PERGAMON

Vision Research 42 (2002) 1225–1235

Vision Research

www.elsevier.com/locate/visres

Binocular integration of partially occluded surfaces

Jason Forte *, Jonathan W. Peirce, Peter Lennie

Center for Neural Science, New York University, Room 809, 4 Washington Place, New York, NY 10003, USA

Received 20 August 2001; received in revised form 19 February 2002

Abstract

Normal binocular vision can provide a view of an object partially occluded so that no part of it is seen by both eyes but all of it is seen by one or other eye. We used two-dimensional filtered noise textures to explore the conditions under which the visual system can piece together the monocular fragments of such occluded surfaces. When the fragments seen by left and right eyes are drawn from a continuous texture with strong horizontal correlation, observers see coherent surfaces reliably located in depth. When textures are discontinuous or have weaker horizontal correlation, or the left and right eyes' views represent unnatural depth relationships, no coherent surface is perceived, and binocular rivalry ensues. The discovery of coherent surfaces under our conditions seems to reflect the operation of a high-level integration process, failures of which drive rivalry. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Occlusion; Depth; Binocular vision; Surface perception

1. Introduction

When objects lie at a range of distances from an observer, and occlude one another, the separation of the eyes ensures that each eye sees parts of the world that are invisible to the other. Partial occlusion can result in corresponding regions of the two eyes' images being completely different, yet this conflict does not intrude on normal vision, and we are generally unaware of it. In the laboratory, superficially similar conflicts between image regions give rise to binocular rivalry, and characteristically unstable perception.

The potential importance of the partially occluded regions ("monocular zones" of Howard & Rogers (1995, p. 512)) has been highlighted by earlier work (e.g. Julesz, 1964) and explored with displays like those in Fig. 1, which depict a surface behind an aperture (Fig. 1A), or standing out against a background (Fig. 1B). Shimojo and Nakayama (1990) showed that the monocular zone tends to prevail in a potentially rivalrous conflict with the image in the other eye. This is ecologically appropriate. Gillam and Borsting (1988) and Nakayama and Shimojo (1990) showed that the monocular zone can help identify depth discontinuities; Liu, Stevenson, and Schor (1994), Anderson (1994) and Gillam and Nakayama (1999) showed that relative depth can be recovered from occlusion cues alone. Anderson and Nakayama (1994) demonstrated that occlusion cues influence stereo matching, and developed a physiological model to explain how occlusion contours are detected and depth relations assigned.

In cases of the kind illustrated in Fig. 1 the image fragments in the monocular zones are generally treated as though they lay in the plane of the adjacent (binocularly seen) rear surface (Julesz, 1964; Anderson & Nakayama, 1994; Hakkinen & Nyman, 2001), but this has not been explored experimentally, and no account has been offered of how it might come about. Moreover, real-world occlusion can give rise to image fragments in monocular zones that have no neighboring binocular regions. These problems are framed sharply by considering the case of an object viewed through a fence of vertical bars (Fig. 2). The bars occlude the face such that no part of it is visible to both eyes, but all of it is visible to one or other the eye (Fig. 2a). Fig. 2b and c show respectively the left and right eyes' views, and Fig. 2d shows the reconstructed composite image.

Can the fragments seen by the two eyes be pieced together to give coherent percepts of extended surfaces? Fig. 3 provides a demonstration that suggests an answer.

---

[*] Corresponding author. Tel.: +1-212-998-3529.
E-mail address: jay@cns.nyu.edu (J. Forte).

PII: S0042-6989(02)00053-6

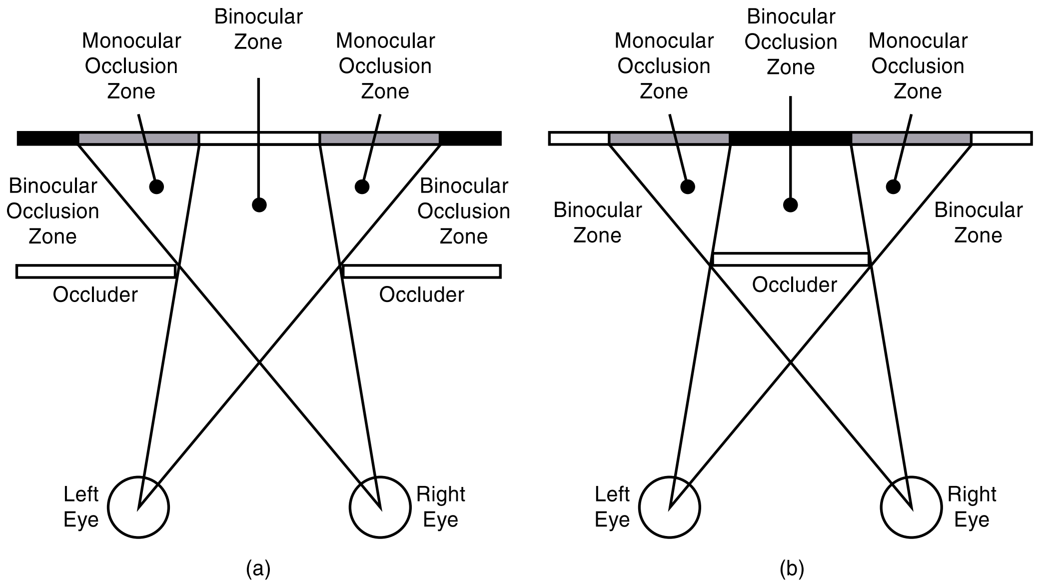

Fig. 1. Surface regions visible to only one eye (monocular zones) arise through partial occlusion of the binocular view of the surface by another lying in front of it. a. Occlusion by a continuous surface creates separated monocular zones at its left and right borders in which the surface behind is visible to left and right eyes respectively. b. Occlusion by an aperture creates separated monocular zones at is left and right edges visible to right and left eyes respectively. In both a and b the monocular zones have neighboring regions of surface that are visible binocularly.

