# Toward a General Theory of Stereopsis: Binocular Matching, Occluding Contours, and Fusion

Barton L. Anderson and Ken Nakayama

Models of stereopsis have focused on developing strategies for identifying common features in the 2 half-images so that disparity may be computed. This emphasis ignores the unpairable features that arise at occluding contours (*half-occlusions*). Most models treat half-occlusions as noise or outliers that are interpreted after disparity processing is completed. A series of experiments reveal that occlusion relationships are sensed during the earliest stages of binocular processing. The authors hypothesize the existence of receptive field structures that sense the local structure of stereoscopic occlusion relationships to account for these findings. Finally, a simple theoretical framework is presented in which fusion, stereopsis, and occlusion are unified. This theory explains the co-occurrence of stereopsis and diplopia and how half-occlusions escape the suppression characteristic of binocular rivalry.

Over a century of research has revealed the richness of stereoscopic vision in generating percepts of three-dimensional (3D) form. Since Wheatstone's (1838) invention of the stereoscope, it has been known that the different views projected to our two eyes contain information that is used to recover depth. Historically, theories of stereopsis have relied on computations of disparity to explain percepts of stereoscopic depth (Dev, 1975; Jones & Malik, 1992a, 1992b; Julesz, 1971; Marr & Poggio, 1976, 1979; Pollard, Mayhew, & Frisby, 1985; Prazdny, 1985). Models of stereopsis differ in their methods of computing binocular disparity, but despite these differences, disparity has been—and remains—the fundamental type of information used in all models of stereoscopic processing.

The foremost theoretical difficulty that arises when attempting to understand how the visual system computes disparity is known as the *correspondence* or *matching* problem. By definition, disparity, the relative difference in (retinal) position of an image feature, can only be defined once corresponding features have been identified in the two eyes. The goal of binocular matching is to link features that arise from the projection of a common surface region. Once identified, the difference in the relative retinal positions of matched features may be computed, generating a local measure of disparity. At first consideration, matching does not seem to pose any great theoretical difficulty. Most natural scenes are replete with distinctive forms that could, in principle, be identified in each eye before binocular matching. However, Julesz's (1960) invention of the computer-generated random-dot stereogram (RDS) revealed that monocular form recognition was not a necessary prerequisite to binocular matching. Rather, matching could be based on rather

simple image tokens, which, in the case of the RDS, seemed to be individual points. Once points were matched, the computation of disparity was straightforward. By computing disparity for all points in the half-images, a disparity "map" could be generated that putatively contained all of the information used by the visual system in constructing representations of stereoscopic depth.

Recently, attention has been drawn to the fact that the matching problem cannot be solved for all features projected to the two eyes. Even if all of the "correct" matches are identified, there remain regions that cannot be assigned an appropriate match and, hence, do not generate disparity signals. These unpaired regions arise from occluding contours, a fact appreciated since at least the time of DaVinci. The geometric conditions that give rise to such unpairable features are depicted in Figure 1. Following Belhumeur and Mumford (1992), we refer to these features as *half-occlusions*, because these features are occluded in one eye but visible to the other. Despite the impossibility of computing disparity for half-occlusions, psychophysical evidence is accumulating that demonstrates that the visual system makes positive use of these regions in generating 3D representations of surface structure. Unpaired image features can generate subjective contours (Lawson & Gulick, 1967; Nakayama & Shimojo, 1990), and they do not undergo the fluctuations of dominance and suppression characteristic of binocular rivalry (Julesz, 1960; Shimojo & Nakayama, 1990). There is even some evidence that unpaired image features can reduce the time needed to perceive depth in RDSs, at least when these unpaired points are consistent with geometric constraints of occlusion (Gillam & Borsting, 1988). These findings present challenges for models of stereopsis that rely solely on measures of disparity, as such models treat unpairable features as noise.

It is currently unclear how these findings should be incorporated in the disparity-based framework so prevalent in models of stereopsis. One approach is to handle the breakdown in correspondence that occurs at occluding contours during the final stages of refining the disparity map (e.g., Jones & Malik, 1992a, 1992b). This strategy does not necessitate any fundamental

Barton L. Anderson and Ken Nakayama, Department of Psychology, Harvard University.

Correspondence concerning this article should be addressed to Barton L. Anderson, who is now at the Laboratory of Vision Research, Rutgers University, Psychology Building, Busch Campus, Piscataway, New Jersey 08854. Electronic mail may be sent to drbart@gandalf.rutgers.edu.
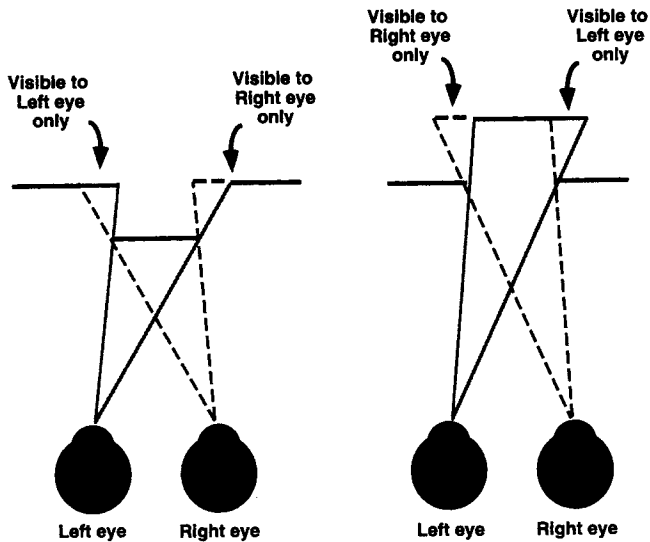
*Figure 1.* Top view showing how occluding contours generate features that are visible to one eye only, features we refer to as half-occlusions. Despite the fact that half-occlusions do not generate disparity signals, they appear stable and in depth at the farthest surface in the cyclopean three-dimensional representation.

change in the central role played by disparity but, rather, simply requires that disparity computations be supplemented by another stage of processing. The determination that a contour is an occluding contour is typically derived from the depth relations in the disparity map. In virtually all models of stereopsis, the process of establishing binocular correspondence is assumed to be the primary "problem" that needs to be solved. Surface relationships such as occlusion are assumed to be products of this matching process, not properties that could be used to facilitate the reconstruction of stereoscopic depth. Despite the ubiquity of this theoretical approach, there is actually no direct evidence to sustain this belief; it simply acts as an assumption that has shaped the development of stereo models. Indeed, it is conceivable that the breakdown in correspondence that occurs at occluding contours could be detected very early in stereoscopic processing if there existed mechanisms that were organized to sense this structure. Whether or not such mechanisms exist can only be resolved by experiments that are designed to determine when information about occlusion impacts on stereoscopic processing.

