PERGAMON

Vision Research 40 (2000) 2265–2276

Vision
Research

www.elsevier.com/locate/visres

Neither occlusion constraint nor binocular disparity accounts for the perceived depth in the 'sieve effect'

Jeffrey J. Tsai *, Jonathan D. Victor

Department of Neurology and Neuroscience, Weill Medical College of Cornell University, Room K615, 1300 York Avenue, New York, NY 10021, USA

Received 11 October 1999; received in revised form 1 March 2000

Abstract

Current notions of binocular depth perception include (1) neural computations that solve the correspondence problem and calculate retinal positional disparity, and (2) recovery of ecologically valid occlusion relationships. The former framework works well for stimuli with unambiguous interocular correspondence, but less so for stimuli without well-defined disparity cues. The latter framework has been proposed to account for the phenomenon of perceived depth in stimuli without interocular correspondence, but its mechanism remains unclear. In order to obtain more insight into the mechanism, we studied the depth percept elicited by a family of stereograms — 'sieve' stimuli, adapted from Howard (1995) [*Perception*, *24*, 67–74] — with interocular differences but no well-defined positional disparity cue. The perceived depth was measured by comparison to references at various depths established by standard retinal disparity and was consistently found to lie behind the fixation plane. Moreover, the magnitude of the depth percept depended on both the horizontal and vertical spatial characteristics of the stimulus in ways that were at odds with constraints of occlusion geometry. In comparison to the depth percept elicited by stimuli with well-defined disparity cues, the precision of the percept from the sieve stimuli was 10–20 times worse, suggesting that a different underlying computation was involved. Thus, neither of the above frameworks accounts for the depth percept arising from these stimuli. We discuss implications of our results for physiologically based computations underlying binocular depth perception.

Keywords: Stereopsis; Binocular disparity; Matching; Ecological optics

1. Introduction

The frontal eye placement in humans leads to a relative horizontal displacement of images projected onto the retinas for objects that are not at fixation. The visual system can use this positional disparity to recover depth (Wheatstone, 1838), even in the absence of monocular depth information (Julesz, 1971). This binocular positional disparity is considered to be an important cue for the range of depths relevant to manual dexterity (McKee, Levi & Bowne, 1990; Fielder & Moseley, 1996) because it provides precise and metrical measures of depth from the fixation plane. Many studies have focused on the mechanisms of binocular depth perception, or stereopsis — from physiological studies of neurons in the visual cortex that might detect disparity (e.g. Barlow, Blakemore & Pettigrew, 1967; Hubel & Wiesel, 1970; Poggio, 1990; Ohzawa, DeAngelis & Freeman, 1990) to computational models that can solve random-dot stereograms (reviewed in Poggio & Poggio, 1984; Blake & Wilson, 1991; Weinshall & Malik, 1995; Qian, 1997). One predominant theme in these works is the idea that stereopsis is predicated on binocular disparity, which in turn depends on the correct matching of features in the left and right eyes' views. Binocular disparity of corresponding points is thus considered fundamental to stereopsis.

In light of the fundamental role of inter-ocular matching in stereopsis, it is highly noteworthy that stimuli that do not possess well-defined disparity signals can also produce an impression of depth (Panum, 1858; Nakayama & Shimojo, 1990a; Anderson, 1994; Liu,

---

* Corresponding author. Tel.: +1-212-7466522; fax: +1-212-7468984.

E-mail address: jtsai@med.cornell.edu (J.J. Tsai).

0042-6989/00/$ - see front matter © 2000 Elsevier Science Ltd. All rights reserved.
PII: S0042-6989(00)00100-0

Stevenson & Schor, 1994). Common to these dichoptic stimuli are luminance elements that are present in only one eye's view and hence have no corresponding interocular matches. It is important to note that the presence of unmatchable features can only be detected by a process that compares the input from both eyes; hence depth from unmatched images cannot be attributed to monocular or pictorial cues. Moreover, since the disparity signal, which is predicated on the existence of matchable elements, is ill-defined in these stimuli, another explanation has to be found to account for the depth perceived.

The aforementioned investigators considered this phenomenon in the context of properties of the visual world. In a world in which opaque objects partially occlude one another, the differential occlusion due to parallax may generate left and right retinal images that are not matchable. Unmatched images can thus provide a clue to three-dimensional surface relationships. Since identifying such surface relationships in the real world is an important step in intermediate visual processing (Gibson, 1950; Marr, 1982; Nakayama & Shimojo, 1990b, 1992), use of these cues may reflect a developmentally learned or evolutionarily selected response (Nakayama & Shimojo, 1990a; Shimojo & Nakayama, 1990). According to this argument, the perceived depth from unmatched images is necessarily consistent with real-world occlusion.

The above 'ecological optics' (EO) heuristics may account for the result of, but does not address, the process by which the depth response is obtained. Part of the difficulty is reflected by the debates on the nature of the depth response. For example, Nakayama and Shimojo (1990a) and Liu et al. (1994) separately reported figural stereograms without binocular correspondence which produced quantitative depth perception consistent with occlusion. However, in the case of Nakayama and Shimojo (1990a), models that allow multiple-matching of features could also account for the depth percept, based essentially on position disparity signals (Gillam, Blackburn & Cook, 1995). In the case of Liu et al. (1994), residual binocularly-corresponding features (Gillam, 1995) or the output of appropriate linear filters (Liu, Stevenson & Schor, 1997) may provide conventional disparity depth cue. For these stimuli, the depth percept driven by 'unmatched image points' may well be explained on the basis of classical stereopsis. In order to determine whether an alternate mechanism is involved, there is a need to examine the characteristics of depth perception from unmatched images in which other binocular depth cues have been eliminated (Gillam & Nakayama, 1999).

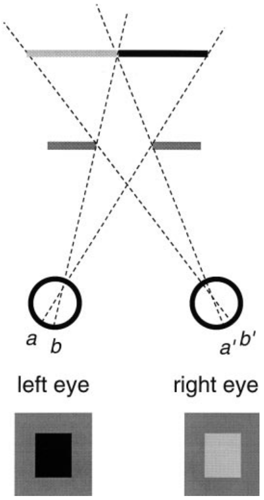

Fig. 1. A bird's eye view of the eyes looking at two fronto-parallel surfaces. The near surface has an aperture through which the far surface can be seen. Different shades of gray represent different brightness on the far surface. The image below each eye shows the respective retinal image. Note that due to occlusion, the inner portion of the retinal images carry opposite contrast polarities.

