

Visual search for direction of shading is influenced by apparent depth

DEBORAH J. AKS and JAMES T. ENNS
University of British Columbia, Vancouver, British Columbia

Recent reports of rapid visual search for some feature conjunctions suggested that preattentive vision might be sensitive to scene-based as well as to image-based features (Enns & Rensink, 1990a, 1990b). This study examined visual search for targets defined by the direction of a luminance gradient, a conjunction of luminance and relative location that often corresponds to object curvature and direction of lighting in naturalistic scenes. Experiment 1 showed that such search is influenced by several factors, including the type of gradient, the shape of the contour enclosing the gradient, and the background luminance. These factors were varied systematically in Experiment 2 in a three-dimensionality rating task and in a visual-search task. The factors combined interactively in the rating task, supporting the presence of an emergent property of three-dimensionality. In contrast, each factor contributed only additively to the speed of the visual-search task. This is inconsistent with the view that search is guided by specialized detectors for surface curvature or direction of lighting. Rather, it is in keeping with the view that search is governed by a number of "quick and dirty" processes that are implemented rapidly and in parallel across the visual field.

Conventional theories of preattentive vision claim that simple features such as size, orientation, luminance, and motion are registered automatically and in parallel, whereas the serial spotlight of attention is required to detect conjunctions of these features (Beck, Prazdny, & Rosenfeld, 1983; Julesz, 1984; Treisman, 1986). This claim is based on data from visual-search and texture-segmentation tasks: preattentive sensitivity is implicated by visual-search rates that are relatively independent of display size (i.e., less than 10 msec per item) and by texture boundaries that are perceived spontaneously (i.e., within 50–100 msec of display onset).

Recently, there have been several reports of very rapid search and/or texture segmentation based on complex conjunctions of these features. Ramachandran (1988) showed that it is possible to segment a texture on the basis of the direction of the luminance gradients within circular texture elements; Nakayama and Silverman (1986) demonstrated rapid search for conjunctions of binocular disparity with motion and color; McLeod, Driver, and Crisp (1988) reported similar results for the conjunction of motion with shape; and Enns and Rensink (1990a, 1990b, 1991a) showed that rapid search was possible for spatial relations among lines and shaded polygons.

How should these results be interpreted? Enns and Rensink (1991a, in press) suggest a framework that distinguishes between features of objects in three-dimensional

(3D) space (i.e., properties of the *scene*) and features of the two-dimensional array of light that is projected from the scene to a visual system (i.e., properties of the *image*). Their analysis suggests that preattentive vision may not only be sensitive to simple geometric properties of the image, but may also be able to recover some properties of the scene. For example, rapid segmentation of textures based on the direction of luminance gradients may reflect preattentive sensitivity to 3D curvature or to the direction of light in the scene (Ramachandran, 1988); rapid search for conjunctions of features with binocular disparity and motion may reflect sensitivity to apparent depth (McLeod et al., 1988; Nakayama & Silverman, 1986); and rapid search for spatial relations among lines and shaded polygons may reflect sensitivity to 3D orientation and the direction of lighting (Enns & Rensink, 1990a, 1990b, 1991a, in press).

The kinds of scene properties that are recoverable in this way, however, are limited by the function and characteristics of preattentive vision. These include the need for processing to be *very rapid* (the output of preattentive vision must be available in 50–100 msec if it is to be useful in guiding eye movements, immediate actions, or attentive processes), *spatially parallel* (local operations are the only kind that can be carried out across the visual field within the available time), and *environmentally relevant* (considerable efficiency can be gained by restricting the interpretations made in each local region to those relevant to the larger goals of the visual system).

This characterization of preattentive vision is called "quick and dirty" by Enns and Rensink (1991a, in press) to emphasize the inherent tradeoff between the speed of a particular visual computation and its validity. Perfect validity (i.e., a correct match between interpretation and

This research was supported by an NSERC (Canada) grant to J.T.E. We are grateful to Ron Rensink, Steve Taylor, and Shuji Mori for comments on earlier drafts of this paper and to Lester Krueger for critical suggestions. Please address correspondence to J. T. Enns, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC, Canada V6T 1Z4 (e-mail: james_enns@mtns.ubc.ca).