## Overview of the Problems

The purpose of this article is to analyze the role of occlusion configurations in binocular vision. In particular, we consider the impact of the structure generated by stereoscopic occlusions on theories of matching and fusion. We argue that a careful consideration of the experimental results on occlusion can resolve a number of recalcitrant debates about the nature of binocular vision. This article is divided into four sections. We begin with a review of the different primitives that have been proposed in models of binocular matching and evaluate these models with psychophysical data to decide which primitives are the most

plausible model of the input to human stereopsis. This review motivates a set of experiments designed to determine if and when information about occlusion impacts on stereopsis. The experiments and a theoretical treatment of the results comprise the second section of the article. The third section summarizes the theoretical consequences of our findings and considers the relationship of these ideas to previous theoretical treatments of stereopsis, particularly cooperative models. The fourth section relates the concept of matching to binocular fusion. In this section, we present a theory of binocular fusion that is general enough to account for the coexistence of stereopsis, diplopia, and half-occlusions.

The fundamental question that we focus on in this article is whether occlusion configurations can influence the matching phase of stereopsis, the earliest process involved in recovering stereoscopic depth. If the answer to this question is "yes," the conceptual problem that emerges is understanding how a failure of binocular correspondence could constrain the process of establishing matches for regions that are visible to the two eyes. It seems natural to think that the failures in correspondence that occur at occluding contours can only be so identified after determining the features that can be matched. After all, how can an "unpaired" feature be so defined without knowing what features are matchable? However, this thinking may reveal more about our theoretical prejudices than the visual system. Indeed, one could invert this question and ask how a match can be defined without knowing what constitutes a nonmatch? These questions may represent two sides of a single problem. Whether or not the visual system establishes matches before sensing occlusion relationships is a question that can only be resolved by experiment. There is currently little or no published research that can guide us in answering this question. The experiments we performed were designed to fill this experimental gap, that is, to determine whether information about occlusion is used by the visual system as "early" as disparity processing, or if occlusion configurations contribute to stereopsis only after some initial processing of disparity. In anticipation, the results we present provide evidence that the breakdown in matching that occurs at occluding contours can actually constrain the correspondence process. This requires that we reconsider the focus that has been traditionally placed on the problem of matching and find a more general representational scheme that incorporates the positive information provided by occluding contours and half-occluded features.

## Section 1: Models of Stereoscopic Depth

Until the recent recognition of the problems generated by occlusions, the problem of stereopsis was to establish correspondence and compute disparity. The goal of any model of disparity detection is to formulate a representational medium in which the units coding for disparity respond only for the correct matches at the appropriate disparity. A number of properties are needed to construct a model of disparity detection. It is not sufficient to posit "cells" that respond to features that are disparate. Rather, there must be a disparity-specific response of the cells underlying disparity detection that allows for the apprehension of different disparities as specific depths. The conceptually simplest way to achieve a disparity-specific representa-

tion is to assume that there exist separate cell populations tuned to the discriminable disparities. This assumption is either implicitly or explicitly adopted by virtually all models of stereopsis. Even for the three "pools" hypothesis of Richards (1970, 1971, 1972), there must be some interactions that allow for the disparity-specific sensitivity observed psychophysically to be preserved in the physiological substrate responsible for disparity detection. Some recent physiological (LeVay & Voight, 1989; Poggio, Gonzalez, & Krause, 1988) and psychophysical (Cormack, Stevenson, & Schor, 1993) evidence favors the multiple "channel" model of disparity sensitivity, not just three pools. Because this is the conceptually easiest way to understand disparity sensitivity, we assume this type of representational medium here. No strong commitment to this particular method of representation is necessary for the majority of ideas that we develop in this article. However, we argue that this representational medium provides a natural way of understanding the coexistence of stereopsis and diplopia, a phenomenon that is rarely if ever addressed in theories of stereopsis. We return to this topic in the last section of this article.

One convenient means of depicting this representational scheme is a Keplerian projection field (see Figure 2). In this framework, different disparities are represented along the ordinate (labeled *depth*). The perceived location of a given feature is represented by the intersections of the projected "rays" from the two eyes of the features that are believed to underlie matching. This representation provides an effective means for capturing the types of problems that arise when assuming different forms of matching primitives and is an important heuristic tool for the discussion that follows.

A number of questions are fundamental to any model of stereopsis. These include an articulation of what features are matched (i.e., the matching "primitives"), how the matching

process is achieved, and the consequences of not finding appropriate matches. The answers to these questions are not independent. What is matched constrains how matching is thought to proceed and the possible problems that may be encountered in establishing matches. The nature of the matching primitives also gives meaning to the concept of a "mismatch" or noncorrespondence, which is central to understanding what is meant by a half-occlusion or an incorrect match. Before we can conduct experiments on the role of occlusion in the matching process, we must understand the salient differences between the kinds of matching primitives that have been proposed and the impact this has on the kinds of processes necessary to achieve matching within each model type. We may then compare the various alternatives to psychophysical data, which will motivate the design of our experiments that evaluate the role of occlusion in binocular matching. We begin with an overview of the different matching primitives that have been proposed.

## Matching Primitives

*Pointwise matching (grey-scale correlation).* The earliest (computational) models of stereopsis simply used points (or pixels) as matching primitives (Dev, 1975; Julesz, 1971; Marr & Poggio, 1976; Sperling, 1970). These points varied along a single dimension (intensity), and in some models, a given point could only assume a value of 0 or 1 (black or white). In part, these models were motivated by Julesz's invention of the computer-generated RDS. It seemed natural to think of displays composed of dots being matched on a dot-by-dot basis, greatly simplifying the notion of correspondence: Two dots could be matched if they had the same intensity. Points are "featureless" matching primitives in the sense that they have no internal structure. Therefore, the only information needed to characterize a monocular point is location and intensity. Because of their inherent simplicity, the only binocular measure that can be computed for individually matched points is the difference in their relative retinal positions, or retinal disparities. A disparity map is generated by performing this computation for all of the "correctly" matched points in the two eyes.