One of those characteristics is the precision of depth judgment. For stereopsis based on matchable binocular images, precision can be high. The stereoacuity threshold for targets presented near the fixation plane is a few seconds, comparable to the threshold for monocular hyperacuity (Ogle, 1953). Discrimination threshold increases as the target moves away from the fixation plane (McKee et al., 1990); but even when the target is far from fixation, depth information consisting of more than mere depth ordering is obtained with a measurable precision. In contrast, real-world constraints of occlusion offer only a minimum bound on the depth of the occluded surface. In the absence of other depth cues, a continuous range of depth configurations can produce the same retinal images. That is, the theoretical limit for the precision of depth localization is inherently worse for unmatched cues than for matchable cues. Thus, finding that unmatched cues are less precise would be consistent with the idea that the visual system recovers depth via a reconstruction of real-world surface relationships. However, we note that a finding that unmatched cues provide an equally precise depth percept would not rule out this explanation, since the visual system may use a priori considerations or 'rules of thumb' to make inferences that are not rigorously implied by the visual stimulus.

Another way to test the EO framework is via its predictions about the perceived depth of the stimulus. EO posits that a real-world stimulus provides all the information necessary to obtain a veridical percept. Fig. 1 diagrams a top-down view of two fronto-parallel surfaces; the near surface has an aperture and partially occludes the far surface. Without loss of generality, let us assume that the eyes are fixated on the near surface; then the image points a and a' and points b and b' fall on corresponding locations on the two retinas. The angle between a and b (or a' and b') equals the horizon-

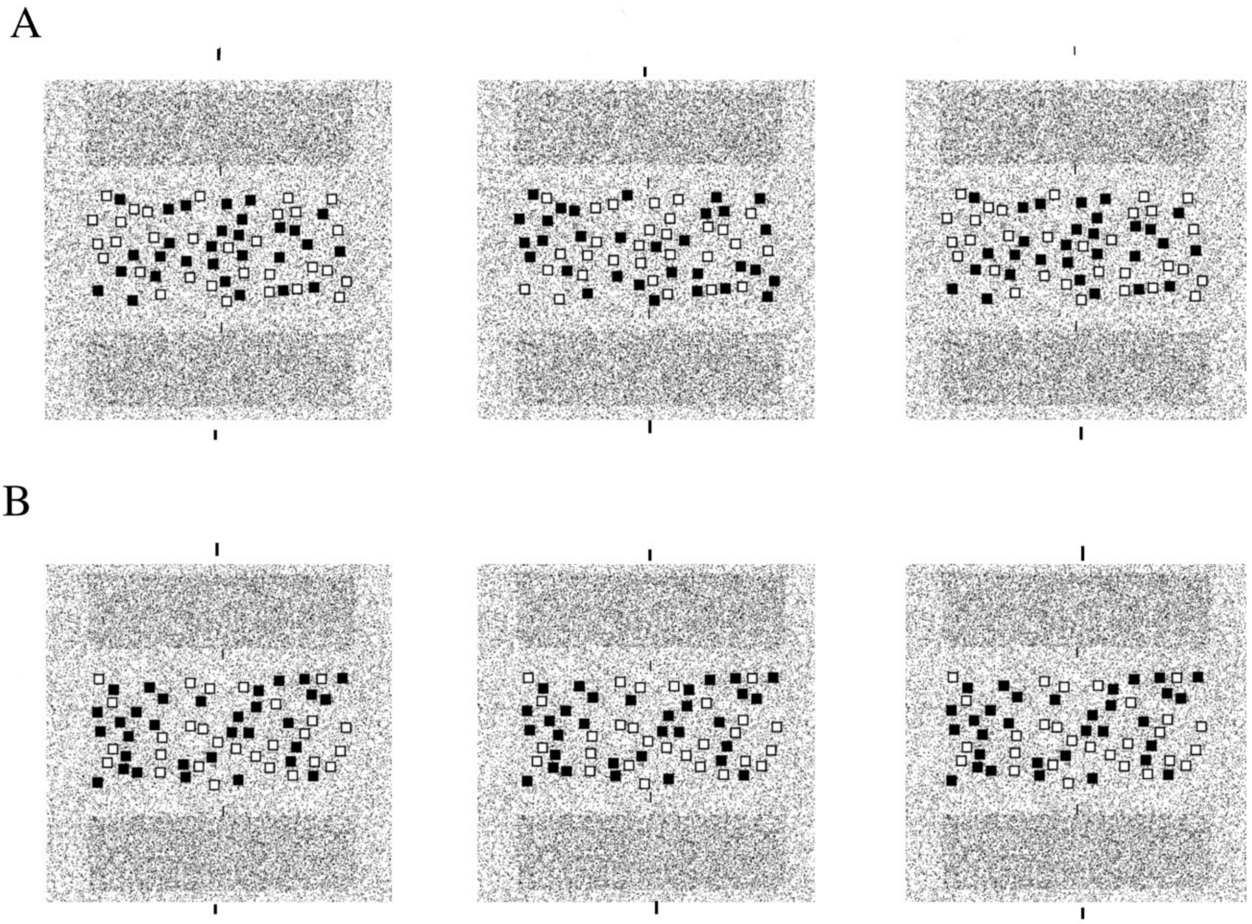

Fig. 2. Examples of (A) the sieve stimulus and (B) the control stimulus. They differ in scale from those used in the experiments and are only intended to approximate the perceptual effects. In some experiments, the upper and lower random-dot stereograms in the control stimulus were replaced with sinusoidal grating stereograms to achieve sub-pixel disparity. Uncrossed-fusers should view the left pair; crossed-fusers view the right pair. The depth probes in the upper and lower thirds of both stimuli have a 3-pixel uncrossed disparity. The disparity of the target in the center of the control stimulus (B) is four pixels uncrossed.

tal width of the aperture. The far surface must be sufficiently distant so that the two eyes' views are non-overlapping. If it were any closer, then parts thereof would be visible to both eyes. This minimum distance constraint has a relative disparity equal, in angle subtended, to the horizontal width of the aperture. Based on this geometric argument, the depth perceived from perfectly unmatched images should be constrained by the horizontal extent of the aperture. Moreover, the perceived depth should be independent of the vertical extent of the aperture.

In these experiments, we measured the perceived depth of a family of stereograms that lacked binocular matches as well as confounding disparity cues. We also measured the precision of the depth percept. We found that: (1) the precision was an order of magnitude poorer than for a comparable classical stereogram; (2) the perceived depths often were not consistent with occlusion geometry; and (3) the vertical extent of the aperture affected the perceived depth, but to a lesser degree than did the horizontal extent. The results suggest that neither horizontal disparity alone nor occlusion can account fully for the depth perceived from unmatched images, and constrain models for neural mechanisms underlying the depth computation.