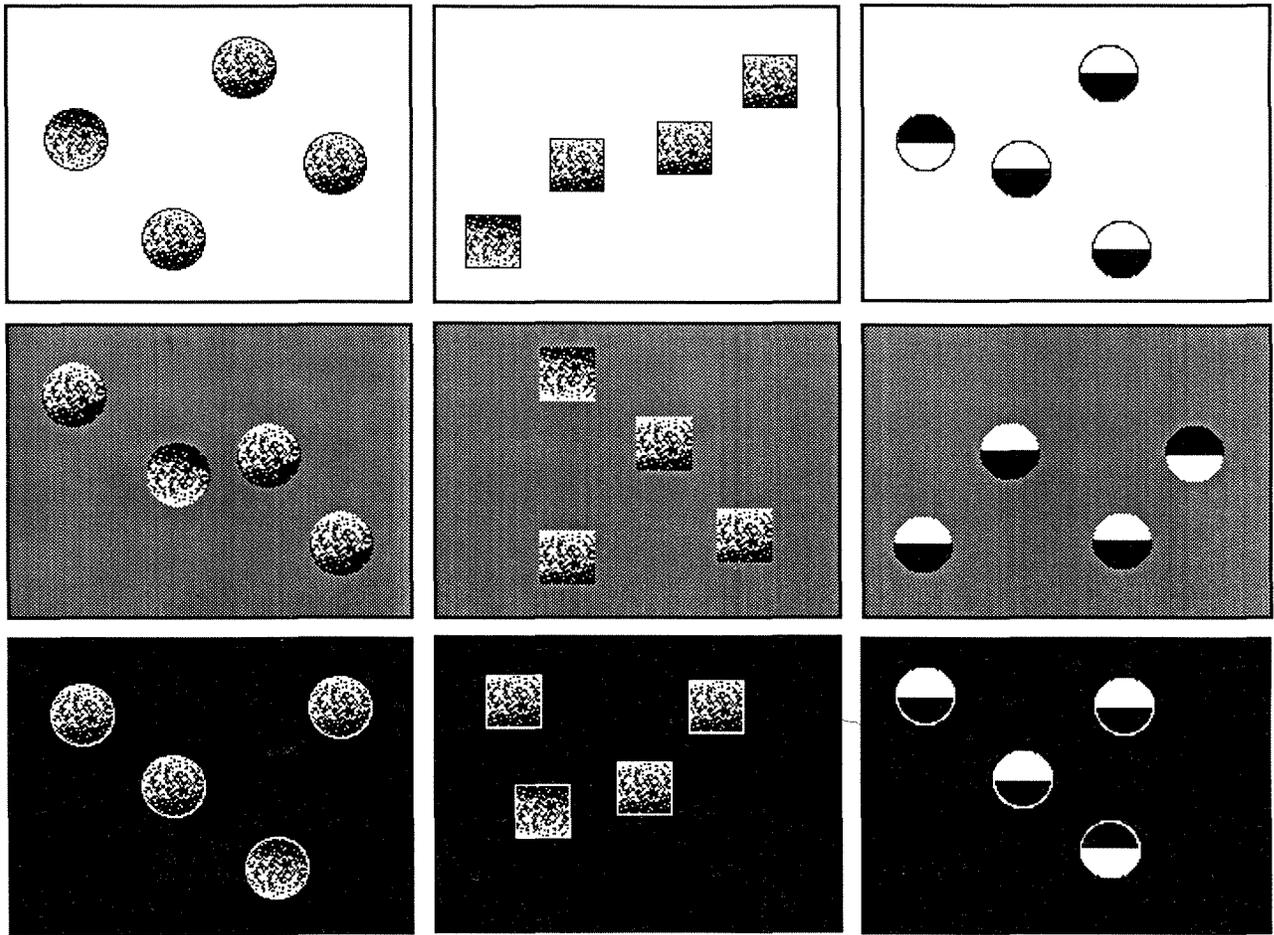


Figure 1. Illustrations of the visual-search displays in Experiment 1. Three different search items (linear-circle, linear-square, step-circle) were examined against three backgrounds (white, gray, black). The target in each display is the item with the luminance gradient of dark to light, moving from top to bottom. Actual displays in the experiment consisted of 2, 8, or 14 items.

image) of computation can only be gained by sacrificing speed and thereby losing the advantage of preattentive processing. Conversely, a “best guess” can be made quickly by a preattentive process, but only at the risk of being wrong some of the time. This can be seen, for example, in that search is easy for line drawings of objects with right-angled corners, but not when the drawings represent corners that are more acute or obtuse (Enns & Rensink, 1991a, in press). Preattentive vision must base its best guesses on assumptions that work much of the time.

In the present study, we used visual search to examine the sensitivity of preattentive vision to gradients of luminance. As pointed out earlier, luminance gradients are conjunctions of luminance and relative location when defined in terms of the image. However, in terms of the scene, they often correspond to simple features of object curvature or direction of lighting. Therefore, our first goal was to determine the conditions under which these gradients could guide rapid search. What does preattentive

vision “know” about the 3D world that lies behind shading in an image?

To illustrate how three different kinds of scene information can be conveyed by luminance gradients, consider a contrived scene in which a light is directed onto objects of constant curvature and matte reflectance, viewed against a background surface of similar reflectance (see middle panel of Figure 1). The 3D shape of the object is signaled in part by the *type of gradient*: a curved object will give rise to a smooth gradient since there is continuous variation in the angle with which the light strikes the surface. An abrupt gradient is more consistent with a sharp edge than with a smoothly curved surface. Surface curvature is also signaled in part by the *outlining contour* of the gradient: a circular outline is consistent with a smoothly curved object, regardless of the vantage point from which it is viewed. A square outline is consistent with some curved objects (e.g., a cylinder), but only from a unique vantage point. Finally, figure-ground segmentation is sig-

naled by the relation between the gradient and the *background*: curved objects will have some regions that are brighter than the background (the top of the sphere is perpendicular to the light source) and some that are darker (the bottom is occluded from direct light).

This example is admittedly artificial, but it illustrates an important point. In any visual environment, some image properties are highly correlated with particular scene properties. If a visual system is able to solve the image-to-scene correspondence problem, it must take these correlations into account at some level. Which of these correlations are already assumed at preattentive levels?

EXPERIMENT 1

In the first experiment, we looked at whether search was affected by the *type of gradient* (linear and step) and by the *outlining contour* (circle and square). Although these factors are simplifications of properties in the natural world, they are sufficiently predictive to allow pictures such as Figure 1 to be readily interpreted. In systematic studies of these relations, subjects interpret linearly shaded circles with light tops as convex (i.e., protruding) solids in a scene with an overhead light source (Benson & Yonas, 1973; Mingolla & Todd, 1986; Rock, 1983; Todd & Mingolla, 1983). With the opposite direction of shading, the impression is either of a concave, top-lit object or of a convex, bottom-lit object. Although there are no similar data for linearly shaded squares or step-circles, these items are clearly not as easily interpreted as 3D objects (see Figure 1 and Experiment 2A). Thus, if apparent three-dimensionality is an important condition for the detection of gradient direction, search should be fastest for the linearly shaded circle.

A third factor examined was the luminance of the *background* against which items were presented (white, gray, black). Relative background luminance is known to have a strong influence on texture segregation (Beck et al, 1983; Beck, Sutter, & Ivry, 1987). Specifically, elements differing in luminance segregate spontaneously when the background luminance is intermediate to them (i.e., when the two elements are reversed in their contrast polarity relative to the background). The same elements segregate only with effort when the background is lighter or darker than all the elements (i.e., the two elements differ from one another by the same amount, but do not differ in sign). In the displays used in this study, a contrast-polarity reversal occurred within each of the items in the gray-background condition (i.e., the luminance of the item changed from positive to negative along the vertical dimension). Therefore, we expected search rates to be faster in this condition than in the white and black backgrounds, where the same change in item luminance did not cross the background luminance value.

Method

Subjects. Seven members of the University community participated in three 30-min sessions to complete four blocks of 60 trials per condition. These subjects were experienced in visual-search

tasks, but did not have specific experience with the targets and distractors used here.

Stimuli and Procedure. Display presentation and data collection were controlled by a Macintosh Plus computer (Enns, Ochs, & Rensink, 1990). The display backgrounds used, shown in Figure 1, consisted of all pixels lit (*white*), alternate pixels lit (*gray*), and no pixels lit (*black*). The luminance, color temperature, and CIE coordinates of these backgrounds were 159 cd/m², 50,000 K, and .25, .24 for white; 58 cd/m², 18,500 K, and .25, .32 for gray; and 12 cd/m², 3,500 K, and .40, .38 for black. The linear luminance gradient consisted of randomly selected black and white pixels, ranging from 5% black at one extreme to 88% black at the other. The step gradient shifted abruptly from white to black along the horizontal meridian of the items. In each condition, the items were a combination of one outline contour (circle, square) with one shading gradient (linear, step). Target items differed from distractor items by a 180° rotation in the image plane.¹

Items were distributed randomly on an imaginary 4 × 6 grid subtending 14° × 19° of visual angle. Each item subtended 1.4° and was randomly jittered in its grid location by ±0.5° to prevent judgments based on item collinearity. The target was present on a random one half of the displays, which contained a total of 2, 8, or 14 items.

The subjects sat 50–60 cm from the screen. Each trial began with a fixation symbol lit for 750 msec, followed by the display, which remained visible until the subject responded. Target presence and absence were reported by pressing one of two response keys, and the subjects were permitted to make their own response-finger assignments. Accuracy feedback (plus or minus sign) was displayed at the center of the screen after each response. The subjects were instructed to maintain fixation throughout the trial sequence, to respond as rapidly as possible, and to keep errors below 10%.

Results

The means and standard errors of the correct search times are shown in Figure 2, along with mean percentage correct. There were fewer than 10% errors overall, and errors differed by no more than 2% between conditions. A nonsignificant correlation between the 54 pairs of reaction times (RTs) and mean percentage of errors in Figure 2 confirmed that, overall, speed was not being traded for accuracy [$r(52) = .06$]. Regression analysis indicated that mean RT and display size were linearly related [mean $r(1) = .94$], as were mean percentage of errors and display size [mean $r(1) = .81$].

