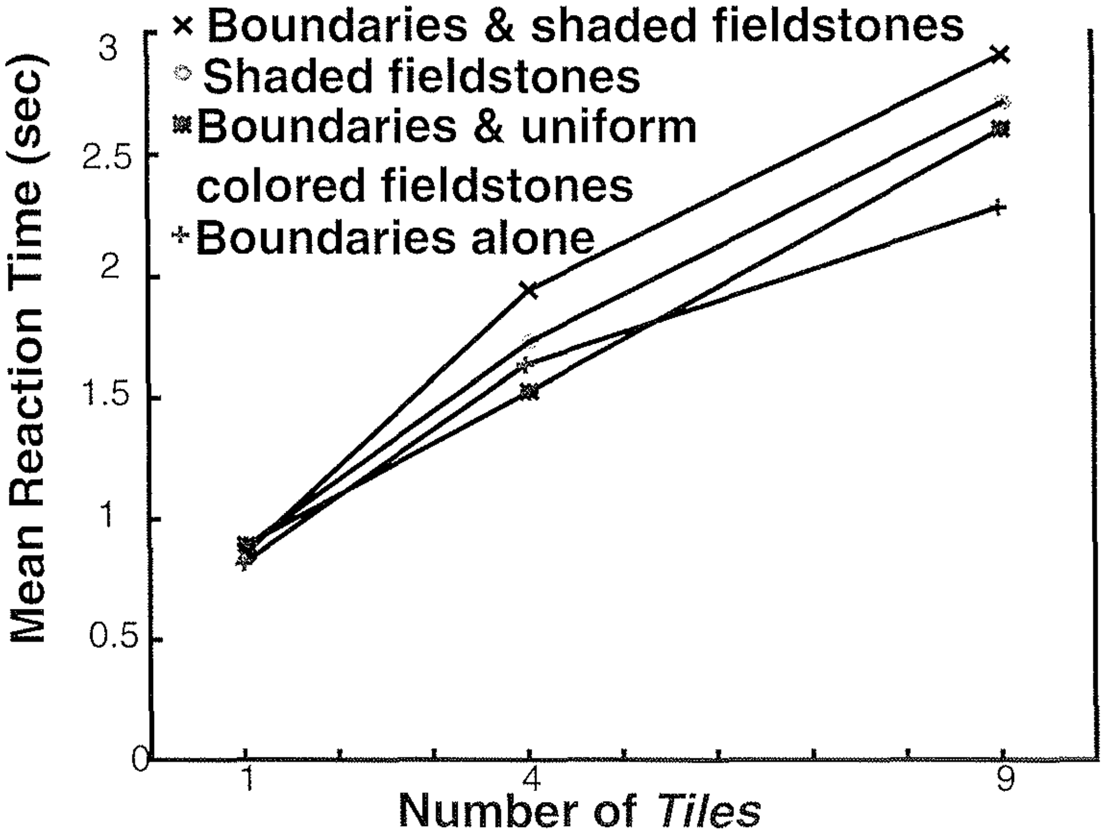


FIGURE 11. Reaction times for Experiment 3, target present conditions. The simplest scenes are the easiest scenes in which to perform visual search.

Discussion. The data shows that the simplest scenes are the easiest scenes in which to perform visual search. While the shaded scene yields a three-dimensional naturalistic percept, the information for the perception of relative depth in these scenes appears to come from boundaries that provide cues to occlusion. The more complex scene may be slower than the boundary scene either because it takes time to develop the boundaries (Grossberg and Mingolla, 1985), or because the added complexity of these scenes requires additional time to identify individual rocks. Since the reaction times for one tile are nearly identical, it is unlikely that retinotopic registration (Grossberg et al., 1994) or categorical feature coding (Wolfe, 1994) of the whole scene requires more time, but rather that the process of searching among rocks in the photo-realistic scene requires more scrutiny than in the boundaries-only scene.

5.4 Summary of Experiments 1-3

The three experiments described so far indicate that, in visual search for an object nearer to the observer than other objects in the scene (1) shadows can significantly speed the search, (2) either hard or soft shadows will speed up search over shading cues alone, and (3) search rate is faster when boundaries alone are used than when realistically shaded objects are used.