2. Methods

2.1. Visual stimuli

The subjects' task was to judge the relative depths in the display in two types of stimuli: 'sieve' and 'control' (see Fig. 2A,B). The sieve stimulus, adapted from Howard (1995), consisted of an array of 'panes' randomly positioned in the middle third of the display, with two random-check stereograms occupying the upper and lower thirds. Each pane consisted of a rectangular or square block of pixels at maximum or minimum display luminance, surrounded by a 2-pixel rim of minimum luminance. The luminance of the pane interior was anti-correlated between the two eyes. In the left and right eye images, each pane occupied the same spatial location and hence had zero binocular disparity. Thus, there was interocular pairing of the panes, but the panes of any pair were not 'matched'. The percept was one of looking through holes punched out in a plane at another surface beyond — hence the name 'sieve effect'. Flanking the array of panes above and below were two random-check stereograms (in which a rectangular surface was rendered in depth by horizontal disparity) whose depth was compared against that of the panes. The two probes were identical in dimensions and disparity. The panes and the probes were embedded in a random check reference field of 10% density and zero binocular disparity. Nonius markers were placed along the vertical meridian of the stimulus. Both the reference field and the nonius markers were continuously visible during a trial. The entire stimulus extended $8.3 \times 8.3°$ and was centered in a $15 \times 11°$ field of mean luminance.

The control stimulus differed from the sieve stimulus in three respects: (1) The interior pane luminance was correlated between the two eyes; (2) Each pane had a common, non-zero binocular disparity; and (3) The random-check depth probe in the upper and lower thirds was replaced by a 1 c/deg vertical sinusoidal grating. The grating was generated with sub-pixel resolution to allow measurement of the sub-pixel thresholds for depth discrimination. We adopted a grating probe because, in pilot experiments with the texture probe that was used in the sieve stimuli, subjects' performance was still close to perfect even at one-pixel disparity.

All stimuli were presented on a SONY 17SEII color display (calibrated using the VSG OptiCal photometer), and viewed at a distance of 114 cm (one screen pixel subtended 1 min). Stereo display was achieved by presenting the image to each eye in alternate video frames using a pair of FE-1 ferroelectric light valves switching at 120 Hz (60 Hz per eye). A VSG 2/3F graphics card generated and displayed the stimuli at 15-bit grayscale resolution. The photometer, the graphics card, and the shutter goggles were manufactured by Cambridge Research Systems, UK. Stimuli were presented in the red channel only because this allowed maximal elimination of cross-talk through the FE goggles; the leakage resulting from a 100% contrast stimulus was undetectable ($<0.5\%$).[1] The mean screen luminance was 9.2 cd/m^2, and was reduced to 1.4 cd/m^2 when viewed through the shutters.

2.2. Subjects and procedure

Six normal subjects (one a deuteranope) were enrolled in the studies. All of them were experienced psychophysics observers, but all were naïve to the hypothesis being tested except for the author. Subjects either wore their prescribed optical correction or had normal uncorrected acuity. All had stereoacuity of 40 s or better on the Titmus Circles Test. All subjects experienced binocular rivalry when viewing the sieve stimulus. Some subjects required prompting before they perceived depth when first presented with the sieve stimulus, as was found by Howard (1995). All subjects practiced the depth judgments for a few sessions prior to collection of data. Two subjects did not achieve a robust depth percept for the sieve stimulus. They were discontinued in subsequent experiments and not included in the data presented.

Each trial was initiated by the subject after nonius markers were aligned. A random-check mask was shown for 200 ms, followed by the stimulus for 5000 ms, then the mask again. In preliminary trials, some subjects were unable to perform the task reproducibly with briefer presentation time. Subjects were allowed to free-view the stimulus, but nonius markers remained visible throughout stimulus presentation and subjects were instructed to use them to maintain alignment. The subject decided whether the depth defined by the collective pane interior was in front of or behind the comparison probes and was allowed to enter a response at any time after stimulus onset, which terminated the trial.

In all experiments, the perceived depth and depth increment threshold were measured by a 1-IFC method of constant stimuli. For each subject and stimulus combination, four to nine disparity values (usually six) were chosen to bracket the point where the depth probe changed from being in front of to behind the pane interior, as determined in pilot runs for each subject. Each disparity value, x, was tested 15 times per run. Each run typically included 180 trials randomized in which two or three conditions were interleaved. One or more runs were completed in a session with rest breaks. For each stimulus condition, four runs were collected and averaged. Because some disparity values that were tested differed from run to run, the number of judgments at each disparity value was not the same. Error bars in the figures indicate 95% confidence limits for the maximum likelihood estimate of the judgment probability $y(x)$, based on a Binomial distribution of responses.

Since the sieve stimulus might not contain robust cues to depth, we took measures to prevent subjects from using the depth of the probes as the sole basis for response. The concern was that, since the judging the depth of the sieve stimulus was difficult, subjects might have ignored the target and responded based only on

---

[1] Cross-talk measured via a nulling method at 1 Hz was undetectable. Based on psychophysical contrast sensitivity, the leakage contrast was less than 0.5%. Direct measurement of the cross-talk with a photometer was within the level of the instrument noise with a 100% contrast display.

the depth of the probes. We therefore introduced some catch trials in which the depth ordering of the probe and the target was the reverse of what would have been the response of a subject paying attention only to the probes. In a catch trial, the probes were rendered at the largest of the disparity values tested within that run, while the target, with the same contrast-polarity in both eyes, was at a disparity slightly larger than that of the probes. If a subject were simply making decisions based on the relative depths of the probes, then he/she would have responded incorrectly. All subjects performed nearly perfectly in the catch trials.

The measured probabilities, $y(x)$, of judging that the target was in front of a depth probe at disparity x (positive disparity is uncrossed), were fitted using a non-linear least square algorithm (MATLAB 5.1 with Statistics Toolbox 2.1.0, MathWorks) to a scaled cumulative normal function of the form

$$y(x) = \frac{c}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-(x'-\mu)^2/2\sigma^2}\,dx', \tag{1}$$

where μ and σ are the mean and standard deviation, respectively, of the underlying normal distribution. c is the ceiling on the proportion of ‘in front’ judgments. The curve fit was weighted by the number of repeats at each value of disparity tested.

We define: (1) the *perceived depth* as the value at which the function $y(x)$ crosses 0.5; (2) the *depth increment threshold* as $1/\sqrt{2\pi}$ times the reciprocal of the maximum slope of the function, i.e. σ/c. If $c = 1$, then perceived depth and increment threshold equal μ and σ, respectively. To convert increment threshold to the threshold change in depth that results in a change in response rate from 50 to 75%, when $c = 1$, we multiply it by 0.6745.