While pointwise matching simplified disparity computations once correspondence was achieved, finding the correct or best match became extremely difficult. The simplicity of the matching primitives meant that there would be many points in the other eye of the same type that did not correspond to the appropriate or "true" match for that feature. This problem has been called the *matching-noise*, or *false target*, problem (Julesz, 1971; Nelson, 1975). Indeed, the primary difference between most models of stereopsis is how false targets are eliminated to determine the "best" matches. This problem is represented in the Keplerian diagram depicted in Figure 2. Note that all possible intersections are all locally equally good matches. Clearly, as the number of points in an image becomes large, the number of false targets becomes enormous. The severity of the false target problem has been the primary theoretical constraint that has shaped virtually all extant models of stereopsis.

*Edge-based matching.* The severity of the false target problem encountered by using points as matching primitives inspired new strategies for establishing binocular correspondence. The motivation behind these approaches was articulated by
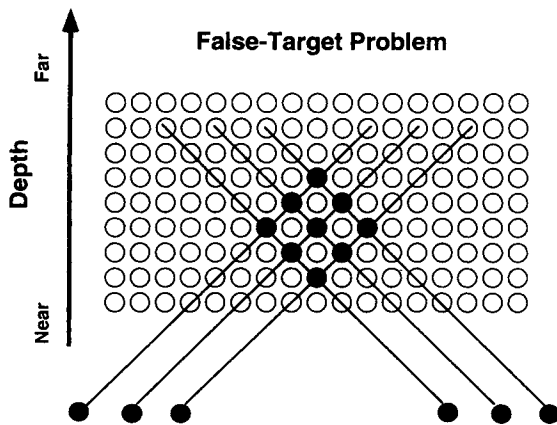


*Figure 2.* The false target or matching-noise problem as it arises for three discrete points in the two eyes. The array of circles define a Keplerian "grid." Each open circle represents a hypothetical disparity detector that would respond when matchable features from the two eyes project into the same circle (or receptive field). Note that when each black dot is considered separately, any intersection with any of the three dots in the complementary eye may be a possible match. This ambiguity as to the "correct" match is known as the false target or matching-noise problem.

Marr (1982), who recognized that matching noise could be eliminated in one of two ways: either by reducing the density of features that were compared; or by using relatively complex matching primitives, so that there would rarely be a compatible false target in the complementary eye. Most models have pursued the former alternative. For example, Marr and Poggio (1979) developed a stereo algorithm that used the zero-crossings of spatially filtered images as matching primitives. Zero-crossings refer to the image locations where the second derivatives of the luminance profiles cross zero. Such signals provide a rough estimate of the locations of luminance edges in a scene, which often correspond to the locations of object boundaries (although this correlation is far from perfect). The benefit of using zero-crossings for matching is that they are relatively sparse features, reducing the number of comparisons needed to establish binocular correspondence. However, zero-crossings are also more complex features than points. The zero-crossings of an image form (piece-wise) continuous contours, which can be characterized by a number of different measures such as orientation, curvature, spatial frequency, and contrast polarity. This means that the computation of disparity is more complex than when using points as matching primitives. Marr and Poggio limited their matching criterion to two properties: the orientation and contrast polarity (direction of contrast) of the zero-crossings. In this model, zero-crossings could be matched if they were similar enough in orientation and had the same contrast polarity (say, if they were both dark–light). Note that zero-crossings only specify the positions at which the luminance varies most rapidly in the two eyes; all information about the magnitudes of contrast is lost in this representation. Edge-based models must also rely on some method of depth interpolation to assign depth to the regions between zero-crossings.

*Patches and linear filter models.* A strategy that is currently popular in the artificial intelligence community is to correlate some patch or window of points (or pixels) with features in the other eye (Geiger, Ladendorf, & Yuille, 1993; Gruen, 1985; Kanade & Okutomi, 1990). This reduces the number of false targets by eliminating the number of compatible matches in the other eye. However, a number of new difficulties emerge. One is determining the optimally sized window for the correlation. If too small, there is little difference between this strategy and matching points; both result in a large number of false targets. If too large, then difficulties arise if one or more discontinuities are included in the patch. In this case, there may be no patch in the other eye that will be sufficiently correlated to declare a successful match. Thus, one unresolved issue is to provide a principled account for choosing the best sized window on which to perform the cross-correlations of the two images.

Another difficulty with the patch approach is their use of a least squares method of correlation. This computation implicitly assumes that the major difference between corresponding patches is the presence of noise, ignoring the possibility that corresponding regions can vary in more systematic ways. For example, different viewpoints can cause differential foreshortening of surfaces that are not strictly fronto-parallel. The appreciation of such systematic variations in the projections of image patches played a large role in the linear filter approach described by Jones and Malik (1992a, 1992b). In this model, local image patches are represented by the outputs of linear spatial filters that sample the image at a range of scales and orientations. The matching primitives in this model are vectors that contain the entire set of filter responses for a given image region. A statistical measure of similarity (in their model, the sum of absolute differences) is used to determine the best candidate match in the complementary eye. Here, the "best match" corresponds to the image patch with the smallest vector difference. This input representation has a number of advantages. For one, the complexity of the matching primitives makes it unlikely that there will be many candidate matches for any given surface patch, reducing the amount of matching noise. Second, this input representation allows for an accurate recovery of properties such as surface slant and tilt, which generate orientation disparity and spatial frequency differences in corresponding image patches. Most models only measure positional disparities, implicitly assuming that other sources of variation of corresponding image features may be treated as noise. This allows such properties as differential foreshortening to provide useful information rather than making the matching process more difficult. However, as with the patch approach, this model will also run into difficulties if the filter output contains a discontinuity or occluding contour. Hence, some strategy must be developed to eliminate the large filter contributions near discontinuities so that such edges will not be inappropriately smoothed.

