Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 11155–11159, November 1995
Psychology

Visual attention to surfaces in three-dimensional space

ZIJIANG J. HE* AND KEN NAKAYAMA

Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138

Communicated by Jacob Nachmias, University of Pennsylvania, Philadelphia, PA, August 10, 1995

ABSTRACT **Although attention plays a significant role in vision, its spatial deployment and spread in the third dimension is not well understood. In visual search experiments we show that we cannot easily focus attention across isodepth loci unless they are part of a well-formed surface with locally coplanar elements. Yet we can easily spread our attention selectively across well-formed surfaces that span an extreme range of stereoscopic depths. In cueing experiments, we show that this spread of attention is, in part, obligatory. Attentional selectivity is reduced when targets and distractors are coplanar with or rest on a common receding stereoscopic plane. We conclude that attention cannot be efficiently allocated to arbitrary depths and extents in space but is linked to and spreads automatically across perceived surfaces.**

How do we direct our visual attention to specific loci in three-dimensional space? Recent studies suggest that the visual system can focus attention to a particular depth defined by binocular disparity (1–3). Initial evidence came from a study on feature-conjunction search experiments by Nakayama and Silverman (1). Fig. 1*a* is a stereogram similar to their stereoscopic stimuli, where search elements are located in one of three fronto–parallel planes (for those who cannot free-fuse the stereogram, see the diagram in Fig. 2*a*). The target is located in the middle-depth array (zero disparity), the "search plane," where all elements have the same color (red or green), except for the target which is of the opposite color (green or red). The observers' task is to find the single odd-colored target within this plane, despite various numbers of distractors of the same color in the nearer and farther flanking planes. Nakayama and Silverman (1) found that search reaction times were independent of the total number of distractors. To explain this result, they hypothesized that the visual system can selectively focus attention on a particular binocular disparity and that isodisparity, or common depth, constituted the basis for attentional selection (isodepth hypothesis).

Nakayama and Silverman's display, however, confounds binocular disparity with the presentation of a well-formed surface composed of coplanar elements. Therefore, an observer's ability to search efficiently could be the result of directing attention to the middle *surface* rather than to a particular disparity (depth) value. The visual system, according to this "surface" hypothesis, can direct selective attention efficiently to any well-formed perceptually distinguishable surface; this hypothesis does not require that all elements be equidistant (same depth) from the observer.

EXPERIMENT 1

Common Depth Is Not a Sufficient Basis for Attentional Deployment

To distinguish between these two hypotheses, we have constructed search displays where the two factors—common surface vs. common depth—could be dissociated. As shown in the stereograms of Fig. 1*b* and the illustrations of Fig. 2 *b* and *c*, we can disrupt the perceptual grouping of the fronto–parallel surface by slanting the individual elements away from the fronto–parallel plane, while keeping their average disparity unchanged. Such manipulation should have very little effect on the search performance according to the isodepth hypothesis but should impede search performance according to the surface hypothesis. The reader should free-fuse the stimuli in Fig. 1, either cross fusing the two left columns or uncross (parallel) fusing the two right columns. Fig. 1*a* is similar to the Nakayama and Silverman (1) case. Notice that without much effort, it is possible to attend to the middle fronto–parallel plane and select the odd-colored target, despite the existence of many distractors in the nearer and farther depth planes. In contrast, Fig. 1*b* depicts a situation where within each plane the individual elements are slanted, while preserving the mean depth of each. Thus, the depth elements are no longer coplanar with other elements at the same depth and are no longer perceived as a well-formed unitary surface. Predicted by the surface hypothesis and against the isodepth hypothesis, it is more difficult to attend to the middle set of elements, and therefore it is more difficult to find the odd-colored target.

To obtain quantitative support for these phenomenological observations, we measured reaction times to detect targets in such common depth, or isodisparity, arrays. Observers were required to search the middle 4×3 element isodisparity depth array, containing either coplanar elements (Fig. 2*a*) or elements slanted out of the plane of the array (Figs. 2 *b* and *c*). According to the isodisparity hypothesis, performance on all tasks should be equivalent, whereas the surface hypothesis predicts selective degradation in the non-coplanar arrays.

Methods

Binocular fusion of separate left eye and right eye images was achieved by a phase haploscope, consisting of liquid crystal shutters synchronized with alternating cathode ray tube frames for each eye. Frame rate was 60 Hz. Viewing distance was 60 cm.