Although search rates were surprisingly rapid in all conditions (mean RT slopes across conditions ranged from 3–12 msec per item for target-present trials and 5–17 msec per item for target-absent trials), there were reliable differences between conditions. Search rates were rapid for the linear-circles under all background conditions (mean search rate = 4 and 5 msec per item in the white condition, 3 and 8 msec per item in the gray condition, and 6 and 5 msec per item in the black condition) and did not differ from one another (Fisher's *LSD* tests were not significant for any pair of these rates [$p > .20$]). Search was also very fast for all items against the gray background (mean search rate = 3 and 8 msec per item for the linear-circle, 3 and 5 msec per item for the linear-square, and 5 and 8 msec per item for the step-circle) and also did not differ reliably between items (Fisher's *LSD* tests were not significant for any pair of these search rates [$p > .20$]). In contrast, search rates were slower

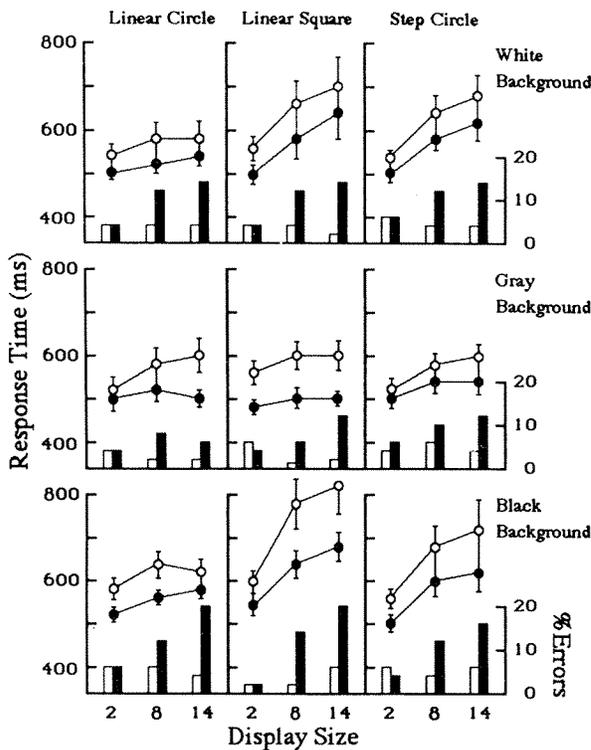


Figure 2. Mean correct reaction times and percentage of errors for the nine search conditions in Experiment 1. Error bars are SEMs. Three different search items (linear-circle, linear-square, step-circle) were examined against three background conditions (white, gray, black). Filled circles and bars represent data from target-present trials; open circles and bars represent target-absent trials.

by a factor of two for the linear-square (mean search rate = 12 and 13 msec per item in the white condition, 12 and 17 msec per item in the black condition) and the step-circle (mean search rate = 8 and 10 msec per item in the white condition, 9 and 15 msec per item in the black condition) under both black and white background conditions (Fisher's *LSD* tests between these conditions and all others were reliable for both target-present and target-absent trials [$p < .01$]).

These comparisons were supported by a factorial analysis of variance (ANOVA) on the correct RT slope data. The analysis revealed significant main effects for items (linear-circle, linear-square, step-circle) [$F(2,12) = 20.50, p < .05$], trial block [$F(2,12) = 4.23, p < .02$], and present versus absent trials [$F(1,6) = 6.65, p < .05$]. The only reliable interaction was items \times background [$F(4,24) = 3.44, p < .05$], reflecting the speed-up in search for linear-circles in all backgrounds and for all items against the gray background.

Discussion

This experiment demonstrated that the visual search rate for a 180° difference in the direction of shading was not

the same in all contexts. In particular, very rapid search was possible under two conditions. The first occurred when a circular outline was added to the linear shading gradient. The impression of three-dimensionality given by this conjunction seemed to allow subjects to search easily for objects that differed in one or more scene properties. The obvious candidates for these properties, based on previous research, include the perceived direction of lighting (i.e., bottom lighting for the target vs. top lighting for the distractors) and surface curvature (i.e., concavity for the target vs. convexity for the distractors) (Benson & Yonas, 1973; Mingolla & Todd, 1986; Rock, 1983; Todd & Mingolla, 1983).

In the second case, search was rapid when there was a contrast-polarity reversal within the search items. The combination of direction of shading with a contrast-polarity reversal allowed for rapid search even in the absence of a linear shading gradient or a circular outline (see the gray background condition in Figure 2 and in Experiment 2B). This suggests that the search items themselves do not have to have a strong 3D interpretation for rapid search to occur. However, it also does not eliminate entirely a role for 3D mechanisms. As pointed out in the introduction, contrast polarity reversal may itself be a signal for relative differences in depth.

Beyond these contextual influences on search for the direction of shading, we were surprised that search was not slower for the linear-square and step-circle (e.g., search rates of 30 msec or more per item). It is possible that slower rates were not observed because these items do have 3D interpretations, albeit atypical ones. The linear-square is consistent with a cylinder whose ends are accidentally aligned with the viewpoint, whereas the step-circle can be interpreted as a sphere lit with an overhead spotlight, thereby causing an attached shadow. It is also possible that the subjects differed in the extent to which they were aware of these interpretations and/or that these items simply produced weaker preattentive signals for three-dimensionality. The latter view is supported by other findings that search is slowed by unconventional views of objects (Enns & Rensink, 1991a, in press). We will explore this relation between preattentive registration and the subjective three-dimensionality of search items in Experiment 2.

Finally, it is worth noting that the relatively rapid search rates found for step-circles are at odds with reports of texture segmentation of these items. Ramachandran (1988), Malik and Perona (1990), and J. Beck (personal communication, August 24, 1990) have studied these items in texture displays and found that they provide only a weak signal for both region segregation and population grouping. We think there is a strong possibility that different rules govern the edge-finding processes of texture-grouping tasks and the feature-detection processes of visual-search tasks, just as different rules appear to govern region segregation and population grouping (Beck, Graham, & Sutter, 1991). We are now investigating this possibility in a series of experiments that build on the present paper.

EXPERIMENT 2

There are at least two different classes of explanation for the finding that preattentive vision combines information from a variety of sources when interpreting luminance gradients. One is that specialized mechanisms exist quite early in the visual stream for the detection of simple volumes (i.e., spheres) and/or their properties (e.g., curvature or direction of lighting). Such mechanisms could have evolved because of the importance of these scene properties for the survival of locomoting visual organisms (Gibson, 1966; Ramachandran, 1985).