## Processing Strategies for Eliminating Matching Noise: Theory and Data

All computational models of stereopsis use some measure of similarity to determine correspondence. In the earliest models (Dev, 1975; Julesz, 1971; Marr & Poggio, 1976), features could be matched if they were deemed "compatible." This criterion was usually invoked to explain matching in a one-bit (black and white) RDS. In this context, compatibility simply referred to whether dots (or pixels) were the same color, that is, black or white. More generally, the complexity of the measure of similarity used by any particular model will increase as the complexity of the matching primitives increases. "Noncorresponding" features are generated when none of the candidate matches meet the model's criterion for declaring a successful match.

Matching errors can arise in at least three ways. First, some features that have matches may be left unmatched by virtue of missing its correct interocular pair. Second, a feature could be matched to the wrong feature in the other eye (matching noise). Third, some regions that do not have correct matches (i.e., half-occluded regions) could be incorrectly matched with a feature in the other half-image. If a feature has a correct match in the other eye, an attempt should be made to find it. Falsely interpreting a nonmatch as a half-occlusion will generate inappropriate breaks in the representation of surfaces. However, if the feature is part of a half-occluded region, it should be left unmatched. If a half-occluded feature is mistakenly assigned a match, it could inappropriately influence the depth of its neighbors and mask the clarity of the edge. These two types of errors—interpreting too many features as half-occlusions or assigning inappropriate matches to half-occlusions—must somehow be balanced to avoid the pitfalls inherent in either extreme. We argue that the need to rapidly identify half-occlusions and avoid the false target problem favors the conclusion that the vi-

sual system uses relatively complex matching primitives. However, before we can make this argument, we need to review some of the strategies proposed for eliminating false targets when using the various matching primitives we described above. We consider the problem of occlusion in the sections that follow.

In pointwise matching models, the false target problem was at its most severe. The source of the problem was the simplicity of the matching primitives: A possible match was any pixel (dot) of the same intensity in the other eye. Because there were many pixels of the same intensity in the two eyes, any candidate match considered individually would be as good as any other. The problem was to somehow determine the best match among all the candidate matches. Clearly, some auxiliary information must be provided to solve this problem. This was accomplished by imposing limiting assumptions about the kinds of outputs that were considered as correct matching solutions. A variety of assumptions has been proposed. Perhaps the most ubiquitous constraint is a *smoothness* constraint. We focus on the role of smoothness constraints because such assumptions bear directly on the problem of identifying occluding contours.

A number of strategies for imposing smoothness were developed. Marr and Poggio (1976) proposed an algorithm in which pointwise disparity-sensitive cells locally excited neighboring cells tuned to the same disparity. In this model, the excitation was strongest between laterally connected same-disparity cells, which biased this algorithm to detect surfaces that were strictly fronto-parallel, not just smooth. More recent models have considered this form of smoothness to be too restrictive and therefore broadened the range of disparities that excite neighboring units (Pollard et al., 1985; Yuille, Geiger, & Bulthoff, 1991). Despite differences in the kind of smoothness invoked, the computational role of this constraint was singular: to limit the kinds of surfaces that would be considered correct matching solutions and thereby eliminate matching noise.

A major limitation of pointwise matching models was that they could only solve RDSs; that is, they could not be applied to natural images. The natural extension of these approaches was to allow for a greater variety of pixel intensities, now referred to as *grey-scale correlation*. However, grey-scale matching has been criticized as being computationally unreliable (Poggio & Poggio, 1984), as a number of variables will cause the same surface patch to project different intensities on the two eyes. For example, differences in brightness arise from viewing surfaces illuminated by a point light source (such as the sun) from two slightly different positions, a fact that is inescapable with binocular viewing. More significantly, a number of psychophysical results have indicated that the human visual system can match image features that differ greatly in luminance. The simplest evidence for this was presented by Julesz (1971), who demonstrated that random-dot half-images of different luminances could yield vivid stereoscopic depth. The only requirement was that half-images have the same direction of contrast, or contrast polarity. The discovery that contours with opposite directions of contrast resist fusion dates back to Helmholtz (1910/1925), and some version of this constraint remains in virtually all models of stereo matching.

The fact that matching apparently tolerates differences in the intensities of image features but not contrast polarity played a formative role in the matching criteria used in edge-based ste-

reo models. Edge-based algorithms reduced matching noise in two ways. First, compared with points, edges were relatively sparse features, so there were fewer candidate matches that needed to be considered. Second, edges (or zero-crossings) could be computed at a number of different spatial scales. The coarse scales would have very few features, whereas the finer scales would have many. False targets could be eliminated by matching the largest scales first and using the matches found at the coarse scales to constrain the search for matches on the finer scales. The insight guiding this theory was the belief that there would be a strong correlation between the scale of the spatial structure and the magnitude of the disparity: Coarse spatial scales could have large disparities, but fine spatial scales would be restricted to relatively small disparities. A coarse-to-fine strategy could be implemented by either driving vergence movements (Marr & Poggio, 1979) or by neural circuits that cortically shift the search for matches on the fine scales (Nishihara, 1987; Quam, 1987). Recent experiments have apparently eliminated the first alternative by demonstrating that band-limited high spatial frequency targets can drive vergence movements (Mowforth, Mayhew, & Frisby, 1981). However, some support for the use of coarse-to-fine matching strategies has been found to constrain fusion limits (Wilson, Blake, & Halpern, 1991) and may in fact play a role in facilitating fusion.

Although edge-based models have enjoyed some computational success, recent experiments have revealed that the human visual system uses more than the contrast polarity of zero-crossings to establish binocular correspondence. Perhaps the strongest evidence for this was provided by Bulthoff and Mallot (1988), who found that stereoscopic depth could be perceived in shaded patterns devoid of zero-crossings. This result implies that interocular variations in local contrast can provide useful information for recovering stereoscopic depth. Thus, contrast differences are not simply tolerated in stereo matching, but can provide information used to reconstruct depth. This result should be compared with Julesz's (1971) observation that random-dot half-images of unequal contrast can be fused. Julesz's finding demonstrated that binocular matching could occur in the presence of global contrast differences and ruled out matching schemes that required strict grey-scale matches. However, recent psychophysical evidence has revealed that such global differences in half-image contrasts deleteriously affects stereoscopic acuity. Halpern and Blake (1988) and Legge and Gu (1989) independently found that disparity thresholds increased when the half-images had identical contrast polarities but differed in contrast magnitude. Reducing the contrast in just one eye caused a much greater elevation in disparity thresholds than reducing the contrast of both eyes. The fact that stereo performance is worse when the contrast is reduced in just one eye demonstrates that the degradation in binocular sensitivity is not due to simply a problem in monocular localization. This would predict that performance should be worse when the contrast is reduced in both eyes. Hence, these results imply that interocular contrast differences impair some binocular process. One likely candidate is the process of binocular matching.