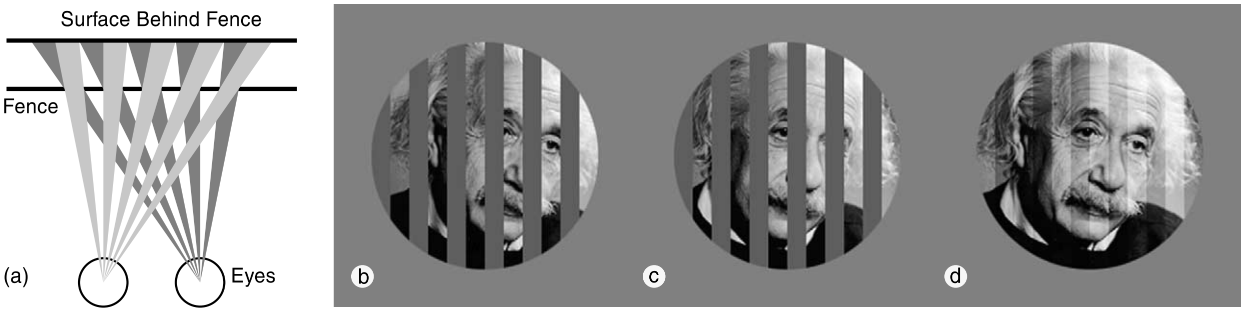

Fig. 2. Partial occlusion of a binocularly viewed surface can result in no part of the surface being visible to both eyes while all of it is visible to one or the other eye. a. Schematic showing the two eyes' views. b and c. Left and right eyes' views. d. Composite view that would result from piecing together the left and right eyes' views.

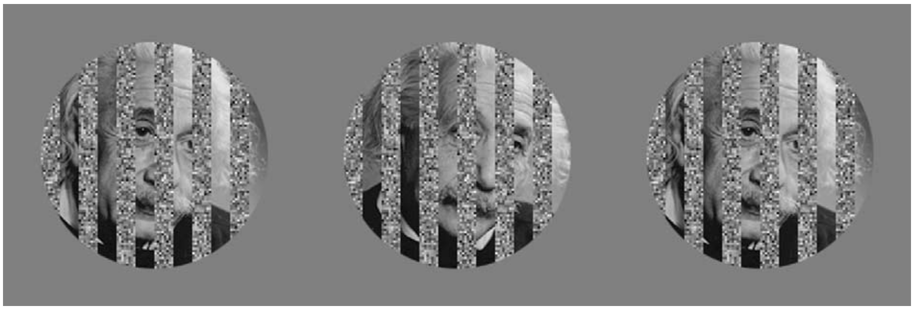

Fig. 3. A coherent surface, partially occluded so that no part is visible to both eyes, can be pieced-together from the image fragments seen by each eye. Uncrossed fusers should view left and center images; crossed fusers should view center and right images. When depth relations are made inconsistent with real world occlusion by reversing the left and right eyes' views (uncrossed fusers view center and right images; crossed fusers view left and center images) the coherent surface can no longer be perceived.

When the left and right eyes' views are consistent with Einstein behind bars the separate image fragments can be stitched together to yield a percept of a continuous object behind the bars, but when the images are reversed (consistent with Einstein in front of the bars, but ecologically invalid), Einstein appears jumbled and the fragments between the bars become rivalrous.

What enables the fragments to be pieced together to yield a stable percept, where do the resulting surfaces appear in depth, and what mechanism provides the protection against rivalry when the depth relations are sensible, but not otherwise? We might not be especially surprised that coherent percepts of familiar images can be pieced together, because (like Einstein) they can be recognized from monocular fragments alone. The interesting cases to examine are those in which the image fragments represent nothing recognizable, and in which we can control the correlation between the two eyes' views. We have therefore explored how the visual system deals with partially occluded surfaces comprised of two-dimensional filtered noise.

2. Methods

2.1. Stimuli

We wanted displays in which the surfaces to be occluded were unidentifiable, and which allowed control of the correlation between the statistical structure of neighboring regions. We created such surfaces from two-dimensional spatial white noise, filtered to control the dominant spatial frequency and orientation. The noise was generated by randomly assigning each pixel a luminance from the 256 equally spaced luminance values available. All luminances were equally likely. Noise arrays were Fourier transformed and the resulting amplitude spectrum was weighted in frequency and orientation by a Gaussian filter of the form

$$w_{f\phi} = \exp\left[\frac{-(f - f_{\mathrm{p}})^2}{2\sigma_f^2}\right] \cdot \exp\left[\frac{-(\phi - \phi_{\mathrm{p}})^2}{2\sigma_\phi^2}\right];$$

where $w_{f\phi}$ is the weight of the amplitude spectrum at spatial frequency f and orientation $f\phi$, f_{p} is the peak frequency, σ_f is the standard deviation of the frequency, ϕ_{p} is the peak orientation, and σ_ϕ is the standard deviation of the orientation. This spectrum was then recombined with the phase spectrum and inverse transformed to create textures of the kind shown in Fig. 4a–c.

We used filters with peak frequencies of 1.9, 3.8 and 7.6 c/deg. The standard deviation of each filter was set to half the peak frequency. The peaks of the orientation filters ranged from 0 to $\pi/2$ radians in $\pi/12$ radian steps, with a standard deviation of $\pi/8$ radians. Filtered