Search elements were either red or green, arranged in a three-dimensional $4 \times 3 \times 3$ matrix and displayed on a black background (see Figs. 1 and 2 *Upper*). The binocular disparity between each common depth array was 27 arc min. The center-to-center vertical distance between two elements in a given common depth plane was kept constant at 2.4° of visual angle, while their horizontal distances varied randomly over a 2.0–2.3°. The size of each element was $\approx 1.2 \times 0.3°$; the individual shapes of the elements were skewed to allow the stereoscopic slant to vary as needed (within a disparity range of 9 arc min). The total number of elements was always 36. Between trials, the subject fixated a point having the disparity of the middle depth array (the zero disparity distance).

The subject initiated a trial by depressing a mouse button, and the search stimulus was displayed after a random period of 1–2 sec. The task was to release the mouse button when the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*Present address: Department of Psychology, University of Louisville, Louisville, KY 40292.

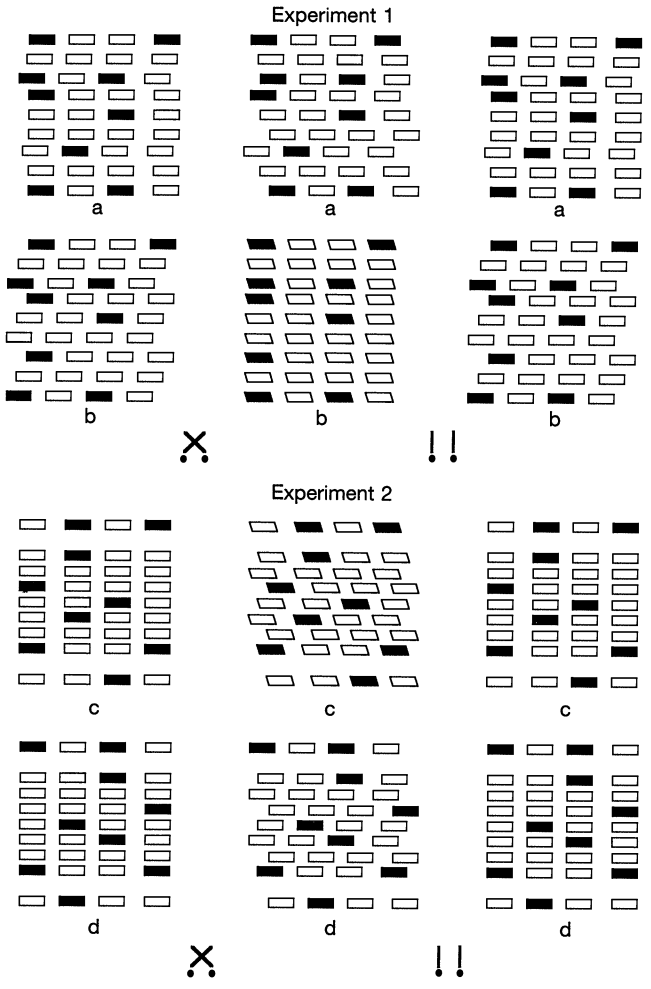

FIG. 1. Stereograms (a,b) and stereograms (c,d), similar to those used in experiments 1 and 2 (and described in Fig. 2 *a*,*b* and Fig. 3*a*, *b* respectively). The reader should fuse the left and center images convergently or the center and right images divergently. The task in Exp. 1 was to find the odd-colored target in the middle fronto-parallel plane (forming rows 2, 5, and 8 in the two-dimensional image). Compare stereograms a and b). Search is much easier when the elements are also oriented in the fronto–parallel plane (a). The task in Exp. 2 was to find the odd-colored target in the middle horizontal stack (forming rows 3,5, and 7). Compare stereograms c and d. Here, the search is easiest when the elements lie in this receding stereoscopic plane (c).

odd-colored target (either red or green) was detected in the middle depth array of 12 elements.

Results

Averaged reaction times for 100 trials are plotted individually for the two observers. Even though all search elements in the middle depth array had the same disparity, search reaction times were greatly prolonged when the implicit surface (Fig. 2*a*) was disrupted by slanting the elements either away (Fig. 2*b*) or toward (Fig. 2*c*) the observer. These results confirm the prediction of the surface hypothesis.