To summarize, whereas the findings of Bulthoff and Mallot (1988) illustrate the positive role of local contrast variations for recovering stereoscopic depth, the Halpern and Blake (1988) and Legge and Gu (1989) studies reveal the deleterious effects

of global interocular contrast differences. Together, these studies provide evidence that the primitives used to establish binocular correspondence contain information about image contrast, not just contrast polarity. This provides a partial answer to the question of what is matched in stereopsis and implies that noncorrespondence refers to interocular differences in contrast. This is not to say that contrast differences prevent stereopsis; rather, as we argue below, our contention is that contrast differences can provide information that could be used to identify half-occluded features.

## Is the False Target Problem a False Problem?

The motivation behind the edge-based approaches was to reduce the number of features that needed to be compared to establish correspondence, thereby reducing the amount of matching noise. However, perhaps the entire concept of matching noise is misleading. The idea that stereopsis must eliminate false targets is predicated on the belief that there will be many identical possible matches for a given feature, that is, a large number of features that are equally good matches, only one of which is correct. This certainly would be the case if the primitives used in binocular matching were individual points. Then, by definition, there would exist many possible matches that were all equally good. This is also true if a dense array of zero-crossings served as matching primitives. Selecting the single best match then requires some additional information, usually in the form of assumptions about the nature of the surfaces that are likely to be encountered (e.g., smoothness) or that the disparity of a feature will be proportional to its spatial scale (as in coarse-to-fine strategies). However, we believe that the fundamental mistake that has been made in the statement of the matching-noise problem is the confusion of a method of stimulus construction with a theory of visual perception. Creating a stimulus by randomly distributing discrete points on a computer screen does not mean the visual system treats this stimulus as an array of discrete points. Indeed, the statement of the false target problem harkens back to the early structuralist accounts of visual perception that treated the visual stimulus as a retinal mosaic of discrete points that must be assembled into coherent wholes. Although current accounts of spatial vision have seemingly advanced beyond this description of the retinal stimulus, this conceptualization remains as the form of input used in many models of stereopsis even to this date (see, e.g., Belhumeur & Mumford, 1992).

Pointwise matching emphasized the role of binocular neural interactions (excitation and inhibition) as a means of disambiguating correct matches from false targets. These models were called "cooperative," because they emphasized the dependence of a given disparity-sensitive cell on the activity of its neighborhood. Cooperativity was the "force" invoked to explain how a coherent 3D percept could be assembled from an array of discrete points. For example, in the Marr and Poggio (1976) cooperative algorithm, all of the possible (pointwise) matches were initially identified (including the false targets), and the neighborhoods of all candidate matches were allowed to interact. In this model, disparities along the same lines of sight would inhibit each other, and excitation would propagate between nearby cells tuned to the same disparity. Thus, the local ambiguity in matching could be resolved by searching the neighborhood of a match for "support." Note, however, that the neighborhoods of features were only considered during binocular matching; the input to the matching process was discrete points, not image neighborhoods.

When considering the problem of matching natural scenes, entirely different problems arise. In the projection of natural scenes, it is very unlikely that even the correct match will appear identical in the two eyes. This is true whether individual points or zero-crossings serve as matching primitives. The examples we have given include illumination differences and differential foreshortening generated when viewing slanted or curved surfaces from slightly different vantage points (which occurs by virtue of one's two eyes being laterally displaced in one's head; see Poggio & Poggio, 1984). In other words, features that should be matched, that is, features that correspond to a common distal source, are typically not identical in natural scenes. Therefore, a more complex method of comparing features must be used that can tolerate some differences but, at the same time, be sensitive to gradations in the similarity of the features to determine the best match. Note that the problem here is not one of having many features in one eye that are identical to a single feature in the other but, rather, that even the correct match will not be identical to its binocular mate. If the matching primitives are sufficiently complex, correspondence can be determined by comparing a given feature to all candidate matches and selecting the most similar feature as the appropriate match. The complexity of the primitives will almost always guarantee that this will result in unique matches, at least for regions of the scene containing a high degree of luminance variation. This is the approach taken by Jones and Malik (1992a, 1992b).

In the filter model described by Jones and Malik (1992a, 1992b), matching noise is reduced by using the outputs of linear spatial filters that sample the image at a range of spatial frequencies and orientations as the input for binocular matching. Each spatial location of the image is characterized by a vector that contains information about the orientation and spatial frequency content of the image patch. Because this characterization of the matching primitives is relatively complex, the number of false matches that arise becomes negligible. Ironically, RDSs generate very little matching noise for this model, as RDSs contain information at all spatial frequencies and orientations at virtually all regions in the images. This fact provides a powerful argument that the amount of matching noise is not necessarily large in RDSs. In contrast to cooperative models that use the neighborhoods of the disparity map to reduce matching noise, the filter approach advocated by Jones and Malik exploits the neighborhoods of the monocular input to stereopsis to reduce the matching ambiguity.

The crucial step needed to eliminate the false target problem is the use of rich monocular neighborhoods as the input to the matching process. This could be done in a variety of ways, including the technique of adaptive windows that has become popular within the artificial intelligence community (Geiger et al., 1993; Gruen, 1985; Kanade & Okutomi, 1990). Thus, the only reason that we are highlighting the approach of Jones and Malik (1992a, 1992b) is that this model is more closely aligned with current knowledge about the monocular processing stages believed to occur before binocular combination in the human

visual system. From a computational perspective, there may be little difference between correlation techniques that use adaptive windows and similarity measures defined on the vector outputs of spatial filters. However, it is somewhat ironic that many stereo models pay so little attention to the types of primitives that seem to be present in area V1 of the cortex, the first site of binocular combination. Cells in V1 are sensitive to a wide variety of image properties, including orientation, contrast polarity, contrast magnitude, and spatial frequency. In representations of this kind, "points" or individual pixel intensities do not even exist. In short, the richness of matching primitives seems to be an inescapable biological fact that must be incorporated within computational models of stereopsis, at least if these models are to qualify as models of human stereopsis. In addition, when such richness is introduced, the false target problem takes on an entirely new character.