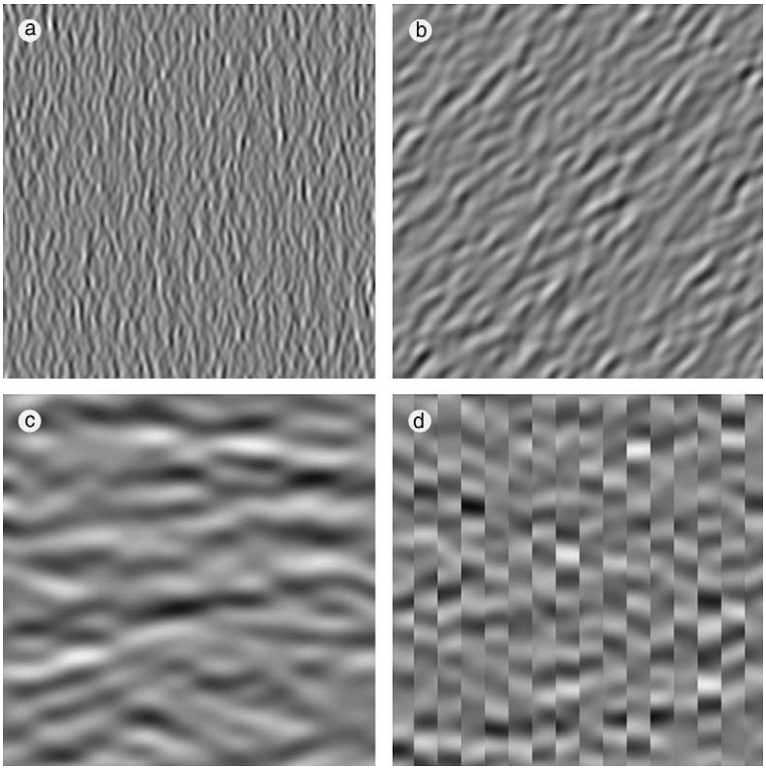

Fig. 4. Filtered two-dimensional noise textures from which surfaces were constructed. a–c. Binocularly continuous surfaces with increasing degrees of correlation in the horizontal meridian. a. Vertical texture of high spatial frequency. b. Oblique texture of intermediate spatial frequency. c. Horizontal texture of low spatial frequency. d. Binocularly discontinuous surface made from horizontal textures of low spatial frequency. Strips seen by left and right eyes were drawn from two independently-generated textures with the same statistics.

textures were then normalized to maximize the contrast range while maintaining the same mean luminance (70 cd/m^2). Our set of twenty one textures consisted of three scaled versions of statistically identical textures at seven different orientations.

Textures were masked by a 3.7° circular aperture with a uniform surround of the same mean luminance (70 cd/m^2). Textures were displayed on the left and right halves of a calibrated Sony G400 CRT monitor (flat screen) and viewed through a mirror stereoscope. The screen displayed 1024×768 pixels, providing each eye with a field that extended 8.5° horizontally by 12.8° vertically. The aperture and surround lay in the plane of fixation (zero disparity); the texture was given a disparity of $16'$ to make the surface it represented appear behind the aperture.

Vertical bars $16'$ wide and spaced $16'$ apart were placed in the aperture to occlude parts of the texture. The bars lay in the plane of the aperture. By having width and spacing equal to the disparity of the texture, the bars exposed completely separate strips of the texture to each eye, while making the whole texture surface potentially available in a binocular view (as in Fig. 2a). In most experiments the bars were filled with two dimensional white noise elements subtending 2 min of arc. We established through preliminary observations that the bar width and spacing were not important, as long as the displays contained several bars and the noise elements on the bars were clearly discernible. The noise had a mean luminance equal to that of the occluded texture and the region enclosing the aperture, and was correlated for the two eyes. The noise made each bar unique, and so ensured that only a single binocular match could be made for the array of bars. In preliminary experiments using plain bars we found that observers could sometimes experience the 'wallpaper' illusion as a result of matching the wrong bars.

By varying the dominant spatial frequency and orientation of the texture exposed through the bars, we could alter the degree of horizontal continuity (or predictability) in the image, and thus the extent to which what was seen by one eye predicted what was seen by the other. A vertically-oriented texture containing spatial frequencies that were high in relation to the width of the bars had the least continuity in horizontal structure; a texture containing coarser spatial frequencies provided more; a coarse texture oriented horizontally provided the greatest continuity. Fig. 4a–c show examples of textured surfaces reconstructed from monocular components that provide progressively more horizontal continuity. As a special case we could completely eliminate the continuity between the two eyes' views by drawing them from statistically identical independent textures, as in Fig. 4d. We call this a *binocularly discontinuous surface*, to distinguish it from the normal *binocularly continuous surface*.

2.2. Observers and general procedure

The authors served as observers. All had normal or corrected-to-normal acuity.

Observers saw two binocular textures presented concurrently, one above the other, centered 4.2° apart. The spatial frequency and orientation for both textures was the same for a given trial. One of them (appearing randomly in the upper or lower position from trial-to-trial) was presented as a binocularly continuous surface and the other as a binocularly discontinuous surface. These remained present on the screen for as long as the observer took to decide which was more coherent and respond with a key press (usually 2–4 s). On each of the twenty trials in a block the texture giving rise to the surfaces was chosen randomly, as was the noise on the bars. From block to block a texture with different spatial frequency and orientation was drawn from the standard set of twenty one. The sequencing and display of stimuli was controlled by the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) running on a Macintosh G4 computer. Observers viewed the display with head held steady by a chinrest and forehead support, but fixation was not constrained.

3. Results

3.1. Appearance

When presented with a coarse *binocularly continuous* texture oriented horizontally, and occluded by the bars so that no part of it was seen by both eyes but all of it could potentially be pieced together from the two eyes' views (Fig. 5a), observers saw a continuous surface lying behind bars. Despite the fact that different images fell on corresponding points in the two eyes, the binocular percept was coherent and stable, though paradoxical. The bars remained visible in front of the surface, but did not obstruct the observer's view of it. As the dominant spatial frequency of the texture was raised, and/or its orientation was rotated towards vertical (Fig. 5b), the texture less often appeared like a surface in depth, losing its clearly defined depth and becoming rivalrous with the bars. When presented with a binocularly discontinuous texture partially occluded by the bars (Fig. 5c), observers never saw a coherent surface lying behind the bars. These displays were always rivalrous.

In the measurements that follow we explore quantitatively the perceived coherence and relative depth of binocularly continuous surfaces viewed through occluding bars. *Coherence* is estimated from the observer's capacity to distinguish a binocularly continuous texture from an binocularly discontinuous one with the same spatial frequency and orientation. We express this