The idea of specialized detectors and corresponding trigger features can be traced to Barlow (1953). Following this tradition, several recent theories of object perception begin with the assumption that the lowest order elements of object perception are 3D or volumetric solids that are registered automatically by specialized mechanisms (Biederman, 1987; Buffart, Leeuwenberg, & Restle, 1981; Leeuwenberg, 1988; Pentland, 1986). For example, Biederman's recognition-by-components theory, which is the most thoroughly developed of these theories, claims that image regions are first assigned to one of 36 volumetric primitives or *geons*. Empirical support for this theory comes largely from speeded object-naming tasks that examine the effects of geon number and line deletion on naming times. The present results can be accommodated in this view by proposing that a "spheroid detector" is activated by the conjunction of appropriate luminance gradient, outlining contour, and/or contrast polarity.

An alternative explanation is that preattentive vision combines information from a number of sources in a "quick and dirty" way (Enns & Rensink, 1991a, in press). In this view, there could be a large number of rapid and spatially parallel processes that make best guesses about the scene based on information in the image. These processes are not necessarily tuned to specific scene properties themselves. However, they may be able to signal important scene properties collectively and stochastically. For example, one process might examine the luminance gradients in the image: a smooth gradient would be interpreted as a curved surface, whereas an abrupt gradient would be interpreted as an edge. A second process might be involved in object-boundary formation: a circular border would be simpler to compute by any number of theories than would a square border (Attneave, 1954, 1967; Kellman & Shipley, 1991; Leeuwenberg, 1971). A third process might examine the local relations of contrast polarity: consistent relations everywhere would correspond to a uniform field of objects, whereas a reversal would signal an object standing out from its neighbors. Taken together, these processes would be able to signal the presence of important scene properties such as surface shape and the direction of lighting. However, note that the processes themselves would not have to "know" anything about surface curvature or the rules of lighting.

One way to distinguish between these two explanations is to consider their predictions for a visual-search exper-

iment in which stimulus factors are combined orthogonally. Sternberg's (1969a, 1969b) additive factors method (AFM) is a framework that has been developed to analyze results from such an experiment. Before describing the experiment, we will briefly review the assumptions of AFM that are relevant to this kind of an analysis (see also Taylor, 1976).

First, it is assumed that successive and independent stages of processing intervene between the presentation of a search display and the subject's response. The relations between proposed stages of processing are established by selectively varying their temporal durations. In our study, the duration of the visual-search task was manipulated by varying the type of luminance gradient (to influence a potential luminance-gradient analysis), the background luminance (to influence a potential contrast-polarity analysis), and the shape of the outlining contour (to influence a potential object-boundary analysis), as well as the usual visual-search variables of display size and target presence versus absence (to influence the number of items that needed to be inspected).

Second, if orthogonal variation in the difficulty of two factors leads to an additive pattern of performance, then the existence of two successive and independent stages of processing is implied. If, on the other hand, factor variation results in an interactive pattern, then a common stage is implied. For example, additive relations between gradient type and background would imply two independent stages of influence for visual search: luminance-gradient and contrast-polarity analyses. Interactive relations between these factors would imply a common source of influence, that is, a stage in which these two analyses are combined.

Third, the analysis of performance must be based on a dependent variable for which interval measurement properties can be assumed. If only ordinal measurement is assumed, then the discrimination between additivity versus interaction becomes very difficult (Loftus, 1978). The primary measure in a visual-search experiment is RT, a measure that satisfies the equal-interval requirement. Note that this assumption precludes analyses based on nonlinear transformations of RT, such as percentage change RT or logarithmic RT, which violate the inherent additivity of a real-time scale (see Sternberg, 1969a, 1969b, for a complete discussion of this issue).

Finally, we note that AFM is not without its critics and competitors (e.g., Eriksen & Schultz, 1979; McClelland, 1979; Turvey, 1973) and that our experiment is not designed to compare AFM directly with alternative models. Nevertheless, these models all agree on the interpretation of an additive performance pattern: they all use this as diagnostic of separate stages. Where the models differ is in their interpretation of interactions. For example, some models predict interactions from separate processing stages that overlap in time (McClelland, 1979). To test hypotheses that distinguish among these models, it would be necessary to perform more stringent tests than those used here (e.g., varying interstimulus interval [ISI] or

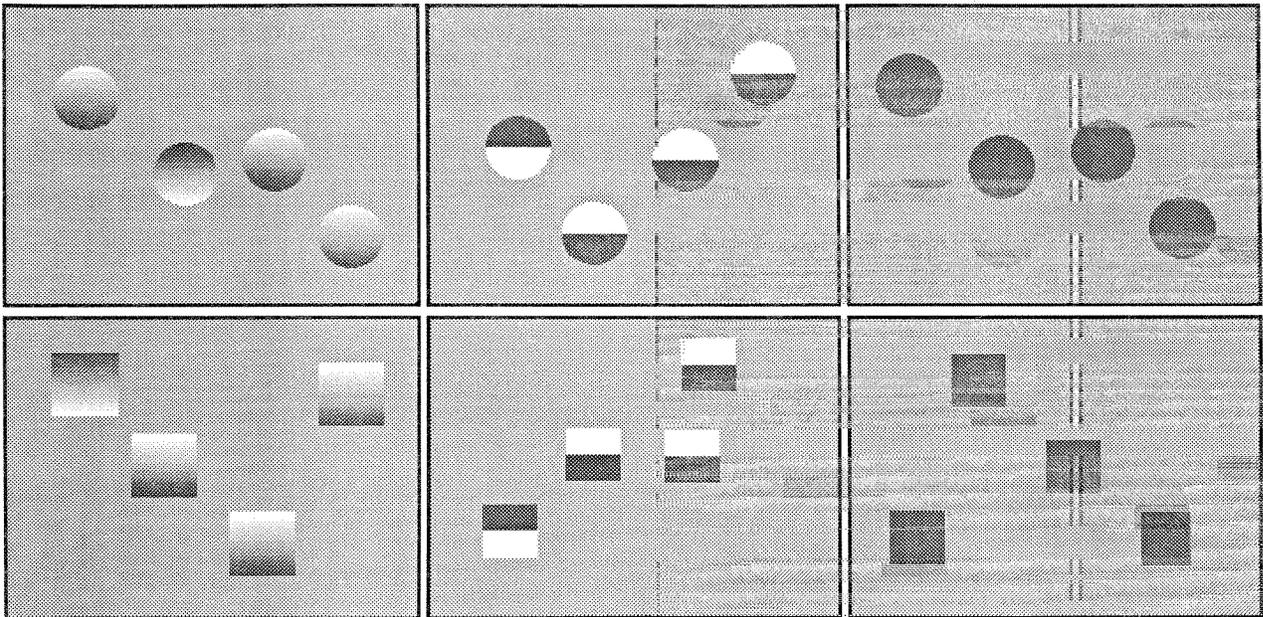


Figure 3. The six display types used in the three-dimensional rating task in Experiment 2A. Items were formed by combining three types of luminance gradient (linear, step, none) with two types of outlining contour (circle, square). Although white and gray backgrounds were tested, only the gray condition is shown.

stimulus onset asynchrony [SOA] in a brief exposure paradigm).