3. Results

3.1. Experiment I: precision of perceived depth in the sieve effect

We measured the depth increment thresholds and perceived depths of two stimuli: one composed of unmatched images (sieve stimulus) and the other of matchable images (control stimulus). Both stimuli had sixty 11 × 11 min panes, and were presented in separate runs. In the control-stimulus runs, the stimulus target was individualized for each subject, and chosen to have a disparity pedestal at least as large as the subject's perceived depth of the sieve stimulus. With increasing disparity pedestal, the threshold for depth discrimination obeys Weber's Law up to a point before it plateaus (McKee et al., 1990). Therefore, the threshold measured at the disparity pedestal chosen is an upper bound and provides a fair comparison with the threshold for the sieve stimulus. Note that one subject (FM) was tested with a different disparity pedestal (25 arcmin) from the rest (20 arcmin).

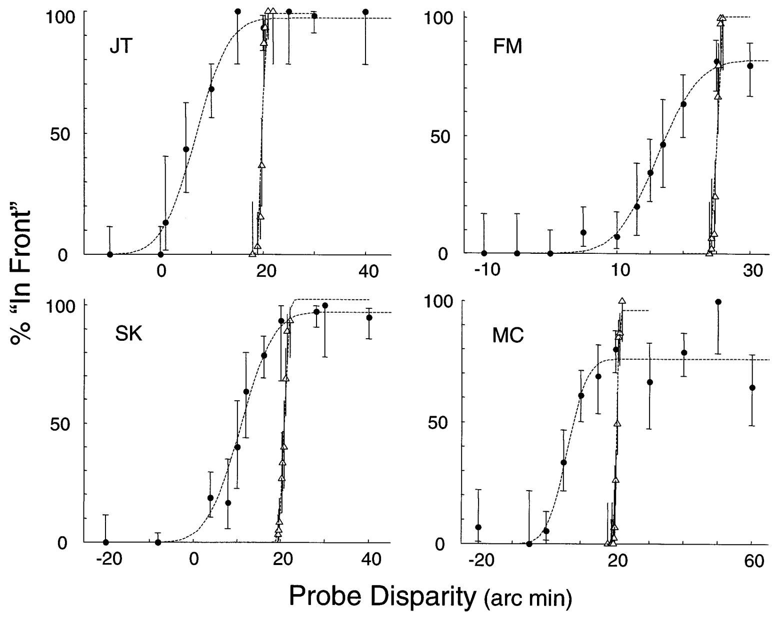

Fig. 3. Psychometric functions for depth comparisons in four subjects. Each graph shows two conditions of the target, a stereogram made of matchable images (△) with a particular non-zero interocular disparity (see text for details), or unmatched images (●) with zero disparity. Each data point represents the mean of 15–90 observations. Here and in subsequent figures, error bars show the 95% confidence intervals of the estimates of judgment probability based on a Binomial distribution of responses. The dotted lines show the best fit curve, weighted by the number of observations at each point. Positive values of disparity represent uncrossed depths.

Table 1
Parameter values of the best-fit curve (Eq. (1)) to the data in Fig. 3[a]

Subject		Perceived depth (arcmin)	Depth increment threshold (arcmin)	Ceiling
JT	Sieve	7.1 ± 0.9	5.4 ± 1.3	0.97 ± 0.05
	Control	19.9 ± 0.1	0.4 ± 0.1	0.99 ± 0.08
FM	Sieve	17.5 ± 1.1	6.0 ± 1.8	0.82 ± 0.08
	Control	25.0 ± 0.1	0.3 ± 0.1	1.01 ± 0.09
SK	Sieve	11.0 ± 0.8	6.2 ± 1.2	0.97 ± 0.05
	Control	20.8 ± 0.1	0.7 ± 0.1	1.02 ± 0.19
MC	Sieve	7.9 ± 1.3	6.1 ± 2.1	0.76 ± 0.05
	Control	20.6 ± 0.1	0.4 ± 0.1	0.96 ± 0.10

[a] Ranges given are the 95% confidence intervals of the parameter estimates.

Psychometric functions for four subjects are shown in Fig. 3. For most of the data, a ceiling c near 1 turned out to be the best fit for both sieve and control. In no case could differences in the ceiling account for the difference in slope or threshold of the psychometric curves. The most striking aspect in the results for all subjects is the difference in the steepness of the psychometric function between the sieve and the control stimuli. The ratio of depth increment thresholds for the two stimuli ranges between 10- and 20-fold (Table 1). Indeed, at scales that provide useful renderings of the psychometric functions for the sieve stimulus, the psychometric functions for the control stimulus resembled a step function (because the value of the parameter σ was very small). This difference in slope was corroborated by subjects' informal reports that the control stimulus trials were much easier, and the confidence in their responses was much higher, compared to the sieve stimulus. The imprecision found with the sieve stimulus cannot be attributed to subjects' inability to perceive depth, for in that case, one would expect a flat psychometric function. Moreover, re-testing of one subject (FM) after a 3-month hiatus found a similar precision and perceived depth, indicating that the subject was not making responses based on short-term recall.

The thresholds for the control stimuli with a pedestal of 20 min, expressed in terms defined by McKee et al. (1990), ranged from 15 to 24 arcsec for our subjects. As a comparison, McKee et al. (1990) found that the threshold was ca. 60 arcsec at a standing disparity of ca. 20 min with a 1000 ms presentation time. They also found that the threshold improved substantially when presentation duration was lengthened from 150 to 1000 ms. Thus, the long presentation time of 5000 ms in our experiment may account for our smaller threshold values. In addition, stimulus differences may contribute to the difference. McKee et al. (1990) used line targets to measure disparity threshold while we used stereograms depicting surfaces.

Once a subject achieved a stable depth percept, he or she invariably saw the sieve stimulus as behind the fixation plane. This was a strong and consistent bias; the fractions of 'in front' responses for all crossed-disparity probes — i.e. for values of the disparity less than zero — were essentially zero. The slight deviation of one subject (MC) from this result was probably due to incomplete performance stabilization. The perceived depth of the sieve stimulus varied from subject to subject, but in all cases was significantly further than the fixation plane (Table 1).

The results of this experiment show that unmatched images generate a systematic uncrossed depth percept, but one that is less robust and less precise than the depth percept produced by matchable images.

3.2. Experiment II: consistency of perceived depth in the sieve effect with occlusion constraints

Experiment I showed that the depth perceived in the sieve effect was behind the fixation plane. In random-dot and figural stereograms, unpaired monocularly-viewed points are always perceived as the far surface (Julesz, 1971; Nakayama & Shimojo, 1990a), which is ecologically valid since a near surface is always visible to both eyes. In the same vein, Howard (1995) pointed out that the retinal images generated by the sieve stimulus are similar to those generated by viewing a randomly checkered surface through apertures in an occluding surface. That is, the ecological optics of occlusion *can* generate retinal images equivalent to the sieve stimulus (Fig. 1). Experiment II addresses whether the perceived depths associated with these stimuli are in fact consistent with occlusion in quantitative detail.