In models using simple, unstructured matching primitives, the false target problem arose because of the existence of too many candidate matches. The designation as a feature as the "correct match" could only emerge after extensive cooperative processing to eliminate all of the false targets. However, if the input to stereopsis contains rich information about the local monocular image structure, then the problem of disparity detection and the identification of half-occluded regions is greatly simplified. Note that the intersections of the Keplerian projection field are only ambiguous under the assumption that the internal structure of the false matches are identical to the correct matches (see Figure 3). If the rays that are projected in the Keplerian model contain sufficient internal structure, then there will typically be only one intersection of a given surface



**False-Target Problem Resolved**

⊞ = Local matching Failures
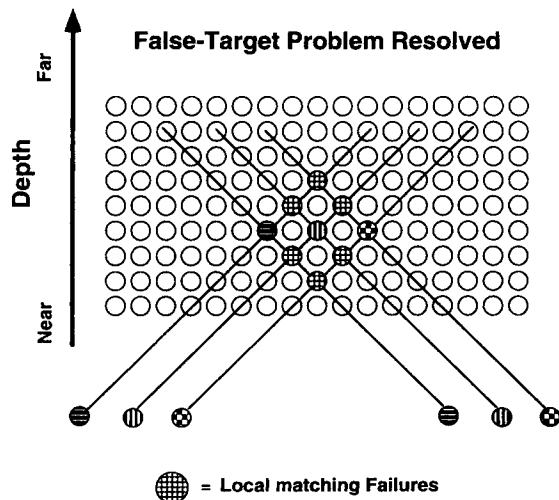
*Figure 3.* A cartoon depicting how the false target problem vanishes if the local structure that is presumed to input into a given disparity-sensitive receptive field is richly structured. Note that the false target problem requires that the local intersections are all equally good matches, which implies that the inputs to binocular matching contain no internal structure. This characterization of the input seems unlikely when viewed in the context of what is known about the spatial processing of the monocular input to stereopsis and the richness of the properties represented in area V1, the first site of binocular combination.

patch with a patch in the other eye that will be highly similar, corresponding to the correct match. The false target problem is essentially eliminated. Indeed, the primary source of matching ambiguity arises when there are little or no luminance variations to specify a given surface region. We discuss this point in greater detail in a later section of this article.

## Noncorrespondence and Occluding Contours

Not surprisingly, all types of models encounter their greatest difficulties at occluding contours. The difficulty in processing occlusion configurations is inherent in models that focus solely on disparity detection, because occlusion configurations contain half-occlusions that (by definition) do not have correct matches. The vast majority of stereo models have simply ignored these regions (Dev, 1975; Julesz, 1971; Marr & Poggio, 1976, 1979; Pollard et al., 1985) or have attempted to extract them in a second pass of an iterative matching algorithm (Jones & Malik, 1992a, 1992b). Models that ignore half-occlusions mask the clarity of occluding contours, either by smoothing over them or incorrectly matching the unpaired features to spurious features in the other eye. These models do not capture the clarity with which these edges appear to human observers.

The use of complex matching primitives helps to rapidly identify both the correct matches and matching failures by virtue of eliminating matching noise. However, it also generates some difficulties that were not present in pointwise matching algorithms. When using either image patches or a range of filter sizes as the input to binocular correlation, some of the filters used to characterize a given image patch may contain a discontinuity, whereas other (smaller) filters representing the "same" image patch do not. It is then possible that no good match will be found for these regions if all of the filters are used. Some process is therefore needed to remove those filters that contain a discontinuity from the matching phase. How can this be accomplished? If the emphasis on disparity detection is retained, this problem can only be resolved after some initial disparity processing has occurred. For example, the approach pursued by Jones and Malik (1992a, 1992b) was to perform post hoc processing of the disparity map to determine the likely locations of discontinuities. They began by computing the local depth variation over a specified region of the disparity map. The size of the region was determined by the size of the largest spatial filter used for stereo matching. If the depth variation in this region exceeded a specified threshold, it was concluded that the filter was too large to be useful for matching in this region (i.e., the filter may contain a discontinuity). This procedure was applied iteratively over the image and over a range of spatial scales (in a coarse-to-fine progression) until the algorithm converged, that is, the filter sizes ceased exceeding their respective thresholds.

Once this algorithm converged and the inappropriately large filters were eliminated, the half-occluded regions could be identified. In this model, half-occlusions were identified through the construction of what they term a *visibility map*. The visibility map was 1 at each image position that is visible in the other eye and 0 if it was a half-occlusion (i.e., a monocular region). Again, it is important to remember that the visibility map was constructed from an initial estimate of the disparity map; half-oc-

clusions were derived from computations of correspondence and then removed from the processing stages that refined the disparity estimates. The half-occlusions would be inserted in the disparity map only after it had been "refined." Indeed, the presence of half-occlusions represents a computational nuisance rather than a positive source of information in this and virtually all other stereo models. Thus, the determination of occlusion relationships did not occur until all decisions about binocular matching had been completed.