6.0 Shading and Shadows

6.1 Experiment 4: Illumination directions

In the next experiment we examined whether illumination source direction could affect reaction time for these stimuli. Illumination was both from above and below the line of sight. There are two ways to construct these illumination configurations; in the first the scenes can be kept constant and the light source can be moved, whereas in the second the light source can be kept constant and the rendered tiles can be inverted. For this experiment the rendered tiles were inverted, because the irregular rocks of Experiment 2 cast shadows of different shape and size when illuminated from the same angle above and below the line of sight. The illumination from above conditions are the same as those in Experiment 2, and the illumination from below conditions are inverted Experiment 2 tiles. (See Figure 12.) It is possible (although unlikely) that inverting the shading patterns may affect reaction time, so the no shadow condition is included in both orientations to serve as a control against this hypothesis. Although the shading in the shadow-present regions of the hard-shadow and soft-shadow conditions differs from that of the no-shadow control, the difference outside of the shadowed regions is subtle. As with experiment 2, there were 16 trials were present in the practice blocks, and 64 were present in the recorded blocks, for a total of 384 recorded trials.

The results, presented in Figure 13, show that for the three conditions' reaction time *no effect of illumination orientation occurs*, F(1,4)=1.54, p>.05. Although this is expected for the no shadow condition, it was not expected for the soft or hard shadow conditions, and is in opposition to numerous findings in the visual search literature where the objects are simple polygonal shapes on uniform backgrounds.

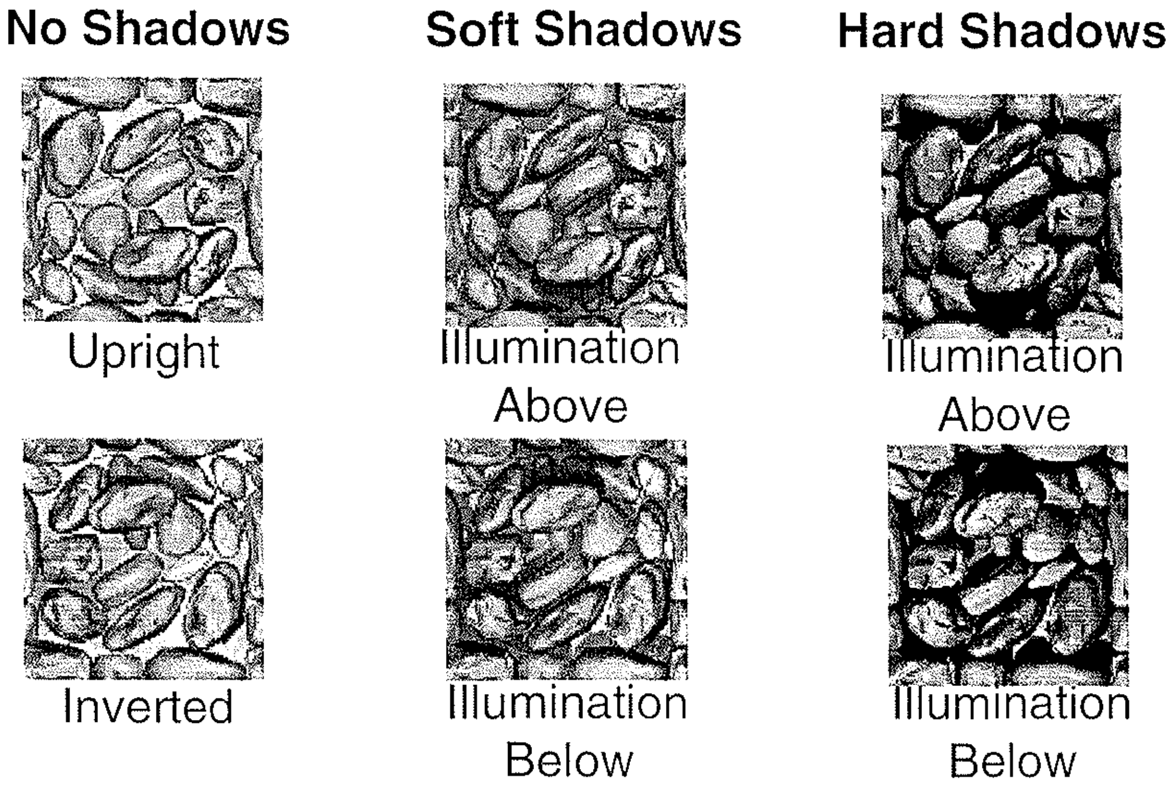


FIGURE 12. Stimuli for Experiment 4. Note that the conditions where illumination is from above are identical to Experiment 2.