In this experiment, two sieve stimuli for comparison were randomly interleaved within a test run. Both stimuli had 40 panes of 16 by 11 min, but differed in the orientation of the panes: vertical (portrait) or horizontal (landscape). The psychometric functions for both sieve stimuli are compared in Fig. 4. Three subjects (JT, FM, and SK) perceived a significantly larger depth in the landscape compared to the portrait orientation. The fourth subject did not show a significant difference

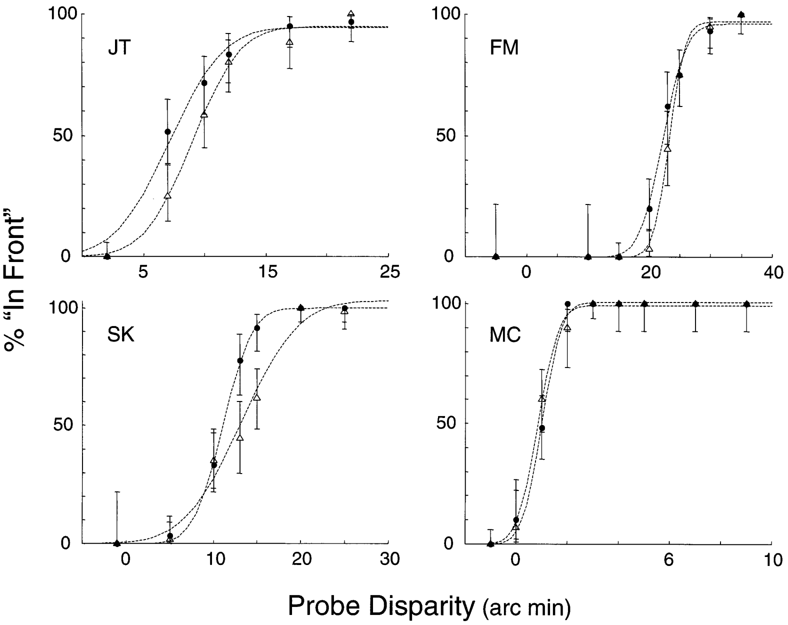

Fig. 4. Effect of shape of sieve pane on perceived depth. Each graph shows psychophysical curves derived from two sieve targets made up of 40 16 by 11 min panes, either in landscape (△) or portrait (●) orientation. Each data point represents the mean of 15–60 observations. The perceived depths and 95% CI (arcmin) are, for the portrait and landscape orientations respectively: (JT) 7.4 ± 0.6 and 9.2 ± 0.5; (FM) 22.4 ± 0.5 and 23.4 ± 0.3; (SK) 11.1 ± 0.4 and 13.0 ± 0.7; (MC) 1.0 ± 0.1 and 0.9 ± 0.1.

between the two stimuli; however, she now saw a much smaller absolute depth for both sieve stimuli compared to the results in Experiment I (completed approximately 4 months earlier). Unlike the other three subjects, her use of this depth cue has apparently changed radically with learning.

Fig. 5 summarizes the effect of the width of the panes on the perceived depth in the sieve stimulus. Ecological optics predicts that the perceived depths should be no less than the width of the pane apertures, i.e. all points should lie above the diagonal in Fig. 5. Clearly, data from three of four subjects failed to obey this prediction. Moreover, the deviation from the EO prediction is not simply a constant offset from the minimal depth, as the slopes of the plotted data differ from that of the diagonal (Fig. 5). As the width of the panes increases, the perceived depth also increases (three of four subjects), but by an amount less than expected from EO. However, concomitant with the increase in width, the height of the panes decreases by the same amount. The next experiment will look at whether the vertical extent may have an effect on the perceived depth as well. In any event, since the vertical extent of the aperture is not expected to affect the depth according to the EO framework, this experiment shows that the sieve effect cannot be completely accounted for by the geometry of real world occluding surfaces for two reasons: (1) The perceived depth is less than the minimal depth consistent with occlusion; and (2) The dependence of the perceived depth on pane width is weaker than expected from EO.

3.3. *Experiment III: dependence of the sieve effect on vertical dimension*

The results so far indicate that the depth arising from unmatched images in the sieve effect is quite different from classical stereopsis in its precision (Experiment I) and also different in magnitude from the prediction based on occlusion (Experiment II). Neither of these explanations predicts a dependence of depth on the

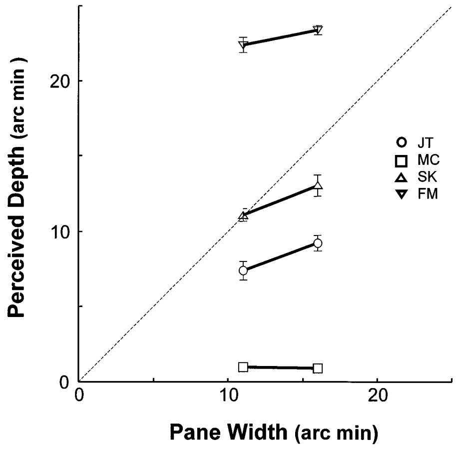

Fig. 5. Summary of data shown in Fig. 4. The perceived depth from the best-fit curve is plotted against the width of the panes for the two stimulus conditions, portrait and landscape. Error bars show the 95% confidence intervals of the depth estimates. The error bars for subject MC are smaller than the size of the plot symbol. The diagonal line represents the minimum depth that would be consistent with the ecological optics of occlusion.