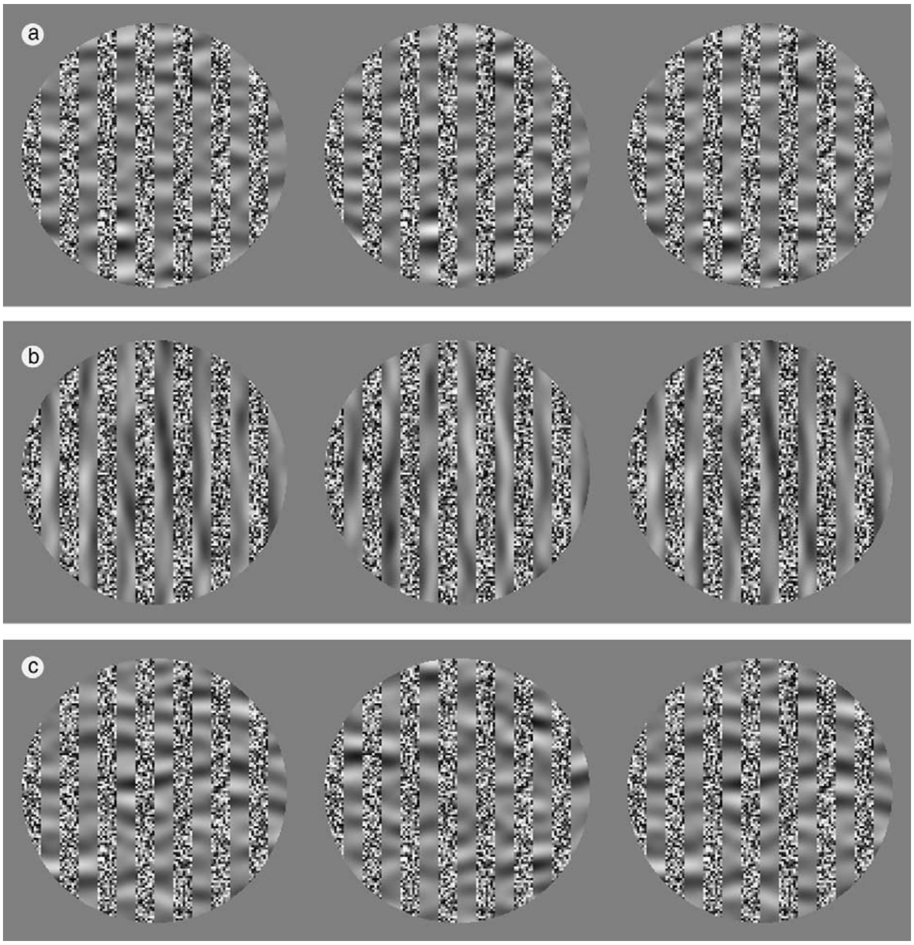

Fig. 5. Partially occluded surfaces that appear coherent (a) or rivalrous (b and c). Uncrossed fusers should view left and center images; crossed fusers should view center and right images. a. Binocularly continuous texture oriented horizontally and of low spatial frequency. b. Binocularly continuous texture oriented vertically. c. Binocularly discontinuous texture oriented horizontally. This has the same statistics as the texture in a.

quantitatively as the fraction of trials on which the observer correctly identified the binocularly continuous textures.

3.2. *Binocular continuity determines surface coherence*

We examined systematically how the coherence of a binocular surface depended on the dominant spatial frequency and orientation of textures. Fig. 6 shows, for two observers, how surface coherence depended jointly on the spatial frequency and orientation. When the binocularly continuous texture was horizontal, and of low spatial frequency, both observers saw it as coherent and could readily distinguish it from the binocularly discontinuous one. Performance was progressively impaired by rotation of the texture towards vertical, with the impairment occurring sooner for textures of high spatial frequency. Observers were never able to distinguish vertical continuous textures from discontinuous ones. This pattern of results is what we would expect if coherence depended on the extent to which the structure of the texture strip seen by one eye predicts the structure of the adjacent strip seen by the other eye.

Variations in the coherence of patterns were closely tied to variations in appearance: textures for which coherence was high appeared as stable and continuous surfaces lying in a well-defined position behind the bars; those for which coherence was low were rivalrous with the bars and appeared to lie at an ill-defined position behind the bars.

3.3. *Real-world depth relations determine coherence*

Our observations and measurements so far have dealt with displays that represented real-world occlusion of surfaces by bars lying in front of them. A disparity-selective mechanism that was coarsely tuned in relation to the width of the occluding bars might be capable of resolving depth in these displays, albeit less reliably with the bars present than absent. Such a mechanism might be capable of localizing a coherent textured surface behind the bars. We would expect it to perform best

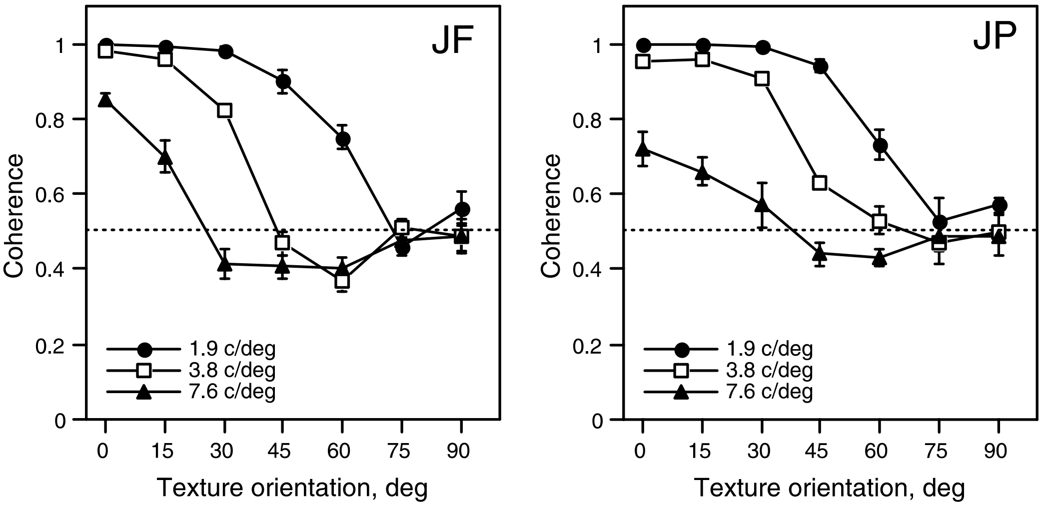

Fig. 6. The coherence of a binocularly continuous surface depends on the spatial frequency and orientation of the texture fragments of which the surface is comprised. Left and right panels show results for different observers. For both observers, coherence declined as texture orientation was rotated away from horizontal, and declined more when spatial frequency was higher. Error bars show ±1 s.e.m.

(giving the strongest coherence) on horizontal textures of low spatial frequency, as in Fig. 6, and we would also expect it to perform as well with near or far disparities relative to the bars. We tested this possibility directly by exploring the coherence of partially occluded surfaces in displays with unnatural depth relations.