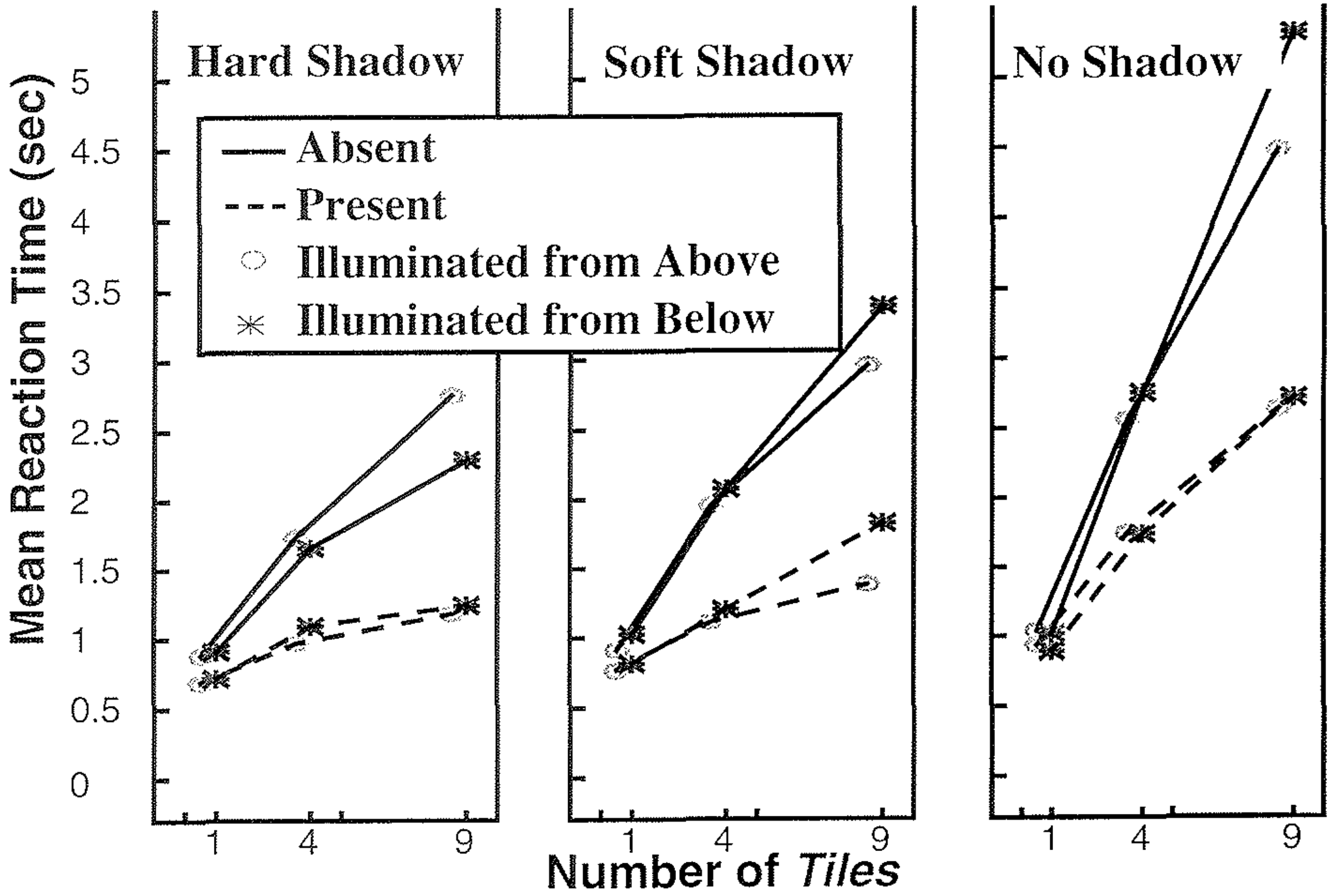


FIGURE 13. Results of Experiment 4. No difference is seen in reaction time between targets illuminated from above and targets illuminated from below, in contrast with other results found in polygonal visual search tasks. Compare the target present reaction times for hard and soft shadows (which would be expected to exhibit an asymmetry) with the no shadow reaction times and the target absent times (which would not be expected to exhibit the asymmetry).

Discussion. Experiment 2 demonstrates that the presence of shadows speeds reaction times for a target in these rich, photo-realistic displays, while Experiment 4 demonstrates that the direction of the illumination did not affect the participant's reaction time. This leads us to conclude that if shadows that appear below the objects that cast them are suppressed in some displays (c.f. Rensink and Cavanagh, 1993), they are not suppressed in the more realistic displays such as those presented here, at least given the task posed to participants. Also it appears that illumination does *not* need to come from above in order to be rapidly interpreted when the displays contain significant context for determination of illumination direction.

6.2 Experiment 5: Isolated shadow illumination directions

In this next experiment we further examine the relationship between these displays and the simpler, polygonal displays used in earlier studies. Unlike the "illumination models" of the earlier polygonal displays (Enns and Rensink, 1990; Kleffner and Ramachandran, 1992; Rensink and Cavanagh, 1993; Sun and Perona, 1996), the illumination of our photo-realistic displays is consistent throughout the entire scene. In the polygonal displays it is almost as if each object has its own source of illumination; that is as if there were a spotlight pointing directly at the objects which rest on a matte grey background. (A single illumination source will not suffice for scenes where the target object is "illuminated" from a different direction than the distractors.) However if each object does have its own spotlight, then there should be a border of light surrounding the object that is illuminated.

We attempted to recreate these rather unusual conditions and the more realistic spotlight illumination conditions by modeling a scene in which each target or distractor rock had a spotlight pointing at it. Additional background rocks still appear in the scenes, but they are not illuminated. To simulate the illumination conditions found in the polygonal displays, we manipulated the illumination of the entire scene so that the spotlit region was not illuminated. Compare the "polygonal" condition with the spotlight and halo conditions found in Figure 14.