It could be argued that the greater search reaction times shown in Fig. 2 *b* and *c* were due to the reduced end-to-end depth differences between slanted elements at the three depths. To control for this possibility, we replicated the coplanar case, this time using disparity differences of 13 arc min so that they were much smaller than the reduced end-to-end depth difference. Despite this reduction, reaction times remained consistently short (see white bars in Fig. 2*a*).

EXPERIMENT 2

Common Depth Is Not a Necessary Basis for Attentional Deployment

In the previous experiment we found that common depth, or isodisparity, alone, was insufficient to support rapid visual search for an odd target. In this experiment, we show that isodisparity is also not necessary. Subjects can selectively attend and search efficiently for targets in a planar surface that span a wide range of binocular depths. As shown in the stereograms of Fig. 1*c* (and illustrated in Fig. 3*a*) an odd-colored target was presented within the middle "horizontal" search plane of a stack of nearly horizontal parallel planes, each spanning multiple depths. Supporting our surface hypothesis, when Fig. 1*c* is fused as a stereogram, we can selectively attend to items in each of these different planes, even though each row in a given plane is interleaved (in the image) with rows from the other planes. Contrast this to the stereogram shown in Fig. 1*d* (illustrated in Fig. 3*b*), where the same 12 elements (in position and depth) no longer constitute a set of coplanar elements. As predicted by the surface hypothesis, attending to this set requires much greater effort.

Again, to evaluate this hypothesis quantitatively, we compared the search reaction times for the three different local arrangements of element orientation within the global search plane's orientation, where the elements were tilted forward by adding 6.7 arc min disparity (Fig. 3*b*) or even more forward by adding 13.8 arc min disparity (Fig. 3*c*).

Methods

The stimulation parameters and procedures were similar to those indicated in Fig. 1, except for the following. The stimulus display consisted of three stacks of slightly slanted horizontal arrays separated vertically from each other by 107 arc min. Each two rows of elements within a horizontal array had a vertical separation of 67 arc min, and a depth difference (disparity) of 22 arc min. Within each row, the horizontal separation between any two elements ranged from 122 to 133 arc min. The size of a search element in condition b was 80 × 20 arc min (Fig. 3*b*). In conditions a and c (Fig. 3 *a* and *c*), because there was a 6.7 arc min disparity between each of their elements' top and bottom edges, the sizes of the retinal image of the elements were skewed accordingly.

Results

Comparison of performance under each condition shows that average reaction times are clearly shorter when the elements within the search array form an implicit surface plane that is congruent with the search array (Fig. 3*a*). This result indicates that observers can easily spread their attention to well-formed surfaces spanning a wide range of depths (disparities). In other words, isodisparity, or common depth, is not required.

Experiment 3: Mandatory Spread of Attention Across Surfaces

So far we have shown that attention can spread selectively over well-formed surface groupings independent of their orientation or slant in three-dimensional space. Here we ask whether this finding reflects a voluntary attentional process or whether it reflects a stronger, more mandatory relationship. In other words, is there an obligatory process that automatically causes attention to spread across surfaces?

To examine this issue, we reversed the logic of our previous experiments and devised a task that requires the confinement of attention, rather than its spread. If attention spreads