Although Jones and Malik's (1992a, 1992b) model retains an emphasis on disparity computations, this emphasis does not seem mandated by their approach. We suggest that one potentially important role of rich matching primitives is the ability to achieve a rapid determination of both matchable and half-occluded features. Note that a measure of similarity is simultaneously a measure of dissimilarity, a fact that can facilitate the identification of both matches and half-occlusions. If there exist mechanisms sensitive to the breakdown in correspondence at occluding contours that are "fed" by richly structured matching primitives, then the identification of occlusion configurations could be performed very early in binocular processing. However, until it can be shown that occlusion configurations can impact on early stereoscopic processes, there is no empirical reason to believe that half-occlusions are not treated in the post hoc manner already contained in the Jones and Malik model. The question that remains to be answered is whether occlusion configurations are used by the human visual system during binocular matching or whether any positive information contributed by these regions is incorporated only after some initial estimate of correspondence has been determined. Currently, there is little information that could be applied to answering this question. This lack of clarity arises from an absence of data on the role played by occlusion configurations in the initial phases of binocular matching. The crucial result that must be demonstrated is that an occlusion configuration, that is, a region containing an occluding contour and half-occluded features, can constrain the correspondence process, which is by definition the earliest phase of stereoscopic processing. Actually, even this is not sufficient. It must also be demonstrated that such information affects correspondence during the initial phases of binocular matching. This would eliminate the possibility that occlusion constraints affect correspondence through a feedback loop. But what kind of evidence would indicate that an occlusion configuration constrains the correspondence process? How can we test whether "second-pass" strategies provide an adequate account of the role of half-occlusions in human stereopsis? In the section that follows, we suggest that the depth relationships at occluding contours may be used to evaluate when these regions impact on the representation of stereoscopic depth.

## Depth Relationships at Occluding Contours

One of the earliest attempts to account for the perceived depth of features that do not have unique matches was described by Panum (1858/1940) and is now known as Panum's limiting case. Panum constructed a stimulus in which one line is presented to one eye and two lines are presented to the other. Note that unique matches cannot be formed for all of the lines

in this stereopair. Nonetheless, observers typically experience relative depth in these stereograms. This perception of depth has been attributed to the single target matching both targets in the two-target half-image, generating two disparity signals. The difference in the magnitudes of the two disparity signals is putatively responsible for the percept of relative depth.

Panum's (1858/1940) experiment suggests that all image features impart a sense of depth because the visual system actually pairs them, even if these features are also matched with another target. This explanation is appealing because it implies that the perceived depth of both paired and unpaired features arises from disparity computations. Indeed, the double-matching thesis suggests that the phrase "unpaired features" is a misnomer: all features are perceptually paired, even if this requires double matches. If this were true, it would provide a unifying framework for understanding how all binocular image features are processed and would not necessitate any fundamental changes in the role of disparity in models of stereopsis.

However, the results of a number of experiments using variations of Panum's (1858/1940) limiting-case stimuli have been mixed. Some studies have shown mixed results across observers, conditions, or both (Gettys & Harker, 1967; Howard & Ohmi, 1992; Nakayama & Shimojo, 1990; Ono & Wade, 1985; Westheimer, 1986), and in some experiments, depth was not seen (Brewster, 1844). Recently, Ono, Shimono, and Shibuta (1992) have argued that these mixed results have occurred because previous research failed to appreciate Panum's limiting case as an occlusion configuration. In the Ono et al. study, the depth of an unpaired bar was found to follow Panum's rule only when it appeared on the temporal side of a fused bar, which is the configuration that leads to the unpaired stimulus appearing as the most distant target. In other words, depth was seen most clearly when the unpaired line appeared either to the left of a fused bar in the left eye or to the right of a fused bar in the right eye. These conditions are precisely those for which the matched bar could be interpreted as occluding the unpaired bar. Reversing this pattern (e.g., placing an unpaired bar on the left of a paired target in the right eye) did not reverse the perceived depth of the targets, as would be expected from the double-matching hypothesis. In the experiments that follow, we demonstrate that Panum's double-matching rule, or any matching rule for that matter, fails to predict the perceived depth of the ambiguous wallpaper patterns we presented. We argue that our results demonstrate the contribution of some other type of information that is used by the human visual system in establishing depth relationships in stereopsis, namely, occlusion relationships.

One of the basic consequences of an occluding contour that generates half-occluded features is that the half-occlusions belong to the surface behind the occluding contour. This simple fact is crucial for understanding the results of our experiments. In anticipation, the experiments described below reveal a bias for an ambiguous pattern to appear at the farthest of two possible depths, even when the "strength" of the matches is not greatest for that depth. Indeed, we demonstrate that schemes that focus solely on the degree of correspondence between the two possible depth organizations not only fail to predict this bias but they also predict the opposite form of depth bias. We contend that the findings described below can be understood by considering the "strength" of the two possible occlusion inter-

pretations: The contour that retains the strongest correlation–decorrelation boundary will be biased to be interpreted as the occluding contour, even if this depth interpretation does not maximize the total amount of matching "energy." This depth bias appears as soon as stereoscopic depth is resolvable, implying that information about occlusion impacts on the earliest stages of binocular processing.

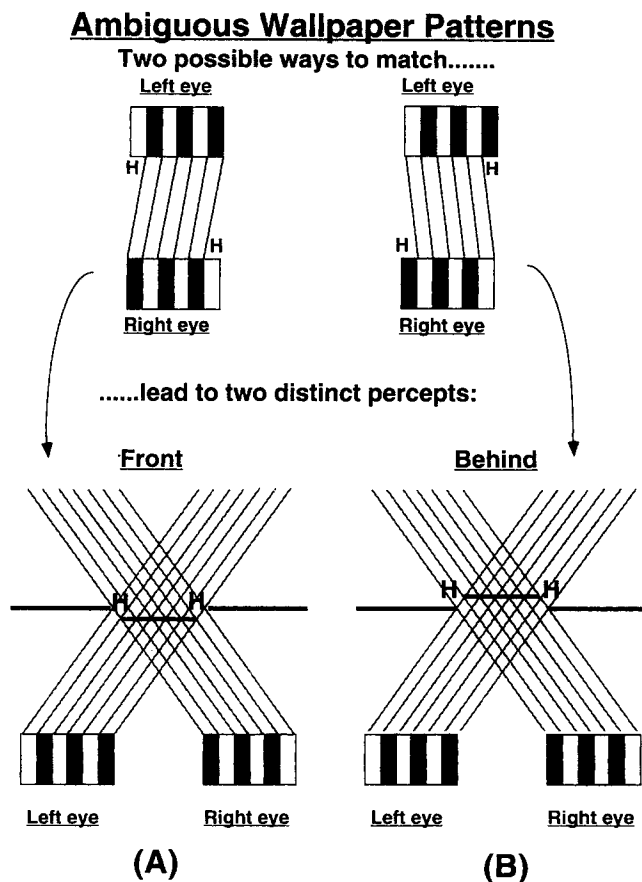### Section 2: Experimentally Testing Whether Half-Occlusions Influence Binocular Matching

*Using Contrast to Bias Binocular Matching*

We used ambiguous patterns—patterns that could be matched in two mutually exclusive ways—to study the role of occlusion constraints on binocular matching. More specifically, we examined the role of contrast in specifying a feature as matchable or unmatchable by measuring its ability to bias the perceived depth of an ambiguous wallpaper pattern to one of its two possible depths. There are two percepts that typically arise when viewing unbiased wallpaper patterns: The pattern can be seen either in front or behind a zero-disparity background (Julesz & Chang, 1976). Only one depth is perceived at any given instant; a simultaneous percept of both depths is never experienced. The exclusive nature of the matching process in these patterns allows us to use the perceived depth of the wallpaper pattern as an indicator of the constraints used in binocular matching.