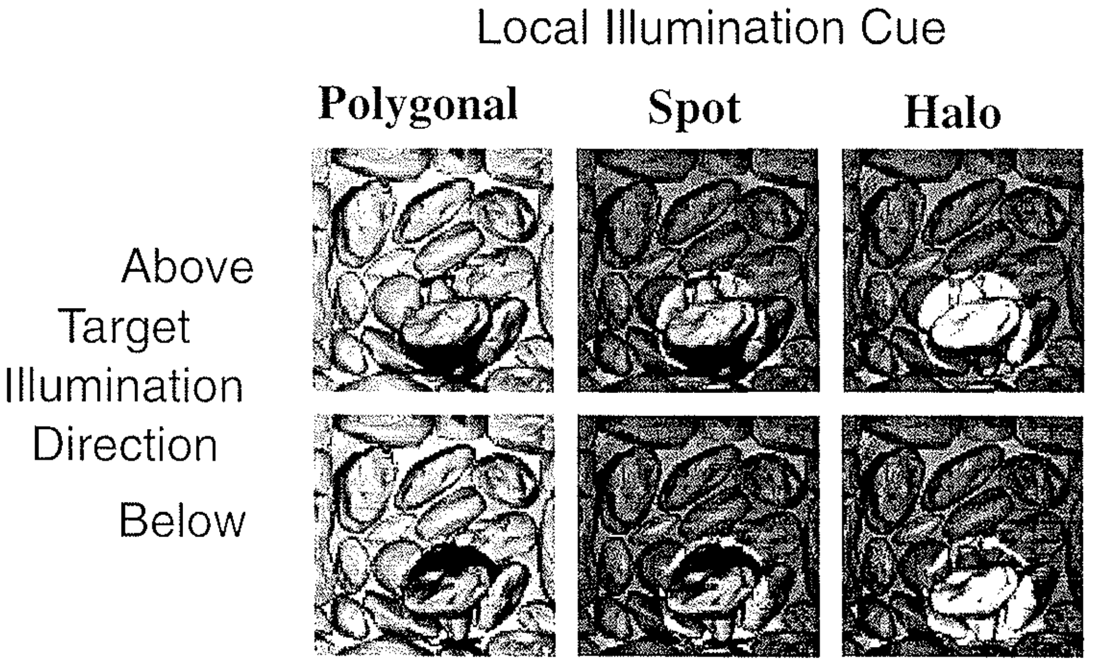
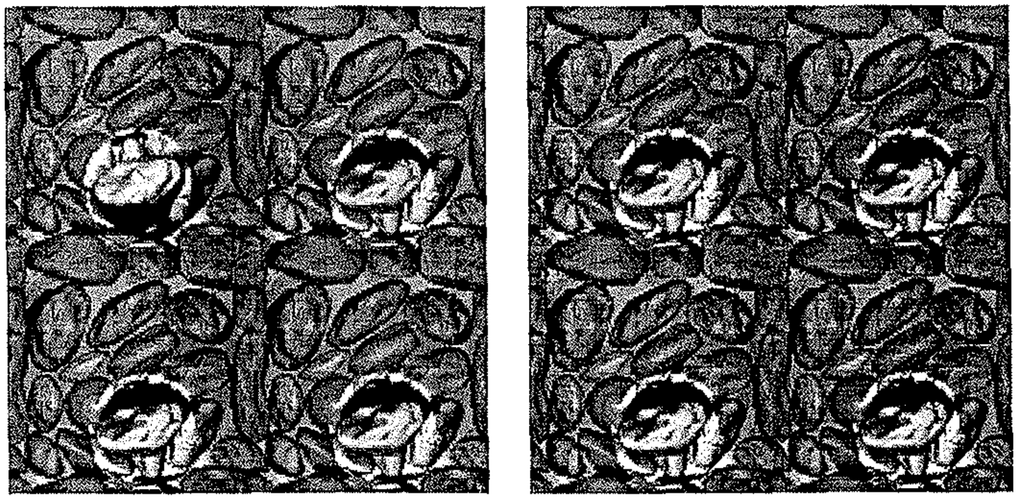


FIGURE 14. Conditions of Experiment 5.

We considered three cases: the reproduced polygonal display conditions, where the background is one color, the target is brighter than the background, and the shadow is darker than the background; the spotlight boundary present condition, in which the target was lighter than the background, the shadow was darker than the background and the spotlight boundary was present; and the halo condition, in which the shadow was the same intensity as the background and the spotlight boundary was present.

Methods. Unlike in previous experiments, illumination for a single set of tiles could not be rendered once and inverted to evaluate reaction time performance for illumination from above and below. The tiling approach used for these experiments requires that boundary rocks be identical for all tiles in a scene, so each target type (illumination from above, illumination from below) must be independently rendered using the same boundary rocks. Previous polygonal experiments used identical shapes and identical cast shadows for both illumination directions; in these experiments the best we can do is select illumination directions that cast shadows of approximately equal size without extending beyond the border of the tile. For this set of experiments illumination was from 20 degrees above and 20 degrees below the line of sight. (See Figure 15.) Each of the six conditions were practiced with 16 trials; order was randomized, scenes were balanced. The six conditions were then presented in two blocks with condition order randomized within each block; 64 trials were presented for each condition in each block. Participants were instructed to find either the target rock illuminated from above or below, depending on the trial.



Present

Absent

FIGURE 15. Four tile search scene from Experiment 5 with target illuminated from above. Left: target present. Right: target absent.