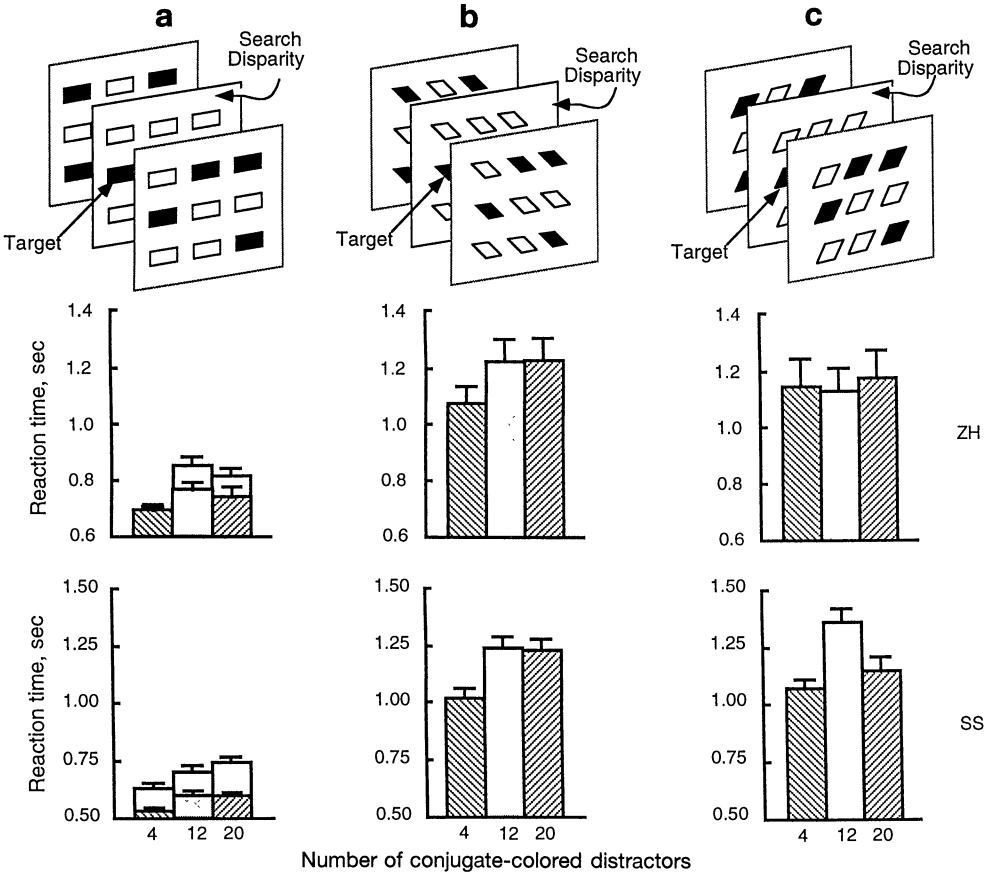

FIG. 2. Visual search in the middle vertical array defined by a single binocular disparity. Three vertical arrays were presented to the observer, whose task was to detect the odd-colored target (colored black in diagram) in the *middle depth* array despite the presence of a variable number of similarly colored distractors in the adjacent depth planes (outlined by rectangles not visible to the observer). Subjects (ZH, the author; SS, a naive observer) find it easiest to detect the odd-colored target when all elements of the middle array are coplanar with respect to the search plane (*a*) than when the targets are either slanted more backward (*b*) or more forward (*c*). Note also that there is no systematic increase in reaction time for increased numbers of same colored distractors. The open histogram bars in *a* are obtained in a separate control experiment, where the disparity difference between the planes is reduced to conform to the end-to-end disparity differences in conditions b and c.

automatically across surfaces, it should be more difficult to restrict it to sections within surfaces as opposed to regions between surfaces.

Methods

We used a cueing paradigm (2, 4) where we required that the observer detect an odd-colored target in either the upper or lower row containing three elements apiece (see Fig. 4). Before each stimulus presentation, a cue identified the row in which the odd target was to appear. The cue was correct (valid) 80% of the time. Looking for a difference in reaction time to detect targets in cued vs. noncued rows as an indicator of selective attention, we varied the binocular disparity between the upper and lower rows. In condition a (Fig. 4*a*), we simply increased the binocular disparity between the upper and lower rows. This change means that, in addition to the increased depth difference, these rows became more clearly defined as lying on two separate fronto-parallel planes with increased depth difference. In condition b (Fig. 4*b*), the same increase in binocular disparity between the upper and lower rows occurred, but the stereoscopic slant of each individual element was adjusted so that each element remained coplanar within a single larger implicit receding plane. In this situation increased depth difference does not alter the coplanar relationship between the upper and lower rows.

During the cueing experiment, the observer fixated a cross located between the two rows of black elements (Fig. 4). To initiate a trial, the observer pushed a button; this was followed by a brightening of either the upper or the lower limb of the fixation cross, which instructed him to direct his (focal) attention to the upper or lower row of elements, respectively. After a 1- to 2-sec random delay, one element (the target) in one of the two rows would change its color to either light-green or dark-green, while the remaining elements in both rows changed to the alternative color. [The target had an 80% chance to appear in the attended row (cue-valid condition) and a 20% chance in the unattended row (cue-invalid condition).] Upon seeing the target, the subject released the button. The time from the display onset to the release of the button defined the reaction time.

Each element subtended 53 × 41 arc min in the zero-disparity condition. The vertical distance between the two rows of elements was 2.5°. The horizontal separation between any two elements in a row was 46 arc min. In condition 4*b*, where individual elements were slanted in depth, the monocular retinal images of the individual elements were skewed so that elements in the upper and lower rows were coplanar.