Our use of AFM, therefore, implies that an interactive (i.e., multiplicative) effect of the three stimulus factors on search rates is consistent with the presence of a common processing stage, perhaps a specialized detector for a spherical object or a curved surface. Search items containing feature conjunctions consistent with such a mechanism would result in rapid search (e.g., a linearly shaded circle against a gray background could excite a "sphere detector"), whereas items composed of feature conjunctions inconsistent with the detector would result in slow, attention-demanding search (e.g., a step-square against a white background might excite an oriented edge detector but not a volume detector). An additive pattern of search rates would be consistent with the operation of several "quick and dirty" processes that each make a best guess for their stimulus property on the basis of information available in the image. A detection decision in the visual-search task would simply involve pooling the information from these independent processes. To the extent that decisions could be made quickly for each putative process, there should be a corresponding increase in search speed.

To test these hypotheses, we examined each of the three stimulus factors at two levels: type of luminance gradient (linear vs. step), background luminance (gray vs. white), and shape of outlining contour (circle vs. square). This meant that a *step-square* item had to be added to the set of search items in Experiment 1 (see Figure 3) and that the black background used in Experiment 1 was no longer

needed. In addition, an improved way of rendering the displays was used in Experiment 2, made possible by an upgrade in the computer program used to run the experiments (Enns & Rensink, 1991b). Luminance gradients now varied linearly in 256 gray-level steps, rather than being simulated with a dithering technique.

We also ran a preliminary experiment to examine our assumption that the linear-square, step-circle, and step-square items were subjectively less 3D in their appearance than was the linear-circle (Experiment 2A). This was important both because previous studies of the subjective interpretations had focused exclusively on the linearly shaded circles (Benson & Yonas, 1973; Mingolla & Todd, 1986; Rock, 1983; Todd & Mingolla, 1983) and because there had been some doubt in our minds about the interpretation that the subjects had given to these items in Experiment 1.

Experiment 2A: Three-Dimensionality Ratings

This experiment served two purposes. First, it examined the extent to which the various items used in the visual-search tasks suggested a 3D interpretation under free-viewing conditions. Second, because of the orthogonal combination of stimulus factors, it allowed us to see whether the rules for combining features in a subjective rating task were the same as those used in the visual-search task.

Method

Subjects. Twenty university students participated in the three-dimensionality rating task. All were unpaid volunteers, and none

had previous experience with these stimuli in a visual-search task or any other psychophysical task.

Stimuli and Procedure. The rated stimulus displays were constructed from an orthogonal variation of gradient type (linear, step, none), outlining shape (circle, square), and background luminance (gray, white); they are shown in Figure 3. Displays contained a total of 2 or 14 items. The homogeneous gray items (no gradient) were included to help anchor the subjects' rating for the absence of three-dimensionality. They contained no target-distractor differences and therefore could not be used in the visual-search task. All displays were 8.5×6.5 in. and drawn with a laser printer.

Luminance was equated across the items at the most extreme dark region for linear, step, and no gradients (Munsell coordinates of 4.3, 4.3, 4.3; reflectance = 14.1%) and at the most extreme light region (Munsell coordinates of 7.4, 7.4, 7.5; reflectance = 48.7%). The items were presented either on a white background (Munsell coordinates of 9.0, 8.9, 9.0; reflectance = 77.4%) or on a gray background (Munsell coordinates of 5.6, 5.6, 5.7; reflectance = 26.1%). All regions had CIE coordinates of .3, .3.

The subject first sorted through a random ordering of the entire set of displays to select the one whose items appeared most "two-dimensional, flat, or picture-like." A second display was then chosen that appeared most "three-dimensional, solid, or object-like." These two displays were assigned ratings of 0 and 10, respectively. The subject then examined each display in turn, assigning values between 0 and 10 to each, according to its apparent three-dimensionality. Two ratings were made for each of the 24 display conditions, with two different displays shown in each condition.

Results

The mean 3D ratings are shown in Figure 4. Ratings were significantly higher for linear gradients than for step or no gradients [$F(2,38) = 113.20, p < .001$], for circular than for square outlining shapes [$F(1,19) = 33.57, p < .001$], for gray than for white backgrounds [$F(1,19) = 108.92, p < .001$], and for large display sizes (mean rating = 4.3) than for small display sizes (mean rating = 3.6) [$F(1,19) = 29.40, p < .001$].

However, there were also significant interactions of shading × outline [$F(2,38) = 12.66, p < .001$], shading × background [$F(2,38) = 44.83, p < .001$], shading × display size [$F(2,38) = 6.63, p < .01$], and shape

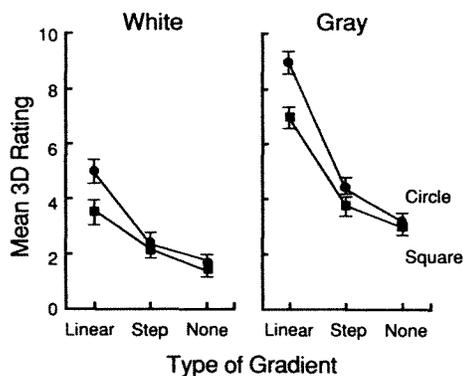


Figure 4. Mean three-dimensional ratings in Experiment 2A, averaged across display sizes of 2 and 14. Ratings were made for two background conditions (white, gray), three types of gradient (none, step, linear), and two outlining contours (circle, square).

× display size [$F(1,19) = 8.99, p < .01$]. The first two interactions indicate that linearly shaded items were judged as more 3D in the context of a circular contour and against a gray background than could be predicted from the individual effects of these factors. In other words, there was an emergent property evident in the subjects' ratings of the three-dimensionality of these displays. The relation between these results and search data will be discussed following Experiment 2B.

Experiment 2B: Visual Search

This experiment examined whether search for luminance gradients was influenced in an additive or interactive fashion by the factors of gradient type, outlining contour, and background luminance. If search is influenced by an emergent property that results from the combination of these factors, then search rates should be a multiplicative function of these factors. On the other hand, if search is determined only by the component factors, then search rates will be an additive function of these factors. Before testing this hypothesis, however, we made sure that the results we had obtained in Experiment 1 with the black-and-white dithered form of shading generalized to the gray-scale form of shading used in Experiment 2.

Method

Subjects. A total of 22 university students participated in two 1-h sessions to complete four blocks of 60 trials in each of four conditions (circle on gray, circle on white, square on gray, square on white). Twelve of the subjects searched for items with linear shading, and 10 subjects searched for items with step shading. The subjects were divided on this factor because the necessary participation time would otherwise have exceeded the subject-pool guidelines. Half of the subjects in each group were experienced in visual-search tasks but had no specific training with these search items. The other half had no prior experience with RT testing. All but three volunteers received course credit or payment.

Stimuli and Procedure. The experiment was run on a Macintosh II computer with a 13-in. high-resolution Apple RGB monitor (Enns & Rensink, 1991b). The luminance of the white background was 98.3 cd/m² and the gray background was 38.7 cd/m². The linear and step luminance gradients each consisted of equal luminance values (72.5 cd/m² and 18.1 cd/m² at the extremes). Thus, both backgrounds were approximately equal in their difference from the extreme values in the items. All luminance values had the same CIE coordinates of .28, .28 and a color temperature of 10,050 K. Instructions to the subject, display dimensions, display sizes, and all other procedural details were otherwise identical to those in Experiment 1.