Results. As in the previous experiments, the reaction times are not affected by the illumination angle, $F(1,4)=0.02$, $p>0.05$. (See Figure 16.) Reaction times for the different cases are different, $F(2,8)=7.74$, $p<0.05$. Post-hoc analysis via Tukey pairwise comparison of all conditions with the same illumination angle reveals that the spot reaction times (both illumination from above and illumination from below conditions) are significantly faster than the halo reaction times, Tukey $q(6,24)$, $p<0.05$. None of the other conditions are significantly different

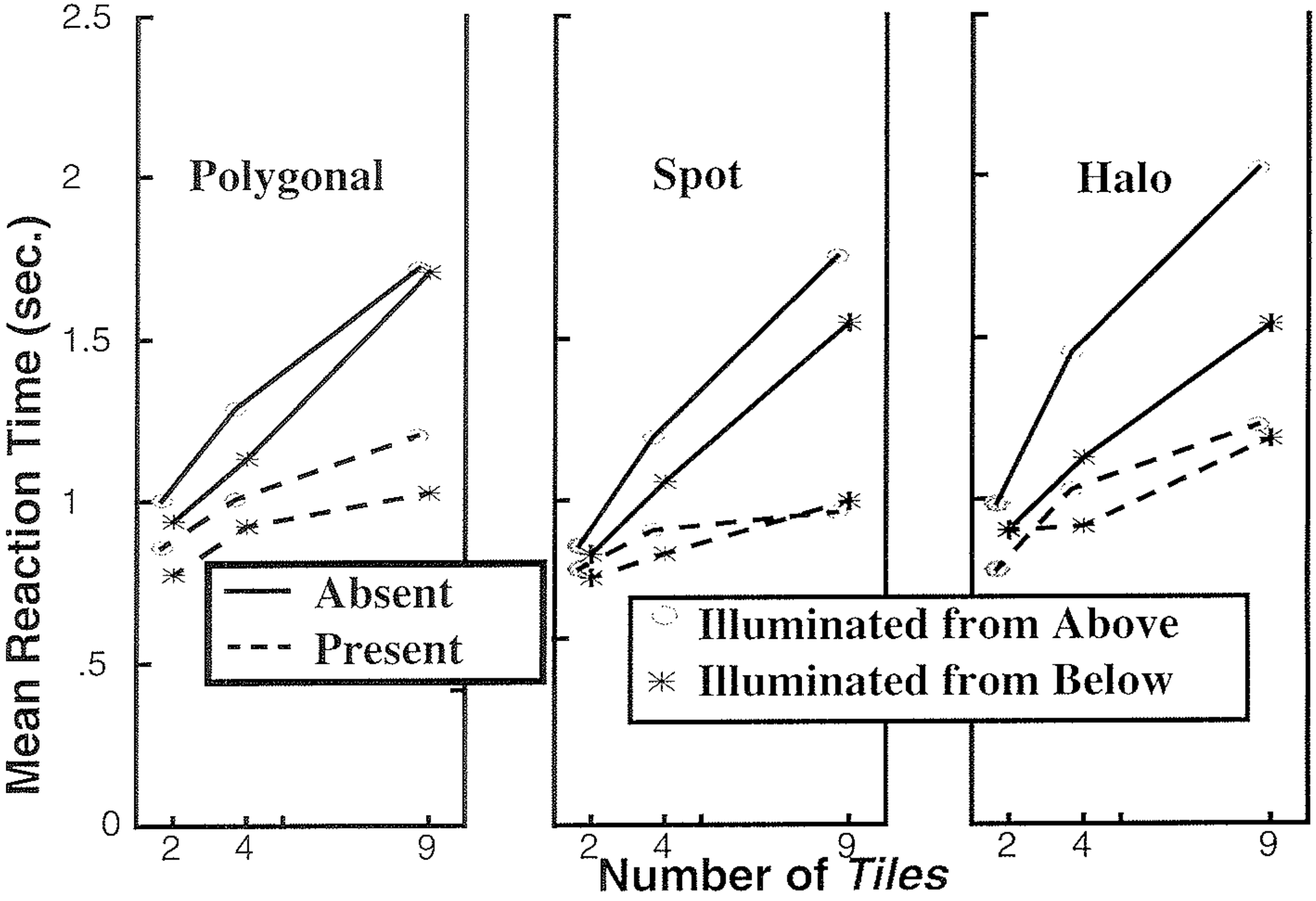


FIGURE 16. Results of Experiment 5. There is no difference between the reaction times for finding a target illuminated from above among targets illuminated from below and vice-versa.