We constructed a display in which the configuration was identical to that used previously, except that the disparity of the textures seen by the two eyes was reversed. In the absence of occluding bars, the fused textured surface appeared in front of the aperture. Adding the occluding bars created a display in which the disparity relationships would place the textured surface in front of the bars that occluded strips of it. This would be consistent with the real world only if the textured surface were semi-transparent. [1]

Were the appearance of the texture being determined by a coarse disparity-selective mechanism for which the occluding bars merely constituted noise, we would expect surface coherence to be as strong as it was when the surface lay behind the bars. This was not the case. Fig. 7 shows, for the same observers and the same textures as Fig. 6, but with textures now having reversed disparities, how surface coherence varied with texture orientation and spatial frequency. Observer JF could distinguish no binocularly continuous texture from a binocularly discontinuous one; observer JP could distinguish them only when the texture was horizontal and of low spatial frequency.

The appearance of the textures was consistent with their low coherence: both observers experienced rivalry, and the texture usually appeared to be vaguely located behind the bars. In the case where JP could distinguish the continuous and discontinuous textures, the continuous one appeared as a fractured surface, clearly lying in front of the bars, but punctuated by them.

The substantial effect of disparity direction on the coherence of otherwise identical textured displays makes it very unlikely that a coarse mechanism of stereopsis determines the coherence or apparent depth of the surfaces. Were coherence determined by such a mechanism we would have expected Figs. 6 and 7 to be similar.

Our results show that stable coherent surfaces can be seen only when texture strips are exposed in a way that is consistent with real world occlusion. Shimojo and Nakayama (1990) showed this to be an important determinant of rivalry at the borders of conventional stereoscopic displays (the eye seeing around the occluder dominates), and Anderson and Nakayama (1994) identified additional phenomena that suggest the visual system is well-adapted to detect natural occlusions.

3.4. Properties of the occluder influence coherence

So far we have shown that the properties of the texture strips seen by each eye profoundly affect surface coherence. We considered whether the properties of the occluding bars affect coherence. The coherence of low spatial frequency patterns was unaffected when the noise element size was changed from 2 min arc to 8 min arc (not shown), but we wondered if coherence depended on having binocularly correlated noise on the bars. Those used so far consisted of binocularly identical two-dimensional noise. Fig. 8 shows how texture orientation

---

[1] An anonymous reviewer pointed out that the photometric relationships in our stimuli are inconsistent with transparency, which would be required of any surface to be seen in front of the bars. We repeated the experiment with dark bars consistent with transparency and found no change in the result.

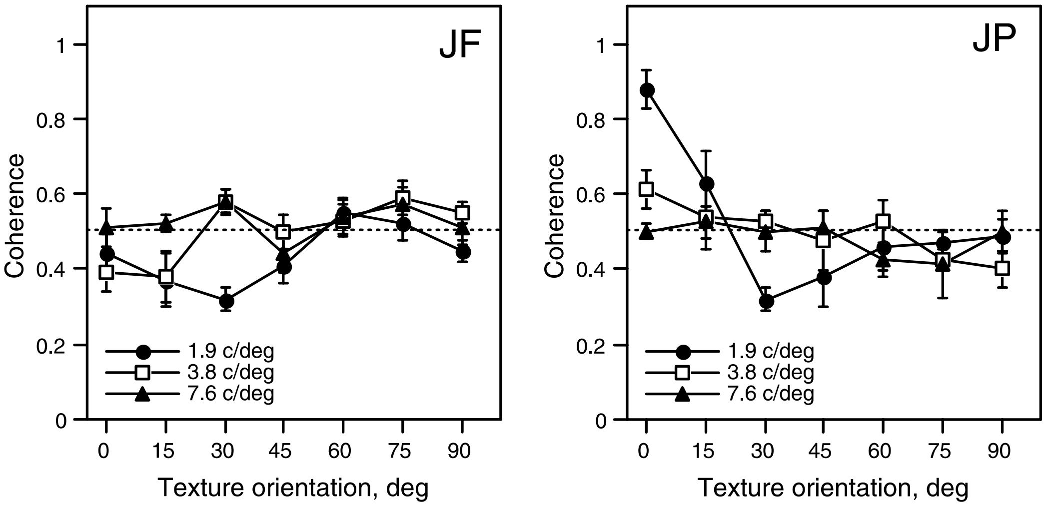

Fig. 7. Binocularly continuous surfaces are not coherent when views are inconsistent with real-world depth relations. Left and right panels show, for two observers, the effects of varying texture spatial frequency and orientation when normal depth relations were reversed, placing texture fragments in front of the occluding bars. Textures were those used in the measurements of Fig. 7. For no orientation or spatial frequency (except JP for horizontal texture of low spatial frequency) could observers distinguish continuous and discontinuous textures. Error bars show ±1 s.e.m.

and spatial frequency affected coherence when the bars in the two eyes consisted of independent noise samples. The results for correlated noise (from Fig. 6) are overlaid as dashed lines.

For both observers the decorrelation of the bars had little effect on the coherence of textures containing low spatial frequencies, but progressively impaired coherence as spatial frequency was raised. Increasing the size of noise elements from 2 min arc to 8 min arc markedly reduced coherence of low spatial frequency textures (not shown), suggesting that the spatial content of the bars does influence the perception of stable surfaces derived from monocular regions. We cannot say from our data whether the content of the bars interferes directly with the matching of monocular regions or the decorrelation destabilizes visual processes such as vergence that may be necessary for combining monocular information in the two eyes.