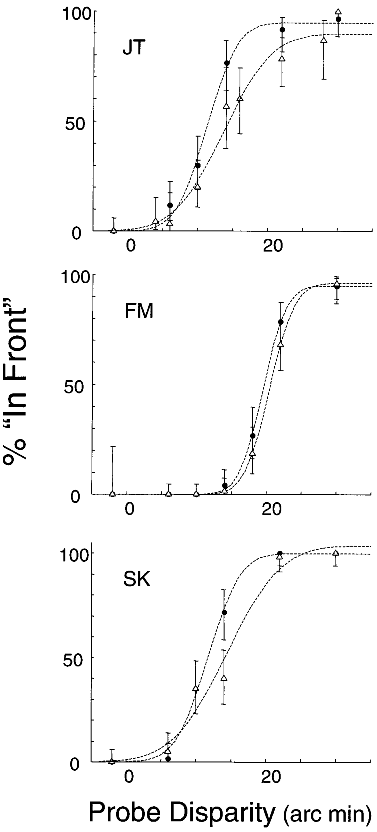

Fig. 6. Effect of height of sieve pane on perceived depth. Sieve stimuli contain panes of the same width but different height — namely, 16 × 26 min (△) or 16 × 7 min (●). Each data point represents the mean of 15–75 observations. The perceived depths and 95% CI (arcmin) are, for the 16 × 7 and 16 × 26 conditions respectively: (JT) 11.5 ± 0.6 and 14.4 ± 1.0; (FM) 19.7 ± 0.4 and 20.6 ± 0.4; (SK) 11.7 ± 0.5 and 14.1 ± 0.9.

vertical extent of the unmatched images; that is, within the sieve stimulus, changes in the pane height do not affect the horizontal disparity cue nor the minimum depth of the occlusion constraint. To further study the relevance of disparity and occlusion to the sieve effect, we examined the effect of the vertical dimension of the panes on the perceived depth. In this experiment, two sieve stimuli having the same pane widths but different heights were compared. Each trial presented either the short panes (16 × 7 min) or the tall panes (16 × 26 min), randomly interleaved in each run. Three subjects participated. For FM and SK, both stimuli contained 25 panes. For JT, the stimulus was of either 70 small or 30 large panes. Psychometric functions for these three subjects are shown in Fig. 6.

The smaller pane height caused the perceived depth to move closer to the fixation plane for all subjects. The differences — from 1 to 3 arcmin — were statistically significant, but less than the change in pane height. There was no significant difference in the slope of the psychometric curves in two of three subjects (JT and FM).

Together with Experiment II, these results show that the perceived depth of the sieve stimulus depends on both spatial dimensions, but apparently more on the horizontal than the vertical. Neither the matching/disparity nor the EO explanation alone can fully account for this kind of dependence. One possibility is that even though panes were always less than 1° in size, and hence in the regime where alternating exclusive rivalry is found, the depth depends on the size or shape of the region engaged in rivalry. This influence on perceived depth might coexist with a cue related to the horizontal dimension. Thus, in Experiment III, the effect of the change in the vertical extent of the pane might be due to a change in its area. In Experiment II, the area of the pane was not varied and perceived depth changed, but not in accord with the extent of the horizontal cue. To account for these results, one would have to postulate either that there is a depth cue related to the vertical extent per se, or that the apparent effect of vertical extent is mediated by binocular rivalry, and the strength of the latter is in turn dependent not only on the size of the pane, but also its shape. In any case, the dependence of perceived depth on the dimensions of the pane is not simply consistent with occlusion.

4. Discussion

4.1. Summary of results

We report a quantitative analysis of the depth percept elicited by a family of stimuli that have interocular differences but no well-defined interocular disparity cue. Several aspects of our results — the precision of the depth percept and the dependence of its magnitude on the horizontal and vertical extent of the unmatched regions — are at odds with two current computational frameworks for stereopsis, namely, positional disparity computation based on matchable image features, and recovery of occlusion relationships from unmatched features. Experiment I showed that the depth increment thresholds based on horizontal disparity cue were an order of magnitude better than those achievable with the sieve stimulus. This is consistent with the results of Cogan, Kontsevich, Lomakin, Halpern and Blake (1995), who measured stereoacuity using bar stereograms and found that relative depths of same-contrast half images could be detected better than reversed-contrast images by an order of magnitude. A somewhat smaller difference (0.5 log units) was found by Cogan, Lomakin and Rossi (1993) using a figural stereogram depicting a pair of dots. This dramatic difference in precision suggests that the computations underlying these two depth processes are likely to be distinct. Experiment II showed that the perceived depths from our unmatched-image stereograms were dependent on the widths of the unmatched images, but in a manner inconsistent with the ecological optics of occlusion. In light of this discrepancy, the notion that perceived depth is based on the real world constraint of surface occlusion appears inadequate to explain this perceptual phenomenon. Experiment III showed that the perceived depths were dependent on the vertical extent of the unmatched images as well. This is counter to predictions based on either of the above two frameworks.

4.2. Interocular matching

Three additional observations suggest that depth perceived in the sieve effect is not driven by horizontal disparity cue from binocular matching. Since the eye-of-origin of a classical stereogram carries the sign of the depth information (Nakayama & Shimojo, 1990b), if some unintended horizontal disparity cue were the basis for the sieve effect, then the sign of perceived depth should reverse upon switching the half images. However, the luminance polarities of the panes are assigned randomly before each trial (so the assignment of the half images to the two eyes is random), but the sign of the depth percept is remarkably consistent. Secondly, even if conventional Wheatstone stereopsis can be driven by 'matching' primitives that are opposite in contrast (Helmholtz, 1909/1962 cited in Cogan et al., 1995; Kaufman & Pitblado, 1969; but see Treisman, 1962; Levy & Lawson, 1978; Cogan et al., 1993), the resulting disparity in the sieve stimulus is zero and cannot produce the depth percepts we found. Finally, the necessity of a longer time for depth perception in the sieve stimulus, compared to < 50 ms for simple random-dot stereograms (Julesz, 1964), suggests that a different mechanism may be involved. These observations further support the hypothesis that horizontal disparity matching per se cannot account for the sieve effect.

4.3. Role of vergence

We consider, but question another possible explanation for the sieve effect, which is based on induced depths resulting from subjects' fixation disparity. Fixation disparity refers to the slight misconvergence of the eyes causing intended fixation targets to have a nonzero disparity. If unmatched monocular features default to the depth of the fixation plane (Howard & Rogers, 1995; McKee, Bravo, Smallman & Legge, 1995; Smallman & Mckee, 1995), then in the presence of fixation disparity, they would be perceived at a different depth relative to the surrounding matchable image points. Some studies have found a relationship between fixation disparity and the perceived depth of unmatched stereograms. O'Shea and Blake (1987) reported that depth was obtained from a random-dot correlogram that had an uncorrelated center against a correlated surround, both without positional disparity. Near depths were reported by subjects who exhibited under-convergence, while far depths were seen by those with over-convergence. These investigators hypothesized that for a subject who habitually under-converges, the failure to bring the half images into registration suggests to the visual system that the depth lies in the opposite direction. We do not believe fixation disparity can fully explain the sieve effect for the following reasons: First, O'Shea and Blake found that among their subjects, near and far depths were preferred equally. Furthermore, although subjects showed a general preference for one direction, fewer than 10% of their 52 subjects had a uniform response to all trials. In contrast, the sieve effect was consistently seen as behind the correlated surround. Secondly, from subjects' verbal reports on monitoring the nonius markers, the vergence changes were not large enough to account for the magnitude of the perceived depths. Finally, it is unclear how misconvergence could account for the effect of pane height on perceived depth.