7.0 Discussion

We have extended visual search for polygonal shapes and letters into the domain of photo-realistic scenes. In this set of experiments we examined the effects of illumination in conditions that were both more realistic and more varied than any experiment heretofore performed, and no search asymmetry due to illumination source position was found. It is unclear if the difference between these and earlier polygonal displays are due to the use of fronto-parallel displays or the realistic nature of these displays.

One possibility is that realistic displays provide sufficient context for perception of illumination direction that is strong enough to override the presumption of illumination from above (Hess, 1947; Hess, 1950; Hershberger, 1970; Benson and Yonas, 1973), whereas the polygonal displays do not. According to Johnston and Hawley (1993), a general characteristic of perception is the inhibition of expected stimuli and the facilitation of unexpected stimuli. If illumination is presumed to be from above in the polygonal figures of Rensink and Cavanagh (1993), then a shadow attached to the top of a figure is novel and unexpected while a shadow attached to the bottom of the figure is familiar and expected. The asymmetry follows from the hypothesis (Johnston and Hawley, 1993). The failure to find an asymmetry with the realistic targets may be due to the cast shadows and shading in the background fixing the illumination direction and overriding the presumption of illumination from above when illumination comes from below. In this case shading and cast shadows of the target and background are consistent with the direction of illumination, for both illumination from above and illumination from below. As a result, targets for both cases are equally expected and thus the reaction times for both cases are not significantly different.

Future experiments should examine the interactions between fronto-parallel displays and illumination direction. There is fertile ground for experimentation between these complex, photo-realistic displays and the polygonal displays that preceeded them.

8.0 References

Beck, J. (1972). Surface Color Perception. Ithaca, NY, Cornell University Press.

Benson, C., and Yonas, A. (1973). "Development of sensitivity to static pictorial depth information." *Perception and Psychophysics* **13** (3): 361-366.

Brewster, D. (1826). "On the optical illusion of the conversion of cameos into intaglios and of intaglios into cameos, with an account of other analogous phenomena." *Edinburgh Journal of Science* **4** : 99-108.

Cavanagh, P., and Leclerc, Y. G. (1989). "Shape from Shadows." *Journal of Investigative Psychology: Human Perception and Performance* **15** (1): 3-27.

Enns, J. T., Ochs, E., and Rensink, R. A. (1990). "VSearch: Macintosh software for experiments in visual search." *Behavior Research Methods, Instruments and Computers* **22** : 118-122.

Enns, J. T., and Rensink, R. A. (1990). "Influence of scene-based properties on visual search." *Science* **247** : 721-723.

Grossberg, S., and Mingolla, E. (1985). "Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading." *Psychological Review* **92** (2): 173-211.

Grossberg, S., Mingolla, E., and Ross, W. D. (1994). "A Neural Theory of Attentive Visual Search: Interactions of Boundary, Surface, Spatial and Object Representations." *Psychological Review* **101** (3): 470-489.

Hald, A. (1952). Statistical Theory with Engineering Applications, John Wiley and Sons, Inc.

Hershberger, W. (1970). "Attached-shadow orientation perceived as depth by chickens reared in an environment illuminated from below." *Journal of Comparative and Physiological Psychology* **73** (3): 407-411.

Hess, E. H., 1947 *The direction of the assumed light-source in the perception of ambiguous relief patterns.* Masters Thesis, Johns Hopkins

Hess, E. H. (1950). "Development of the chick's responses to light and shade cues of depth." *Journal of Comparative and Physiological Psychology* **1950** (43): 112-122.

Johnston, W. A., and Hawley, K. J. (1993). "Perceptual Inhibition of expected inputs: The key that opens closed minds." *Psychonomic Bulletin & Review* **1** : 56-72.

Kleffner, D. A., and Ramachandran, V. S. (1992). "On the perception of shape from shading." *Perception and Psychophysics* **52** : 18-36.

Metzger, W. (1936). Gesetze des Sehens. Frankfurt, Germany, W. Kramer.