Results

Reaction times for the cue-valid and cue-invalid condition are seen in Fig. 4 *Lower*. The curves corresponding to the valid case always show a shorter reaction time than the invalid case. Thus, in all conditions we see a main effect of selective attention for all values of binocular disparity. The overall pattern of results with increased binocular disparity is revealing. In Fig. 4*a* we see increased attentional selectivity with increased binocular

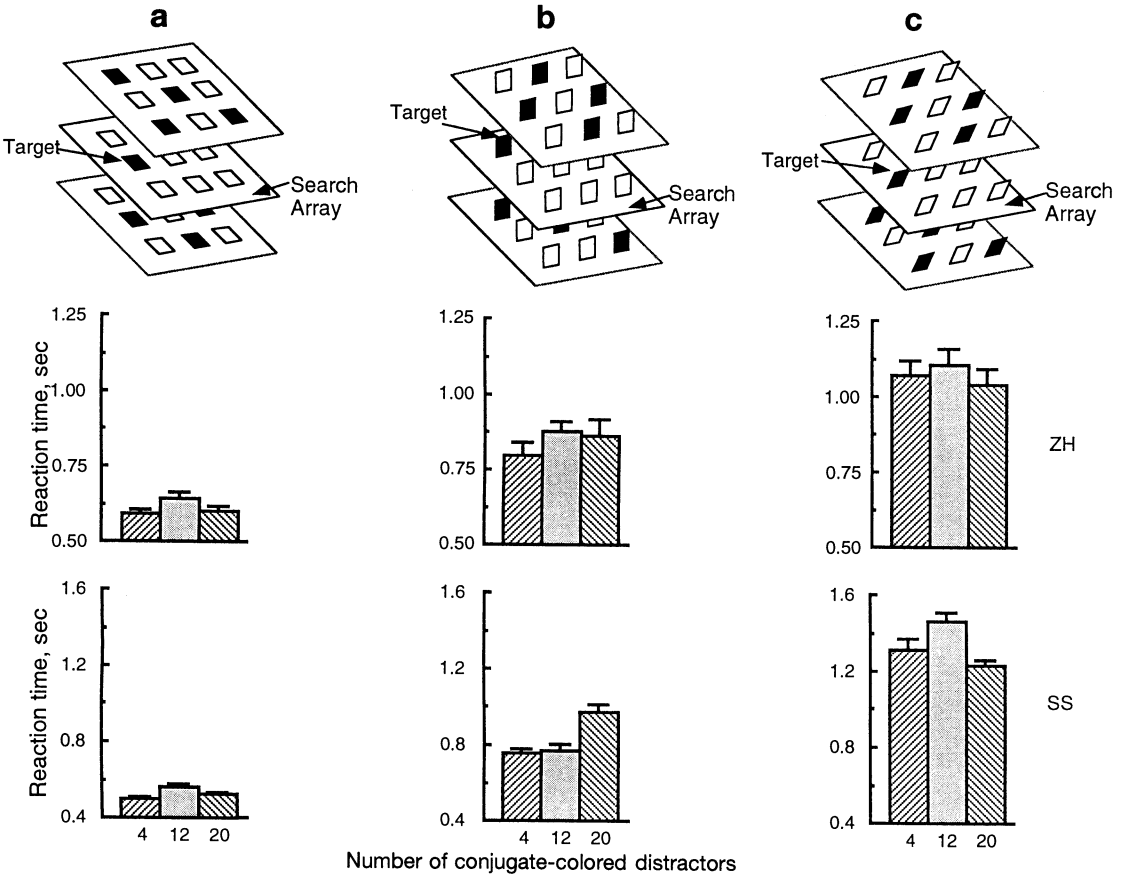

FIG. 3. Visual search confined to the middle of three near-horizontal arrays defined stereoscopically, each array spanning an extreme range of stereoscopic depths. Comparison of search reaction times between condition a, where the slanted "horizontal" search elements' array is coplanar with respect to the horizontal search plane and conditions b and c, where such coplanarity was disrupted. Average reaction times indicate that the search is fastest when the elements within the search array form an implicit surface plane that is congruent with the search array (*a*). Note that in this figure (and in Fig. 2), the outline rectangle illustrates the spatial layout of the search array; this rectangle itself is not visible to the observer.

depth differences (the cue-invalid condition's reaction time increases). Several interpretations are possible. According to the isodepth hypothesis discussed earlier (1, 2), binocular disparity differences alone are responsible for allowing greater attentional selectivity. According to our surface hypothesis, increased disparity also leads to an increase in perceptual segregation of the two planes, thus allowing greater selectivity between the upper and lower rows.