Results

The results in this experiment were examined in two ways. First, the mean correct RTs were analyzed for the influence of the three stimulus factors, along with the usual visual-search manipulations of display size and target presence. Second, the mean RT slope was examined as a function of the same stimulus factors. This measure allows us to speak directly of search rates rather than referring to them indirectly via the interaction between display size and various stimulus factors. The reader should note that the choice of dependent measure does not affect any of

Table 1
Mean Correct Reaction Times and Percentage of Errors
in Experiment 2B

Trial Type	Display Size	Circle				Square			
		Gray		White		Gray		White	
		<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>
Linear Gradient: Reaction Times									
Present	2	474	12	560	22	544	14	616	20
	8	490	13	620	27	618	21	729	27
	14	495	14	663	30	702	30	862	45
Absent	2	543	18	624	27	618	21	658	19
	8	587	25	715	43	781	30	935	50
	14	631	33	797	52	917	59	1,147	71
Linear Gradient: Percentage of Errors									
Present	2	3.3	0.9	2.5	0.8	2.3	0.8	3.5	0.8
	8	3.5	0.9	5.0	1.1	7.5	1.3	13.3	1.8
	14	4.8	1.2	7.7	1.3	9.1	1.3	12.7	1.9
Absent	2	4.2	1.1	2.9	1.0	1.9	0.6	1.9	0.7
	8	2.9	0.9	1.9	0.6	2.1	0.7	1.9	0.8
	14	1.9	0.6	2.9	0.9	2.3	0.7	2.9	0.9
Step Gradient: Response Times									
Present	2	540	31	593	25	554	23	634	36
	8	611	37	757	48	627	29	814	50
	14	654	41	792	50	696	42	921	70
Absent	2	575	31	634	37	582	27	661	28
	8	714	63	841	76	731	65	980	73
	14	833	86	1,050	113	933	117	1,257	107
Step Gradient: Percentage of Errors									
Present	2	4.8	1.4	2.8	1.2	4.1	2.6	3.7	1.9
	8	6.0	2.0	6.2	2.0	4.5	1.9	9.5	2.6
	14	5.5	1.5	9.2	2.6	6.8	2.4	13.8	2.9
Absent	2	4.8	1.9	2.8	0.8	3.4	2.1	1.6	0.6
	8	1.6	0.9	1.6	0.8	1.4	0.8	0.4	0.3
	14	1.9	1.1	2.6	0.8	1.9	1.5	3.0	1.5

Note—Stimulus factors included trial type (present, absent), display size (2, 8, 14), type of gradient (linear, step), outlining contour (circle, square), and background (gray, white).

the conclusions that are reached. This is because the first analysis shows that all the effects of interest involve display size and because, in all conditions, RT increases linearly as a function of display size.

The mean correct RT and percentage of errors are shown in Table 1. There were fewer than 5% errors overall, and they differed no more than 1% between conditions. A small positive correlation between the mean RTs and mean percentage of errors in Table 1 showed that speed was not being traded for accuracy [$r(46) = .22$]. Regression analyses indicated that mean RT and display size were linearly related [mean $r(1) = .85$ for step gradients and $r(1) = .74$ for linear gradients] and that mean percentage of errors and display size were related, albeit less strongly [mean $r(1) = .32$ for step gradients and $r(1) = .26$ for linear gradients].

The mean RTs in the visual-search task correlated very well with the mean 3D ratings in Experiment 2A [$r(10) = .91$]. This indicated that the same factors that led to impressions of three-dimensionality in a free-viewing task also resulted in faster search rates. However, the most

striking finding was the strong effect of all three stimulus factors on search rates. A mixed-design ANOVA of the mean search times revealed significant interactions of display size with all three of the stimulus factors, display size \times shading [$F(2,40) = 3.72, p < .05$], display size \times outline [$F(2,40) = 11.48, p < .001$], and display size \times background [$F(2,40) = 20.05, p < .001$]. However, none of the three-way or higher order interactions involving these factors even approached significance (all $ps > .20$).

In addition to these effects, the ANOVA revealed significant main effects for all within-subject factors, outline [$F(1,20) = 8.57, p < .01$], background [$F(1,20) = 18.78, p < .01$], display size [$F(2,40) = 91.73, p < .001$], and target presence [$F(1,20) = 75.67, p < .001$]. The between-group factor of shading was not reliable as a main effect [$F(1,20) = 1.67$]. The remaining significant effects involved the expected interactions of display size \times target presence [$F(2,40) = 40.04, p < .001$], which reflected larger display-size effects on target-absent trials than on target-present trials, and display size \times tar-

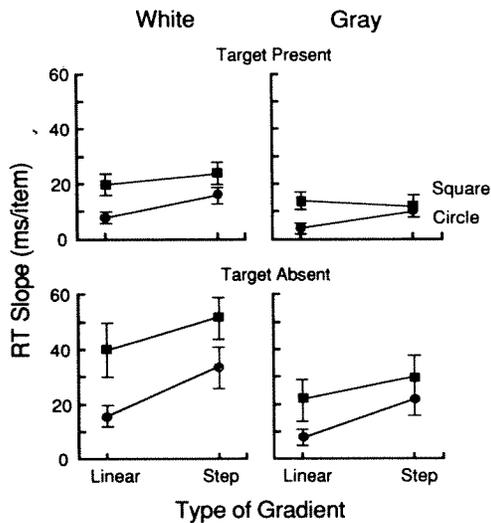


Figure 5. Mean search rates (reaction time slopes) in Experiment 2B, shown separately for target-present (top panel) and target-absent (bottom panel) trials. Search performance was examined against two background conditions (white, gray), two types of gradient (linear, step), and two outlining contours (circle, square). Error bars are SEMs.

get presence \times shading [$F(2,40) = 3.34, p < .05$], display size \times target presence \times outline [$F(2,40) = 5.39, p < .01$], and display size \times target presence \times background [$F(2,40) = 4.14, p < .05$]. The latter three interactions indicated that each of the three stimulus factors also had a larger influence on target-absent trials than on target-present trials.

The results of the RT slope analysis are shown in Figure 5. Each of the main effects, shading [$F(1,20) = 4.44, p < .05$], outline [$F(1,20) = 11.46, p < .01$], and background [$F(1,20) = 19.69, p < .001$], was significant, and no interactions between these factors even approached significance ($p > .20$). Furthermore, each of the factors influenced search rate by approximately the same amount, a factor of two.

Of the three possible two-way interactions among the stimulus factors, the only one that even hinted at a possible underlying effect was that of outline \times background [$F(1,20) = 1.16, p < .30$]. The advantage of a circular outline over a square outline was 15 msec per item against the white background, but only 10 msec per item against the gray background. We do not take this trend too seriously, both because it is statistically unreliable and because it may be an artifact of a "floor" effect in the gray-background-target-present condition (see Figure 5).

As would be expected by almost all accounts, target-absent trials had slower search rates than did target-present trials [$F(1,20) = 55.50, p < .001$], and this factor interacted with outline [$F(1,20) = 4.76, p < .05$] and background [$F(1,20) = 5.85, p < .05$]. However, these ef-

fects were of no consequence in the present discussion, so they were not examined further.