4.4. Relationship to physiological modeling

We begin by emphasizing that motivation in these studies is to advance the understanding of the computations underlying perception, not to address its 'purpose'. Our distinction between the *purpose* and the *process* of vision follows Dodwell (1975): the former is generally clear from behavioral considerations and not altered by changes in understanding of the latter. In particular, the utility of apprehending occlusion relationships remains, even if occlusion geometry does not form the basis of the underlying computational mechanism. Moreover, postulating that visual system responds to unmatched stereograms insofar as they are consistent with the real world constitutes in essence a statement of purpose, but does not directly bear on the question of process. Ultimately, we are interested in the neural mechanism by which the visual system calculates depth. The results presented here provide some new constraints for computational models and their neural implementation. We will show that the current physiological model of stereo perception does not meet these constraints. A model that will account for both these and conventional depth stimuli is currently under investigation.

Like classical stereopsis, the depth from 'unmatched images' depends inherently on the comparison of the input from both eyes. This process requires that the two monocular visual input be distinguishable. Since monocular neural signals remain segregated before reaching the striate cortex, where they converge onto single neurons, the computation that underlies this interocular comparison most likely begins at this stage. For this reason, our consideration of physiologically plausible mechanisms begins with the available binocular cells in the striate cortex. (However, we note that it is possible that depth computations relying on monocular information take place in other visual areas as well, since some V2 neurons receive monocular input (Burkitt & Ts'o, 1999), and the presence of disparity information in later visual areas (Poggio, 1990; DeAngelis, Cumming & Newsome, 1998) implies that signals from the two eyes are treated distinctly, even if not kept separate.)

At present, the best-defined physiological model is based on data from simple and complex cell responses in primary visual area of the cat (DeAngelis, Ohzawa & Freeman, 1991; Ohzawa et al., 1990). When studied with stimuli that vary only in one dimension (horizontal), the responses of these cells are reasonably well-described by a so-called 'energy' model (reviewed in Ohzawa, 1998). The energy model produces a stimulus-phase independent measure of disparity, by squaring and summing outputs from a quadrature pair of linear binocular filters. The left and right monocular stages of the filter may differ (in spatial position, phase, or a combination), resulting in an optimal binocular response to a stimulus disparity that produces the smallest interocular phase difference in the monocular outputs (Fleet, Wagner & Heeger, 1996). Moreover, in response to a one-dimensional sinusoidal stimulus, the output of a quadrature unit is a cosinusoidal function of stimulus disparity with a frequency equal to the stimulus frequency (Qian, 1994; Fleet et al., 1996). Thus the responses of a disparity energy neuron ('quadrature unit') to a given stimulus disparity depends on its receptive field organization and frequency selectivity. From a population of such quadrature units, in one implementation, 'disparity' is computed from the receptive field parameters of the maximally responsive unit (Qian, 1994). For a stimulus with the same contrast polarity in both eyes and well-defined disparity, this method gives an unambiguous answer, provided that the stimulus disparity is restricted to the range $-\pi/\omega_0$ to π/ω_0 where ω_0 is the preferred spatial frequency of the unit (Qian, 1994; Qian & Zhu, 1997). Alternatively, pooling the responses of units of different orientation and spatial scale preference also reduces the ambiguity in the disparity measure (Fleet et al., 1996). The energy method of disparity computation has been tested on random-dot stereograms (reviewed in Qian, 1997), and performs well on these stimuli.

Fleet et al. (1996) showed that the sinusoidally modulated component of the response of a quadrature unit is similar to a cross-correlation operation on the output of the intermediate monocular stages. Hence the energy model predicts that if the contrast polarity of one of the half images is reversed, the disparity tuning of each energy unit will invert also (Ohzawa et al., 1990; Eq. 2.13 of Qian, 1994; Cumming & Parker, 1997; Qian & Mikaelian, 2000). That is, the disparity-tuning function will be shifted by one half of the preferred spatial period of the unit. We can consider the sieve stimulus as a pair of images with zero disparity and locally inverted interocular contrast polarity. In qualitative terms, then, this stimulus should maximally drive two groups of binocular cells — those that have a peak disparity tuning, in response to matched-contrast stimuli, at half of a cycle (comparable to the width of the panes) on either side of fixation. Thus, the existence of a consistent, but less precise, depth percept reflects how these ambiguous signals are synthesized into a 'disparity' signal. In particular, this bimodal population response results in a perceptually unambiguous sense of uncrossed depth. This perceptual outcome is not readily understood in terms of a winner-take-all approach to reading out the population activity, even though the activity of the population of binocular neurons necessarily contains the information needed to account for the perceived depth in the sieve stimulus.

We also found that the magnitude of the depth percept depended on the vertical extent of the panes. This seems to require another kind of elaboration on the current model. Since the model is one-dimensional, changing the vertical extent of the stimulus might alter the size of the response of disparity-sensitive simple and complex cells, but it would not influence which cells have the largest response. Thus, 'disparity', as read out by the tuning of the cells that respond the best, would not be altered by the vertical extent of the stimulus. To account for the observed dependence of perceived depth on the vertical extent of the panes, the fundamental disparity-sensitive computational element must make use of the vertical dimension in a subtle way.

In summary, the results reported here further characterize the nature of depth perception from unmatched images. This percept cannot be readily explained by conventional disparity nor by ecological occlusion. Strict implementation of current neurally-based computational schemes also cannot account for depth in the sieve effect. Nevertheless, it may be possible to account for this depth percept with the same neural elements that are used for classical stereopsis, provided that both the read-out and the front end of the latter process are suitably elaborated.

Acknowledgements

We thank Mary Conte for assistance with and participation in the experiments and for help with the preparation of the manuscript, Ning Qian for helpful comments on an earlier draft of this manuscript, and Steve Kalik and Ferenc Mechler for participating as observers. J.T. thanks Aries Arditi for discussion on stereo display and preliminary results. A portion of this work was presented at the 1999 meeting of the Association for Research in Vision and Ophthalmology in Ft. Lauderdale, FL (Tsai & Victor, 1999). This work was supported by NIH grant EY7977 (J.D.V.) and Medical-Scientist Fellowships from the Lili Foo-Hing Li, the Andrew W. Mellon, and the Iris L. and Leverett S. Woodworth Foundations (J.J.T.).

References

Anderson, B. L. (1994). The role of partial occlusion in stereopsis. *Nature*, *367*, 365–368.

Barlow, H. B., Blakemore, C., & Pettigrew, J. D. (1967). The neural mechanisms of binocular depth discrimination. *Journal of Physiology*, *193*, 327–342.