3.5. *Binocularly continuous surfaces are reliably positioned in depth*

The binocularly continuous surfaces seen behind bars were not only coherent, but appeared well-localized in depth. Were the monocular fragments of texture pieced together in a way consistent with real-world geometrical constraints, the surface would appear at the depth of the unoccluded texture. To establish the depth at which the

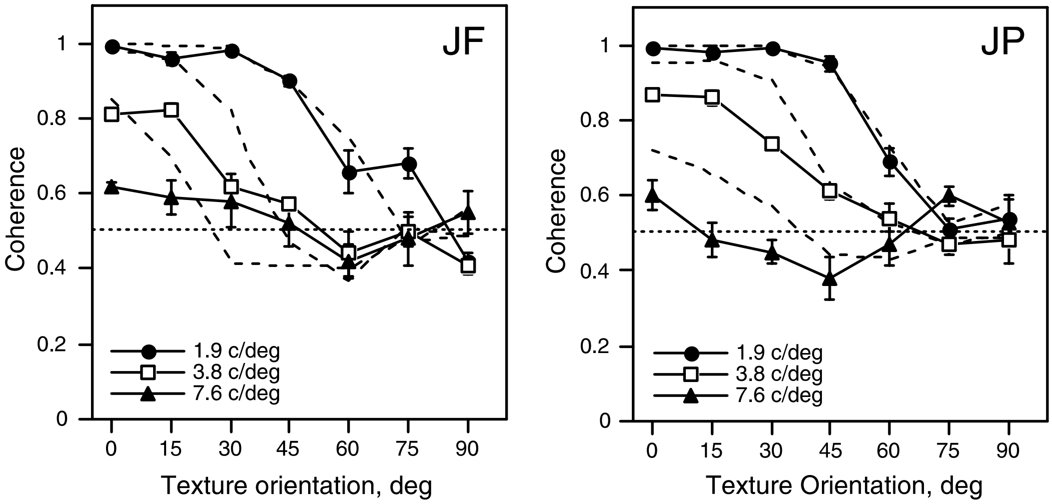

Fig. 8. Decorrelating the two eyes' views of the occluding bars impairs the coherence of textured surfaces seen behind them. Left and right panels show, for two observers, the effects of varying texture spatial frequency and orientation. Textures were those used in the measurements of Fig. 6; the results from Fig. 6 are overlaid as dashed lines. Error bars show ±1 s.e.m.

coherent surface lay we adapted the display and procedure as follows. The upper part contained a binocularly continuous texture occluded by bars, as had been used in the measurements already described. The lower part contained an identical texture, free of occluding bars, whose relative disparity could be adjusted (in steps of 1 min) by the observer moving the mouse. On each trial (unlimited in duration) the observer adjusted the disparity to place the lower surface at the same depth as the upper (occluded) one. Observers also made control settings for the case where both the surfaces were free of bars. Observers made 20 settings in each of 5 sessions. Fig. 9 shows, for JF and JP viewing textures of 1.9 and 3.8 c/deg, the average depths at which the surfaces lay.

In the control task where there were no occluding bars both observers made reliable, accurate matches at both spatial frequencies. In the presence of bars, the observers made reliable matches but in each case with a constant error. JF saw the surface too near to the bars; JP saw the surface too far away.

To characterize performance more precisely we replaced the adjustment task with a forced-choice one in which the lower surface was presented at a range of disparities around the point of apparently equal depth, and the observer indicated whether it lay in front of or behind the occluded upper surface. As before, we also made control observations with the upper surface free of bars. Fig. 10 shows, for three observers, the frequencies with which lower surfaces having different disparities were judged to lie in front of the upper surface.

In the absence of occluding bars, all observers' judgments were narrowly dispersed around the correct disparity. When the bars were present the judgments became somewhat more dispersed, and for JF and JP became biased in the directions found earlier with the method of adjustment (Fig. 9). PL's judgments were less biased. In all cases the biases were reliable, greatly exceeding the dispersion of the judgments.

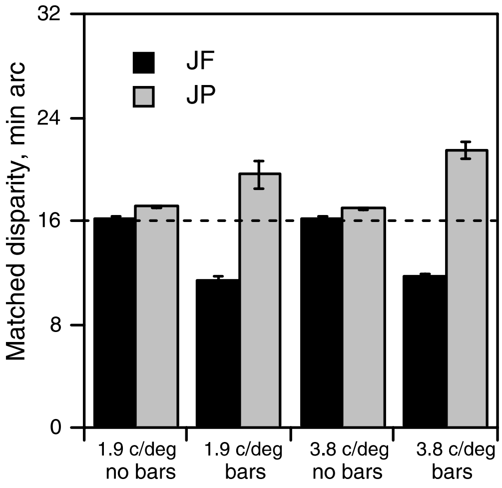

Fig. 9. Binocularly continuous surfaces are located reliably but not accurately in depth. Histograms show the depths behind the aperture at which two observers located occluded and unoccluded surfaces comprised of textures of low and medium spatial frequencies.

4. Discussion

4.1. Appearance

In our displays the two eyes never saw the same texture fragments so, no matter how observers converged their eyes, pointwise matches were impossible. A texture strip in one eye's view was always potentially rivalrous, either with the correspondingly positioned strip in the other eye's view (if the observer converged on the plane of the bars) or with a bar (if the observer converged on the plane of the textured surface). Whether or not observers experienced rivalry depended on the particulars of the monocularly-viewed texture strips. When observers viewed a binocularly discontinuous texture, the texture strips were strongly rivalrous, with characteristically variable lustrous appearance. This is not surprising. The more interesting cases concern the appearance of binocularly continuous surfaces that looked coherent. What they usually saw was paradoxical: the surface looked continuous yet the bars were clearly visible in front of it. This is inconsistent with a conventional cyclopean view in a Euclidean space, but indicates an appropriately rich representation of the surfaces that would actually be present in a natural scene. We refer to this as 'stable diplopia'. When the display was identical except that the left and right eyes' views of the texture were reversed (which would have placed the binocularly continuous surface in front of the bars) the texture strips became strongly rivalrous; the general appearance was the same as if it the textures had represented a binocularly discontinuous surface.

The absence of rivalry in mismatched image regions that are consistent with real-world occlusion has been described before. In the case studied by Shimojo and Nakayama (1990) an image fragment that had no counterpart in the other eye was stably visible only if it could represent a partially occluded surface; it was suppressed otherwise. The present work extends this observation in two ways. First, observers see stable, coherent surfaces when none of the texture fragments making up the binocularly continuous surface have counterparts in the other eye. Second, and more important, in our configuration the perceptual stability does not generally result from suppression of one eye by the other. We know that because a texture fragment seen by one eye was aligned (depending on vergence) with a different fragment or a noise bar in the other eye, yet the observer saw a continuous surface behind bars.