Reichel, F. D., and Todd, J. T. (1990). "Perceived Depth Inversion of Smoothly Curved Surfaces Due to Image Orientation." *Journal of Experimental Psychology: Human Perception and Performance* **16** (3): 653-664.

Rensink, R. A., and Cavanagh, P. (1993). Processing of Shadows at Preattentive Levels. ARVO, Sarasota, FL.

Rensink, R. A., and Enns, J. T. (1995). "Preemption Effects in Visual Search: Evidence for Low-Level Grouping." *Psychological Review* **102** (1): 101-130.

Sun, J. Y., and Perona, P. (1996). "Preattentive Perception of Elementary Three-dimensional Shapes." *Vision Research* **36** (16): 2515-2529.

Treisman, A. (1982). "Perceptual grouping and attention in visual search for features and objects." *Journal of Experimental Psychology: Human Perception and Performance* **8** : 194-214.

von Fieandt, K. (1938). Ueber sehen von tiefengebilden bei wechselnder beleuchtungsrichtung., Psychological Institute of the University of Helsinki.

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

Wolfe, J. M. (1994). "Visual search in continuous, naturalistic stimuli." *Vision Research* **34** (9): 1187-1195.

Wolfe, J. M., Cave, K. R., and Franzel, S. L. (1989). "Guided search: An alternative to the feature integration model for visual search." *Journal of Experimental Psychology: Human Perception and Performance* **15** (3): 419-433.

Woodland, R., Perkins, K., Summerill, S., and Irvin, K. (1994). Strata Vision 3D.

Zelinsky, G., Hayhoe, M., and Ballard, D. (1995). "Memory for spatial position in a realistic search task." *Perception* **24** supplement : 26.

Appendix 1: Generating Scenes

Two alternate approaches exist to full-screen rendering of a visual search scene. The first is to pre-render and save many possible scenes, and the second is to render portions of scenes that can be assembled in many different ways. Selecting an approach involves a trade-off between available storage space, computational capacity, and experimental flexibility. Of these options, only off-line tile rendering of realistic scenes was feasible given the computational capacity and persistent storage available to us when the experiments were performed.

Tile rendering reduces the experimental flexibility available to the experimenter, because each tile must be constructed so that when the top of one tile is butted against the bottom of another, no seam appears. This limits target object placement to tile interiors. Furthermore, if shadows are present in a scene, then neither the object nor its shadow may cross over the boundary. Multi-tile scenes are then constructed in a manner similar to that described by Wolfe (1994a), where tiles are selected at random, butted together and displayed in a scene. The random positioning of the tiles in the scene increases the number of scenes viewed by a subject to many more than the number of tiles. If a given distractor or target tile can occur in any position in a scene any number of times, then

$$A^n P^m C_m^{(n+m)} \qquad \text{(EQ 1)}$$

represents the number of possible scenes that can be constructed, where A is the number of target absent distractor tiles from which the scene is constructed, P is the number of target present tiles, n is the number of distractor tiles in a scene, and m is the number of target present tiles in the scene. If the type of tiles selected for a display is balanced, that is no tile appears more than one more time in a display than any other tile, then the number of possible scenes is

$$\left(\prod_{f=0}^{n-1}(A-(f_{\mathrm{mod}A}))\right)\left(\prod_{g=0}^{m-1}(P-(g_{\mathrm{modP}}))\right)C_m^{(n+m)} \qquad \text{(EQ 2)}$$

where the modular arithmetic describes the balance. For these equations,

$$C_r^o = \frac{(o)!}{r!(o-r)!}. \qquad \text{(EQ 3)}$$

Thus a typical nine tile scene with eight distractors (n=8) selected from among four tiles (A=4) and one target (m=1) selected from among two tiles (P=2) could be constructed 1,179,648 possible ways in an unbalanced experiment, or 864 possible ways in a balanced experiment. Either way, it would be unlikely for a participant to see the same scene twice. Using the tiling approach only 49 kilobytes of storage is required, whereas if one were to construct 864 276x276 pixel greyscale scenes and select from among uncompressed copies of them for display, approximately 63 megabytes would be required for scene storage. For an unbalanced experiment, more than 83 gigabytes would be required!