The results from condition b (Fig. 4*b*) support the surface hypothesis, as there is no observed increase in selective attention. This result is expected because two perceptually separate planes are not evident. The stereoscopic slant of each element is adjusted so that all elements remain coplanar, forming a single plane receding back in depth. Binocular disparity by itself does not aid in the sequestering or confinement of attention. Attention appears to spread across the regions of a surface despite efforts to restrict it to a single row.

The pervasiveness of this involuntary spread of attention can be further seen in the experimental configuration described in Fig. 4*c*. Here and similar to condition a, each row is fronto–parallel to the observer. As such, each row would have the same chance as being grouped and thus perceived as separate surfaces. Added, however, is a stereoscopic plane defined by random dots upon which each of these fronto–parallel rows of elements is perceived to rest. The reaction time to detect the unattended target (in the cue-invalid row) showed no increase with binocular disparity, even though the perception of two separate planes would be presumed to increase as in condition a. This result suggests that attention not only flows freely and automatically within a surface but that this surface-based spread is so pervasive that it can also extend to systems of contiguous surfaces.

General Discussion

We have demonstrated that attention in three-dimensional space spreads preferentially and automatically over perceived surfaces. Attention is not easily moved to arbitrary points or volumes in abstract space but is bound to perceived surfaces. Thus the operation of visual attention depends heavily on a perceptual representation.

These findings are consistent with the view that the spread of attention in three-dimensional space is generally not content free. In this regard the results are related to the work of Duncan (5), who argued that attention was selectively shifted to different attributes within an "object" (a line), which overlapped another (a rectangle) in space [see also Neisser and Becklen (6)]. Thus one could argue that our results provide an example of object- rather than surface-bound attention. We acknowledge that the set of criteria as to what distinguishes an object or a surface is, in part, semantic. Yet, we can preserve the intuitive notions of each by considering them as extremes along a continuum. At one end are known recognizable objects or object classes formed through stored items in visual memory; at the other end are emergent perceptual units of a more primitive nature, surfaces existing independent of specifically known objects. Objects of many kinds exist for humans, and the set is very large. A smaller set of surface elements composes these objects, all subject to fairly restricted sets of properties, local coplanarity of surface orientations, collinear-