Were the unpaired texture fragments in our displays pieced together in correct alignment, geometrical con-

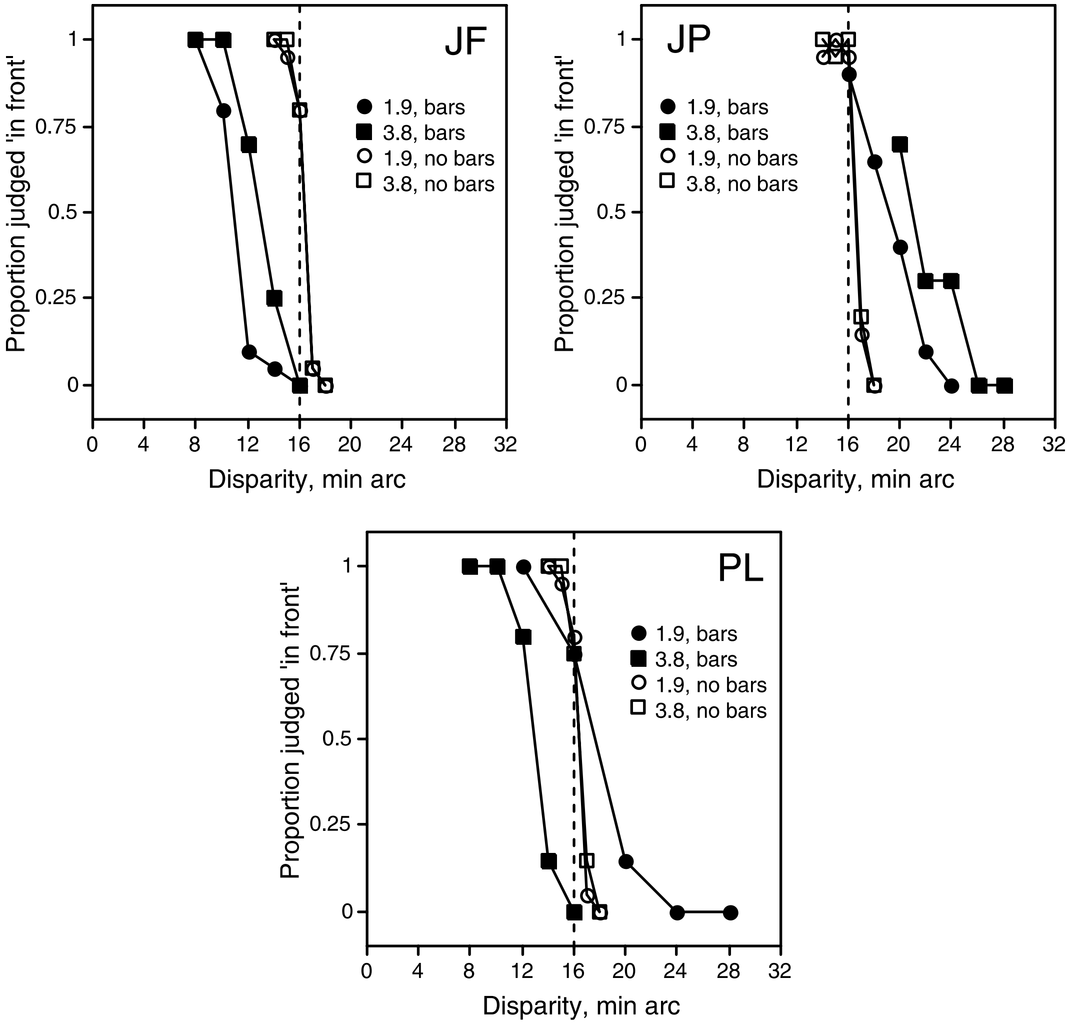

Fig. 10. Precision of judgments of the relative depth of a pair of binocularly continuous textured surfaces, one of which was partially occluded by bars, the other not. Each panel shows results from a different observer. Filled points show the frequency with which the occluded surface was judged to lie in front of the unoccluded one. The dashed vertical line marks the disparity of the unoccluded surface. Each point is based upon 20 trials in each of 5 sessions. Open points show control judgements made when neither surface was occluded. Legend numbers refer to the texture spatial frequency in cycles per degree.

straints would cause the resulting binocular texture to lie a specific distance behind the bars. Observers saw the binocularly continuous surface lying at a well-defined position behind the bars, near but not at its geometrically correct depth. This error implies that correct depth relations need not be resolved for the perception of a coherent surface. The reverse might be the case: surface coherence is first resolved and then constrains depth.

4.2. Discovering continuous surfaces

In a cyclopean view, our task at its simplest amounts to deciding whether neighboring strips of texture belong to each other. If the observer converges on the plane of the textured surface, these strips abut in the cyclopean view; if the observer does not converge precisely on that plane, the task can be conceived as one of establishing continuity for texture strips that will be horizontally misaligned. One way to discover whether the mechanism for establishing binocular continuity is special, or part of a more general class of mechanisms for detecting continuity in surface structure, is to frame the surface coherence problem for monocular viewing: when an observer views horizontally separated strips of texture that may or may not form part of a continuous surface, how do the spatial frequency and orientation of the texture influence performance as the separation of the strips is varied? We made a preliminary study of this. The observer was presented with a pair of texture strips separated horizontally by a noise strip of variable size, and had to decide whether the texture strips were parts of a continuous surface. We explored two cases: first, where the strips would abut were it not for the intervening noise; second, where the noise obscured a central strip of a continuous texture. In both cases observers coped with the largest separations when textures were oriented horizontally. Performance fell to chance at progressively smaller separations as the orientation was rotated towards vertical. These observations suggest that the machinery required to establish binocular continuity operates on the same principles as those used generally to establish that separated fragments of texture belong to the same or different surfaces. It is not clear whether the machinery is actually shared. When performance falls in the case of strips viewed monocularly, the observer accepts the texture strips as belonging to the same surface; when performance falls in the stereoscopic case the strips become rivalrous. The stereoscopic case is also special in another respect: when observers must discover a textured surface that would lie in front of occluding bars performance is profoundly worse than when they must discover one that would lie behind the occluding bars.

If establishing continuity in stereoscopic and monocular conditions depends on common machinery, then the different outcomes in cases where the observer fails to establish continuity imply either that an earlier stage can resolve depth relations to veto certain candidate matches, or some later stage can resolve the three-dimensional relationships into those that are geometrically possible and impossible, with the possible ones leading to stable diplopia and the impossible ones leading to rivalry. We take up this issue in Section 4.3.