Discussion

The results of Experiment 2 indicate that all three factors—type of shading, outline shape, and background luminance—had a strong influence on the perceived three-dimensionality of items in the rating task and on the speed of discrimination in the visual-search task. This suggests that search is indeed influenced by pictorial cues to depth. But in what way does this influence come about?

We think an important clue to this question can be seen in the different ways in which the stimulus factors were combined in the two tasks. Although the three factors contributed interactively to perceived three-dimensionality in the rating task, they contributed only additively to search rates in the visual-search task. What does an interactive versus an additive pattern signify?

The interaction of factors in the rating task is consistent with the experience observers report when they first see these displays: linearly shaded circles give a compelling 3D impression, whereas the other items give a considerably weaker impression of depth. We believe most researchers, including us, have interpreted performance on visual-search and texture-segmentation tasks under the mistaken assumption that the emergent properties we see under casual viewing conditions have an important influence in these tasks.

The visual-search data, however, show that this interpretation is not necessary. Speeded decisions based on the same three factors showed no evidence of an emergent property or a common stage of processing. The additive pattern of results indicates that search rates (RT slopes) could be predicted directly from the separate component factors of outline shape, type of shading, and background luminance. No specialized detectors for 3D objects or their properties need to be invoked. Instead, search appears to be based on separate stages of processing that precede those in which surface curvature and the direction of lighting are explicitly represented. This account of how pictorial 3D cues are combined in visual search suggests that preattentive vision may not look for 3D objects or properties so much as it looks for good predictors of these in the image.

The only evidence for a common stage of processing occurred when the relations between display size and each of the stimulus factors were examined separately: display size interacted significantly with outline shape, type of shading, and background luminance. Finding interactions between these variables suggests that the stimulus factors slow search in essentially the same way as does increasing the total number of search items. One way to think of this is that search rate is determined by the amount of time required to process a single item. Thus, the total search time for a given display can be increased, either by increasing the number of items to be inspected or by

increasing the difficulty of processing each individual item.

GENERAL DISCUSSION

The experiments reported here indicate that rapid search for a 180° difference in the direction of shading is influenced by a number of factors. Specifically, search is faster when the luminance gradients are smooth rather than abrupt, when the gradients appear in circular outlines rather than in square shapes, and when the gradients involve a reversal in contrast polarity relative to the background luminance.

Each of these image factors is related in an interesting way to the ecological relevance and/or computational complexity of the corresponding scene properties. For instance, smooth image gradients correspond to continuous changes in surface orientation. With the assumptions of an overhead light source and uniform surface reflectance, the darker part of the gradient can readily be assigned to a portion of a surface that has a different orientation than that corresponding to the lighter region. In contrast, abrupt gradients often correspond to surface edges. However, there is little to constrain the depth interpretation of the dark versus light portions of an edge. A similar analysis can be made for outlining contours. Circular shapes in the image are related more robustly to curved solids in the scene than are square shapes (i.e., an image square corresponds to a curved surface only if it is a cylinder being viewed from a unique perspective). Circular shapes are also easier to compute (Attneave, 1954, 1967; Kellman & Shipley, 1991; Leeuwenberg, 1971). Finally, a reversal in contrast polarity is a very good signal for figure-ground segmentation. Although a gradient that is entirely above or below the luminance of the background is a clear signal for a surface discontinuity in either reflectance or depth, a reversal in contrast polarity across the gradient strongly suggests that an *object* lies in front of the surface. A solid object in front of a surface of similar reflectance will project some regions in the image that are brighter than the background (e.g., the top, because it lies in front of the background) and some regions that are darker (e.g., the bottom, because light is occluded by the upper parts of the object).

Given the sensitivity of preattentive vision to these contextual factors, it was interesting to observe the way in which the factors were combined to determine visual-search rates. The data yielded no evidence that the speed of search was influenced by an emergent property. Each of the factors simply contributed a constant amount to overall search speed. In sharp contrast to this finding, the 3D ratings produced strong evidence for emergent properties. Here the advantage of combining a linear gradient with a circular contour or a contrast-polarity reversal was clearly larger than could be predicted from the separate components.

This pattern of results suggests that visual search is guided by representations that do not explicitly code for

surface convexity/concavity or for the direction of lighting. Instead, the data support the view that search is based on representations that are better described as precursors to a rich 3D representation. These include the bounding contours of objects and the presence of reversals in contrast polarity, but they do not contain the recovered scene information in an object-centered form. One example of such an intermediate representation is Marr's (1982) "2½-D sketch." It contains viewer-centered information about surfaces in a scene, such as their depth, tilt, and slant relative to the viewer, but it does not contain the volumetric dimensions of the objects to which these surfaces belong. Such representations are only found at a subsequent stage where 3D solids are explicitly represented in object-centered coordinates.

These results thus contribute further to the view that visual search for 3D objects is not based directly on specialized detectors for volumetric solids. Two other recent reports suggest that it is not possible to base rapid visual search on the 3D properties that distinguish geons from one another (Brown, Weisstein, & May, 1992; Ju, 1990). The available data point instead to the view that visual search is based on features that are as high level as multiline junctions and local regions of shading (Enns & Rensink, 1990b, 1991a), but not as high level as object-centered 3D representations (Biederman, 1987). This conclusion also appears reasonable when one takes into account the combinatoric explosion that results from proposing early visual detectors for each of the thousands of objects that are encountered daily by the human visual system (Tsotsos, 1988). Thus, both the data and the logic argue for a preattentive visual system that looks for predictors of object properties in the image rather than looking directly for objects.

If preattentive vision is indeed sensitive to the direction of shading in an image and increases its sensitivity as contextual cues for object curvature are increased, where along the visual pathway might these computations be carried out? To our knowledge, there is no direct physiological evidence on this question, but work with simulated neural networks suggests that it could occur as early as the primary visual cortex (Lehky & Sejnowski, 1988). A neural network trained on luminance gradients at the "input" layer was able to "learn" to match these inputs to the correct "output" surface curvatures in the corresponding scene. When the authors inspected the behavior of the "hidden" units (i.e., those that connect the input and output layers), they found that these units bore a striking resemblance to the simple edge-detecting neurons of cortical area 17 in cat and monkey. This suggests that units in the earliest stage of cortical visual processing may already be tuned to surface curvature. We believe that further simulations like those of Lehky and Sejnowski, in conjunction with direct physiological investigations and psychophysical data such as those reported here, will eventually be able to suggest the regions of visual cortex involved in the rapid visual detection and perception of shaded objects.

The present results are also relevant to computational theories of shape-from-shading. One lesson they suggest is that relatively local analyses (e.g., local luminance-gradient analysis) and relatively distributed analyses (e.g., computing the shape of the outlining contour) should both be considered by models that strive for biological plausibility. At present, computational algorithms for shape-from-shading range from those that examine only local regions in the image (Horn, 1977; Pentland, 1984) to those that also take into account the shape of the outlining contour (Grossberg, 1983; Koenderink & van Doorn, 1980). Our data suggest, on the one hand, that a local luminance gradient analysis is not sufficient on its own to account for rapid visual search; the local context in which the gradient appears is also taken into account. On the other hand, the data also suggest that the computational goal for early vision need not be the complete recovery of an object-centered description. A more reasonable first step might be simply to combine information from several independent and relatively low level image analyses.