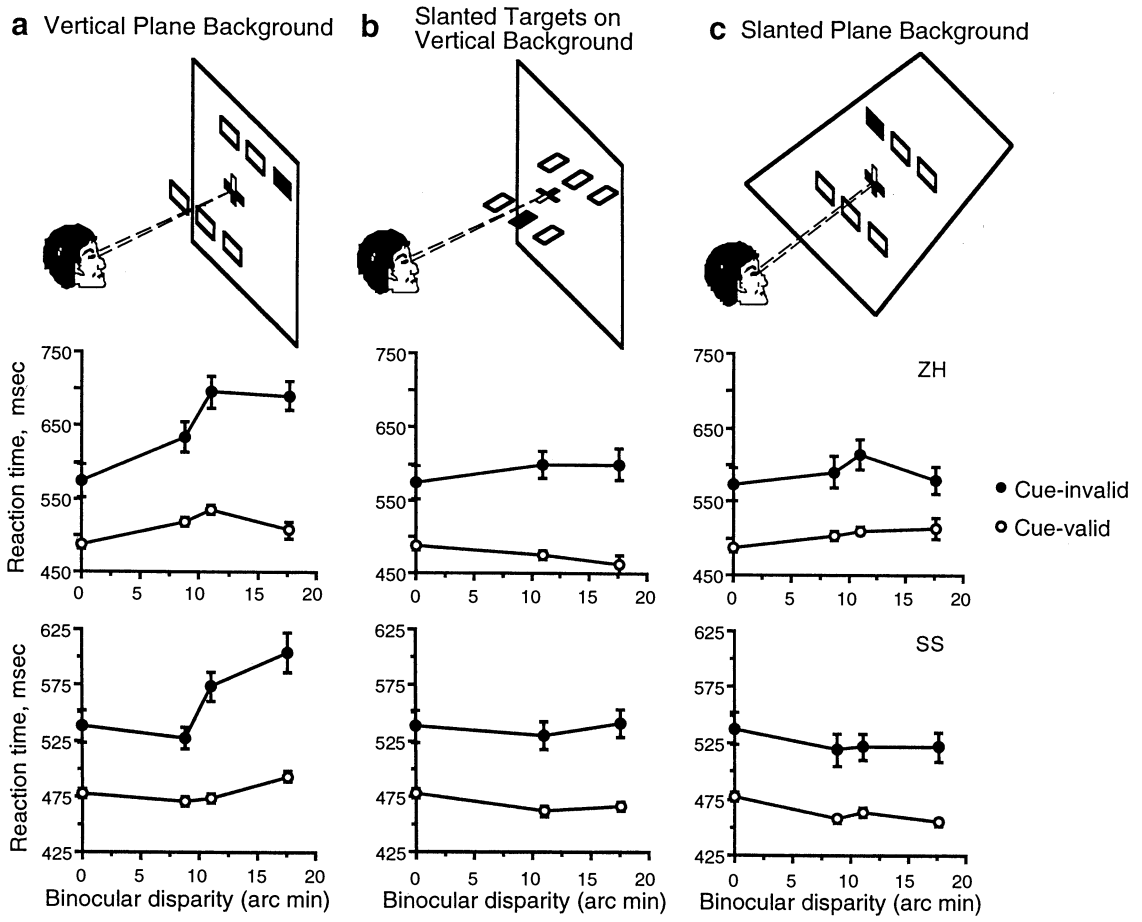

FIG. 4. Cueing experiment where reaction time is plotted as a function of the binocular disparity between the fixation cross and elements. When the cue is invalid (●), reaction time increases with increased binocular disparity only in condition *a*, not in *b* and *c*. When the cue is valid, reaction time does not significantly change in all three cases. [Two-way ANOVA was done on the data. A strong significant interaction was found in condition a ($P < 0.001$) but not in condition b or in condition c ($P > 0.05$).]

ity of surface edges, etc. We argue that our configuration and even the configuration of Duncan (5) appear more related to the primitive notion of surfaces than to the class of specifically experienced objects or object classes.

Whatever the label we attach to the process, we suggest that efficient deployment of attention to perceptual representations rather than to simple loci in space has implications for the understanding of attention at a more mechanistic level. Implied in a wide range of visual attentional studies (1, 7, 8) is the idea of attending to features and locations, which appears to fit closely with the prevailing concept of the visual system as a hierarchy of receptive fields (9). Attention in this framework can be seen as the facilitation or inhibition of such receptive fields that are retinotopically organized. Against this view, the present results suggest a need for an examination of neuronal signals that represent emergent perceptual representations rather than an examination of such signals in terms of local receptive field properties (10, 11). Our view is that a surface representation provides an intermediate level of perceptual organization required for this deployment and spread of visual attention.

This work was supported, in part, by a grant from Air Force Office of Scientific Research to K.N., and a Sloan Research Fellowship to Z.J.H.

1. Nakayama, K. & Silverman, G. H. (1986) *Nature (London)* **320,** 264–265.
2. Downing, C. & Pinker, S. (1985) in *Attention and Performance 11*, eds. Posner, M. I. & Marin, O. S. M. (Erlbaum, Hillsdale, NJ), pp. 171–187.
3. Andersen, G. (1990) *Percept. Psychophys.* **47,** 112–120.
4. Posner, M. I., Snyder, C. R. & Davidson, B. J. (1980) *J. Exp. Psychol. Gen.* **109,** 160–174.
5. Duncan, J. (1984) *J. Exp. Psychol. Gen.* **113,** 501–517.
6. Neisser, U. & Becklen, R. (1975) *Cognit. Psychol.* **7,** 480–494.
7. Moran, J. & Desimone, R. (1985) *Science* **229,** 782–784.
8. Hillyard, S. A. (1995) *Neural Systems Mediating Selective Attention* (MIT Press, Cambridge, MA).
9. Hubel, D. H. & Wiesel, T. N. (1968) *J. Physiol.* **195,** 215–143.
10. Nakayama, K. & Shimojo, S. (1992) *Science* **257,** 1357–1363.
11. He, Z. J. & Nakayama, K. (1992) *Nature (London)* **359,** 231–233.