Establishing surface continuity is a form of perceptual grouping, so it is worth asking how the mechanisms uncovered in our experiments might be related to those revealed through other perceptual tasks whose solution requires integration of separated image fragments. Field, Hayes, and Hess (1993) explored quantitatively the perception of continuity, and found that an observer's capacity to discern a path defined by separated, oriented, image fragments in a field of similar randomly oriented elements depended on the local alignment of adjacent path elements, and rather less on their separation and local difference in orientation. Although it is hard to compare results from the different experiments, Field et al. found that observers could integrate effectively over distances at least as large as the separations used in the present work. Furthermore, collinear pattern elements are more likely to be bound together in rivalrous displays (Alais & Blake, 1999).

4.3. Real-world depth constraints and rivalry

The sensitivity of appearance and performance to real-world depth constraints raises the question of how these are brought to bear. To the extent that the same mechanism can be used to establish surface continuity in stereoscopic or monocular viewing (previous section), one need not resolve depth relationships to establish surface continuity. Nevertheless, depth constraints might act early, to veto the matching of texture fragments in configurations that represent unnatural occlusions.

Anderson and Nakayama (1994) suggested the visual system might rely on relatively low-level mechanisms to distinguish natural and unnatural occlusions. They postulated a mechanism with a receptive field that had different properties in its left and right halves. In one half the mechanism required matching (correlated) inputs from the two eyes, and in the other half it required input from one eye only. Of four possible configurations of left and right halves only two would be sensitive to simple natural occlusion boundaries: left side driven monocularly by the left eye and right side by both eyes; right side driven monocularly by right eye and left side by both eyes. The three principal display configurations we used (a binocularly continuous surface behind the bars, a binocularly continuous surface in front of the bars, and a binocularly discontinuous surface) had identical occlusion geometry, so detectors of the kind postulated by Anderson and Nakayama would be expected to give identical signals. However, only the texture fragments representing a binocularly continuous surface behind bars formed a coherent surface; the others were rivalrous.

The occlusion geometry in our displays was more complex than that in Anderson and Nakayama's displays: ours contained valid unpaired texture fragments seen by the left eye to the *right* of occluding bars, and valid unpaired fragments seen by the right eye to the *left* of occluding bars. These would be invalid in the configurations considered by Anderson and Nakayama, so it is perhaps not surprising that a detector that handles the simpler cases well fails on ours. We have tried to elaborate the low-level occlusion detector into one that might cope with the relationships in our displays, but have been unable to devise one that could do the job. We have also been concerned that physiological observations on binocular neurons in cortical areas V1 and V2 provide little indication that receptive fields could distinguish natural and unnatural occlusions. To a first approximation most neurons in monkey V1 (Smith, Chino, Ni, & Cheng, 1997) and V2 (Kraft, Peirce, Forte, Krauskopf, & Lennie, in preparation) combine signals from the two eyes linearly, and prefer very similar visual stimuli. Destructive interference of signals from the two eyes is rare.

The difficulty of constructing simple detectors that could resolve the depth relationships in our displays, coupled with the physiological evidence on binocular receptive fields, leads us to think it unlikely that depth relationships in displays of the kind we used could be resolved by low-level mechanisms. We think our results point instead to an organization in which the structure of individual texture fragments seen by the two eyes determines whether or not they can be paired. If the fragments can be pieced together, the surface is resolved and the depth relations are determined. If these depth relations are geometrically possible in the real world, surfaces are seen in depth; if the depth relations are impossible, and cannot represent real-world occlusion, binocular rivalry results.

One omission from this account is how rivalry comes about. We know that in the anesthetized monkey almost all neurons in V1 and V2 have binocular receptive fields, and that signals from the two eyes seldom interfere destructively. To the extent that monocular signals are combined in V1 and V2 the visual system has no direct access to the signal from either eye, so rivalry must reflect some differential regulation (controlled by feedback from higher areas) of the gain of the monocular inputs to the early binocular neurons. This regulation might happen in V1 (Polonsky, Blake, Braun, & Heeger, 2000), or even in LGN, which has rich descending connections from cortex.

5. Conclusion

We have shown that monocularly visible texture fragments are sufficient for the perception of a stable surface with quantitative depth, if the textures fragments are binocularly continuous and consistent with the surface lying behind occluders. Such surfaces escape rivalry. Texture fragments that are not continuous (but still consistent with real-world occlusion) are susceptible to rivalry, provide only a qualitative sense of being behind occluders and do not give rise to the perception of a surface.

Our results show that simple occlusion rules based on identifying correlation/decorrelation boundaries cannot account for surface appearance. Furthermore, physiological results suggest that neurons at the early stages of visual processing cannot distinguish natural from unnatural occlusions. These findings lead us to believe that the perception of coherent surfaces from monocularly visible texture fragments depends on higher-level mechanisms. However, the present study tells us little about the nature of these mechanisms or why observers are unable to see transparent surfaces in front of occluders when the same information is readily used to construct stable surfaces behind occluders.

Acknowledgements

This work was supported by NIH grants EY 04440 and EY13079.

References

Alais, D., & Blake, R. (1999). Grouping visual features during binocular rivalry. *Vision Research, 39*, 4341–4353.

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

Anderson, B. L., & Nakayama, K. (1994). Toward a general theory of stereopsis: binocular matching, occluding contours, and fusion. *Psychological Review, 101*, 414–445.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system. *Vision Research, 33*, 173–193.

Gillam, B., & Borsting, E. (1988). The role of monocular regions in stereoscopic displays. *Perception, 17*, 603–608.

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

Hakkinen, J., & Nyman, G. (2001). Phantom surface captures stereopsis. *Vision Research, 41*, 187–199.

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis.* New York: Oxford.

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

Kraft, J., Peirce, J., Forte, J., Krauskopf, J., & Lennie, P. (in preparation). Binocular combination of signals in V1 and V2 of macaque.

Liu, L., Stevenson, S. B., & Schor, C. M. (1994). Quantitative stereoscopic depth without binocular correspondence. *Nature, 367*, 66–69.

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

Pelli, D. G. (1997). The VideoToolbox software for psychophysics: transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Polonsky, A., Blake, R., Braun, J., & Heeger, D. J. (2000). Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nature Neuroscience, 3*, 1153–1159.

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

Smith, E. L. III, Chino, Y., Ni, J., & Cheng, H. (1997). Binocular combination of contrast signals by striate cortical neurons in the monkey.