Blake, R., & Wilson, H. R. (1991). Neural models of stereoscopic vision. *Trends in Neuroscience*, *14*, 445–452.

Burkitt, G. R., & Ts'o, D. Y. (1999). Dependence of V2 disparity processing on stimulus axis and V1 monocular inputs in the macaque monkey. *Society for Neuroscience Abstracts*, *25*, 2060.

Cogan, A. I., Lomakin, A. J., & Rossi, A. F. (1993). Depth in anticorrelated stereograms. *Vision Research*, *33*, 1959–1975.

Cogan, A. I., Kontsevich, L. L., Lomakin, A. J., Halpern, D. L., & Blake, R. (1995). Binocular disparity processing with opposite-contrast stimuli. *Perception*, *24*, 33–47.

Cumming, B. G., & Parker, A. J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature*, *389*, 280–283.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1991). Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature*, *352*, 156–159.

DeAngelis, D. C., Cumming, B. G., & Newsome, W. T. (1998). Cortical area MT and the perception of stereoscopic depth. *Nature*, *394*, 677–680.

Dodwell, P. D. (1975). Contemporary theoretical problems in seeing. In E. C. Carterette, & M. P. Friedman, *Handbook of perception. Seeing*, vol. V (pp. 67–75). New York: Academic Press.

Fielder, A. R., & Moseley, M. J. (1996). Does stereopsis matter in humans? *Eye*, *10*, 233–238.

Fleet, D. J., Wagner, H., & Heeger, D. J. (1996). Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Research*, *36*, 1839–1857.

Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton-Mifflin.

Gillam, B. (1995). Matching needed for stereopsis. *Nature*, *373*, 202–203.

Gillam, B., & Nakayama, K. (1999). Quantitative depth for a phantom surface can be based on cyclopean occlusion cues alone. *Vision Research*, *39*, 109–112.

Gillam, B., Blackburn, S., & Cook, M. (1995). Panum's limiting case: double fusion, convergence error, or 'da Vinci stereopsis'. *Perception*, *24*, 333–346.

Helmholtz, H. von (1909). *Physiological optics*. New York: Dover (J. P. C. Southall (1962), Trans. from *Handbuch der Physiologischen Optik* (3rd ed.). Hamburg: Voss).

Howard, I. P. (1995). Depth from binocular rivalry without spatial disparity. *Perception*, *24*, 67–74.

Howard, I. P., & Rogers, B. (1995). *Binocular vision and stereopsis* (pp. 518–523). Oxford: Oxford University Press.

Hubel, D. H., & Wiesel, T. N. (1970). Stereoscopic vision in macaque monkey. Cells sensitive to binocular depth in area 18 of the macaque monkey cortex. *Nature*, *225*, 41–42.

Julesz, B. (1964). Binocular depth perception without familiarity cues. *Science*, *145*, 356–362.

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago: University of Chicago Press.

Kaufman, L., & Pitblado, C. B. (1969). Stereopsis with opposite contrast contours. *Perception & Psychophysics*, *6*, 10–12.

Levy, M. M., & Lawson, R. B. (1978). Stereopsis and binocular rivalry from dichoptic stereograms. *Vision Research*, *18*, 239–246.

Liu, L., Stevenson, S. B., & Schor, C. W. (1994). Quantitative stereoscopic depth without binocular correspondence. *Nature*, *367*, 66–68.

Liu, L., Stevenson, S. B., & Schor, C. W. (1997). Binocular matching of dissimilar features in phantom stereopsis. *Vision Research*, *37*, 633–644.

Marr, D. (1982). *Vision*. New York: Freeman.

McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research*, *30*, 1763–1779.

McKee, S. P., Bravo, M. J., Smallman, H. S., & Legge, G. E. (1995). The 'uniqueness constraint' and binocular masking. *Perception*, *24*, 49–65.

Nakayama, K., & Shimojo, S. (1990a). da Vinci stereopsis: depth and subjective occluding contours from unpaired image points. *Vision Research*, *30*, 1811–1825.

Nakayama, K., & Shimojo, S. (1990b). Toward a neural understanding of visual surface representation. *Cold Spring Harbor Symposia on Quantitative Biology*, *55*, 911–924.

Nakayama, K, & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, *257*, 1357–1363.

Ogle, K. N. (1953). Precision and validity of stereoscopic depth perception from double images. *Journal of the Optical Society of America*, *43*, 906–913.

Ohzawa, I. (1998). Mechanisms of stereoscopic vision: the disparity energy model. *Current Opinion in Neurobiology*, *8*, 509–515.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science*, *249*, 1037–1041.

O'Shea, R. P., & Blake, R. (1987). Depth without disparity in random-dot stereograms. *Perception & Psychophysics*, *42*, 205–214.

Panum, P. L. (1858). *Physiologische Untersuchungen über das Sehen mit zwei Augen*. Keil: Schwers.

Poggio, G. F. (1990). Cortical neural mechanisms of stereopsis studied with dynamic random-dot stereograms. *Cold Spring Harbor Symposium on Quantitative Biology*, *55*, 749–758.

Poggio, G. F., & Poggio, T. (1984). The analysis of stereopsis. *Annual Review of Neuroscience*, *7*, 379–412.

Qian, N. (1994). Computing stereo disparity and motion with known binocular cell properties. *Neural Computation*, *6*, 390–404.

Qian, N. (1997). Binocular disparity and the perception of depth. *Neuron*, *18*, 359–368.

Qian, N., & Zhu, Y. (1997). Physiological computation of binocular disparity. *Vision Research*, *37*, 1811–1827.

Qian, N., & Mikaelian, S. (2000). Relationship between phase and energy methods for disparity computation. *Neural Computation*, *12*, 279–292.

Smallman, H. S., & Mckee, S. P. (1995). A contrast ratio constraint on stereo matching. *Proceedings of the Royal Society of London*, *B*, *260*, 265–271.

Shimojo, S., & Nakayama, K. (1990). Real world occlusion constraints and binocular rivalry. *Vision Research*, *30*, 69–80.

Treisman, A. M. (1962). Binocular rivalry and stereoscopic depth perception. *Quarterly Journal of Experimental Psychology*, *14*, 23–37.

Tsai, J., & Victor, J. D. (1999). The precision of depth from unpaired image points, *Investigative Ophthalmology and Visual Science*, *40* (Suppl.), 802 (Abstract).

Weinshall, D., & Malik, J. (1995). Review of computational models of stereopsis. In T. V. Papathomas, C. Chubb, A. Gorea, & E. Kowler, *Early vision and beyond* (pp. 33–42). Cambridge: MIT Press.

Wheatstone, C. (1838). Contributions to the physiology of vision. I. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, *2*, 371–393.