REFERENCES

- ATTNEAVE, F. (1954). Some informational aspects of visual perception. *Psychological Review*, **61**, 183-193.
- ATTNEAVE, F. (1967). Criteria for a tenable theory of form perception. In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form* (Vol. 19, pp. 56-67). Cambridge: MIT Press.
- BARLOW, H. B. (1953). Summation and inhibition in the frog's retina. *Journal of Physiology*, **119**, 69-88.
- BECK, J., GRAHAM, N., & SUTTER, A. (1991). Lightness differences and the perceived segregation of regions and populations. *Perception & Psychophysics*, **49**, 257-269.
- BECK, J., PRAZDNY, K., & ROSENFELD, A. (1983). A theory of textural segmentation. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 1-38). New York: Academic Press.
- BECK, J., SUTTER, A., & IVRY, R. (1987). Spatial frequency channels and perceptual grouping in texture segregation. *Computer Vision, Graphics & Image Processing*, **37**, 299-325.
- BENSON, C., & YONAS, A. (1973). Development of sensitivity to static pictorial depth information. *Perception & Psychophysics*, **13**, 361-366.
- BIEDERMAN, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, **94**, 115-147.
- BROWN, J. M., WEISSTEIN, N., & MAY, J. G. (1992). Visual search for simple volumetric shapes. *Perception & Psychophysics*, **51**, 40-48.
- BUFFART, H., LEEUWENBERG, E., & RESTLE, F. (1981). Coding theory of visual pattern completion. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 241-274.
- ENNS, J. T., OCHS, E. P., & RENSINK, R. A. (1990). VSearch: Macintosh software for experiments in visual search. *Behavior Research Methods, Instruments, & Computers*, **22**, 118-122.
- ENNS, J. T., & RENSINK, R. A. (1990a). Scene-based properties influence visual search. *Science*, **247**, 721-723.
- ENNS, J. T., & RENSINK, R. A. (1990b). Sensitivity to three-dimensional orientation in visual search. *Psychological Science*, **5**, 323-326.
- ENNS, J. T., & RENSINK, R. A. (1991a). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review*, **98**, 335-351.
- ENNS, J. T., & RENSINK, R. A. (1991b). VSearch Color: Full-color visual search experiments on the Macintosh II. *Behavior Research Methods, Instruments, & Computers*, **23**, 265-272.
- ENNS, J. T., & RENSINK, R. A. (in press). A model for the rapid interpretation of line drawings in early vision. In *Proceedings of the Second International Conference on Visual Search*, London: Taylor & Francis.
- ERIKSEN, C. W., & SCHULTZ, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics*, **25**, 249-263.
- GIBSON, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
- GROSSBERG, S. (1983). The quantized geometry of visual space: The coherent computation of depth, form, and lightness. *Behavioral & Brain Sciences*, **6**, 625-692.
- HORN, B. K. P. (1977). Understanding image intensities. *Artificial Intelligence*, **8**, 201-231.
- JU, G. (1990). *The role of attention in object recognition: The attentional costs of processing contrasts of non-accidental properties*. Unpublished doctoral dissertation, State University of New York, Buffalo.
- JULESZ, B. (1984). A brief outline of the texton theory of human vision. *Trends in Neuroscience*, **7**, 41-45.
- KELLMAN, P. J., & SHIPLEY, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, **23**, 141-221.
- KOENDERINK, J. J., & VAN DOORN, A. J. (1980). Photometric invariants related to solid shape. *Optica*, **27**, 981-996.
- LEEUEWENBERG, E. L. J. (1971). A perceptual coding language for visual and auditory patterns. *American Journal of Psychology*, **84**, 307-346.
- LEEUEWENBERG, E. L. J. (1988, November). *On geon and global precedence in form perception*. Paper presented at the meeting of the Psychonomic Society, Chicago.
- LEHKY, S. R., & SEJNOWSKI, T. J. (1988). Network models of shape-from-shading: Neural function arises from both receptive and projective fields. *Nature*, **333**, 452-454.
- LOFTUS, G. R. (1978). On the interpretation of interactions. *Memory & Cognition*, **6**, 312-319.
- MALIK, J., & PERONA, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America*, **7**, 923-932.
- MARR, D. (1982). *Vision*. San Francisco: W. H. Freeman.
- MCCLELLAND, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, **86**, 187-330.
- MCLEOD, P., DRIVER, J., & CRISP, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature*, **332**, 154.
- MINGOLLA, E., & TODD, J. T. (1986). Perception of solid shape from shading. *Biological Cybernetics*, **53**, 137-151.
- NAKAYAMA, K., & SILVERMAN, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, **320**, 264-265.
- PENTLAND, A. P. (1984). Local shading analysis. *IEEE Transactions: PAMI*, **2**, 170-186.
- PENTLAND, A. P. (1986). Perceptual organization and the representation of natural form. *Artificial Intelligence*, **28**, 293-331.
- RAMACHANDRAN, V. S. (1985). The neurobiology of perception. *Perception*, **14**, 97-103.
- RAMACHANDRAN, V. S. (1988). Perceiving shape from shading. *Scientific American*, **259**, 76-83.
- ROCK, I. (1983). *Logic of perception*. Cambridge: MIT Press.
- STERNBERG, S. (1969a). The discovery of processing stages: Extensions of Donder's method. In W. G. Koster (Ed.), *Attention and performance II* (pp. 276-315). Amsterdam: North-Holland.
- STERNBERG, S. (1969b). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, **57**, 421-457.
- TAYLOR, D. A. (1976). Stage analysis of reaction time. *Psychological Bulletin*, **83**, 161-191.
- TODD, J. T., & MINGOLLA, E. (1983). Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 583-595.
- TREISMAN, A. (1986). Features and objects in visual processing. *Scientific American*, **255**, 106-115.
- TSOTSOS, J. K. (1988). A "complexity level" analysis of immediate vision. *International Journal of Computer Vision*, **1**, 303-320.
- TURVEY, M. T. (1973). On peripheral and central processes in vision: Interference from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, **80**, 1-52.

NOTE

1. An additional experiment tested the importance of having an explicit border in the search items. Five subjects from Experiment 1 searched for the linear-circle and linear-square on a white background, with and without a surrounding border. The results were very similar in the two border conditions [$F(1,4) = .13$] (linear-circle: mean slope for target present = 4 msec per item with border and 3 msec per item without border, mean slope for target absent = 5 msec per item with border and 4 msec per item without border; linear-square: mean slope

for target present = 11 msec per item with border and 9 msec per item without border, mean slope for target absent = 11 msec per item with border and 13 msec per item without border). Regardless of the presence or absence of the border, search for the linear-circle was more than twice as fast as search for the linear-square [$F(1,4) = 12.90, p < .05$].

(Manuscript received June 14, 1990;
revision accepted for publication January 23, 1992.)