There were two goals to the experiments presented here: (a) to gain an understanding of the matching primitives used to determine whether features were matchable or unmatchable and (b) to determine whether occlusion relationships influence binocular matching. Wallpaper patterns provide the opportunity to examine both of these questions. To see how, we consider each question in turn.

First, consider the problem of deciding which stripes of the wallpaper patterns should be matched. In an unbiased wallpaper pattern, either set of pairings can be formed. (Our definition of bias is any information that will result in the wallpaper pattern appearing predominantly at one of the two possible depths.) We have already described existing psychophysical data demonstrating that magnitudes of image contrast are used during binocular matching. We therefore should be able to bias the depth of the wallpaper pattern by introducing contrast differences for the matches in the two depth planes. If contrast is used during binocular matching, then the depth that retains the highest degree of contrast similarity should be the organization that forms.

A second problem is determining the occlusion relationships in the images. This problem is illustrated in Figure 4. In Figure 4, the central wallpaper figures are presented in a Keplerian diagram, depicting the two possible percepts that can arise. Note that the occlusion relationships are different for the two possible organizations. When the wallpaper pattern appears in front (Figure 4a), the right-most stripe in the left half-image and the left-most stripe in the right half-image form the occluding contours, whereas the other two stripes that border the wallpaper pattern form the half-occlusions. However, when the wallpaper pattern appears behind the background, the background



**Ambiguous Wallpaper Patterns**
**Two possible ways to match.......**

......lead to two distinct percepts:

**Front**          **Behind**

(A)                (B)

*Figure 4.* The possible matching schemes for an ambiguous wallpaper pattern. We used these patterns to explore the constraints underlying the matching process. Two possible depths may be perceived with wallpaper patterns. The lines connecting the edges of the two eyes' views represent possible matching schemes, and the Hs represent the regions that will be interpreted as half-occlusions. The two depths may be distinguished by either the set of matches selected or by which regions are interpreted as the half-occluded regions. Virtually all extant theories treat occlusion relationships as properties derived from the matching process, not as properties capable of influencing binocular matching. Our results challenge this view.

becomes the occluding contour. If the depth of the wallpaper pattern is simply determined by the strength of the best matches, the occlusion relationships would simply emerge as a result of matching the most similar (or largest) contrasts.

As we have previously discussed, the determination that features are half-occluded or that a contour is an occluder is usually considered to be a rather high-level property of visual processing. In virtually all extant models, surface properties such as occlusion would not be expected to influence the more primitive process of binocular correspondence. But again, this may be a theoretical bias that is not borne out by experiment. In addition to maximizing the correlation between features that can be matched, the correlation–decorrelation boundary created by occluding contours could be sensed in parallel with disparity processing and could be used to help drive correspondence. Indeed, there is already one small study that suggests this

(Gillam & Borsting, 1988). This runs counter to the theoretical treatments of stereopsis that have been developed to date. Nonetheless, there are good reasons to believe that the locations of occluding contours are extremely important features that should be sensed as early as possible. Perhaps the most basic problem confronted in any visual task is to segment the images into natural objects. If occlusion relationships could be rapidly identified, at least crudely, it would greatly facilitate the process of image segmentation (or "grouping"), which in turn may help solve the correspondence problem.

In the sections that follow, we provide evidence that information about occlusion can strongly influence the matching process. More specifically, we document biases in a bistable pattern that cannot be understood by the strength of the matches at that depth. We argue that these results may be understood by a process that selects the strongest correlation–decorrelation boundary as the occluding contour. Remarkably, this information can overwhelm matching schemes that rely on the polarity and location of zero-crossings. Indeed, we will show that if we apply Panum's (1858/1940) double-matching thesis to our patterns, then in one of our patterns the visual system must be matching contours with opposite contrast polarities, even when the possibility exists for matching contours with the same contrast polarity. All of these effects can be experienced directly by the reader by fusing the stereopairs we present. We proceed by presenting demonstrations of the main effects, followed by experimental documentation of these results.

## Demonstration 1: Contrast Polarity as a Matching Constraint

Demonstration 1 illustrates how contrast can bias an otherwise bistable pattern to appear at a specific depth. The wallpaper pattern consisted of alternating dark and light stripes (i.e., a luminance square wave) embedded in a zero-disparity, random-dot background. The zero-disparity background defined a reference disparity that could be compared with the depth of the wallpaper pattern. Binocular disparity was introduced by shifting the wallpaper pattern in one of the two eyes by one stripe width, or one half cycle. Thus, if the left eye contained a square wave of dark–light–dark–light bars, the right eye would be light–dark–light–dark. As mentioned earlier, two percepts are possible with this pattern, illustrated in Figure 1. This bistability is evident when viewing repetitive patterns having no contrast difference between elements, such as that presented in Figure 5a, a finding originally demonstrated by Julesz and Chang (1976). With a little practice, observers can usually flip the wallpaper pattern between the two organizations (i.e., in back or in front of the background).

However, a bias can be observed when the wallpaper pattern consists of luminance stripes. By simply altering the contrast and polarity of the light and dark bars relative to the background, a specific percept will emerge as dominant. This can be experienced by fusing the stereograms presented in Figures 5b and 5c. Note that one of the stereopairs is predominantly perceived to lie in front of the ground plane (Figure 5c for crossed fusion), whereas the other is seen predominantly behind. The only difference in these two patterns is that the half-images have been interchanged. As we explain below, the contrast relation-
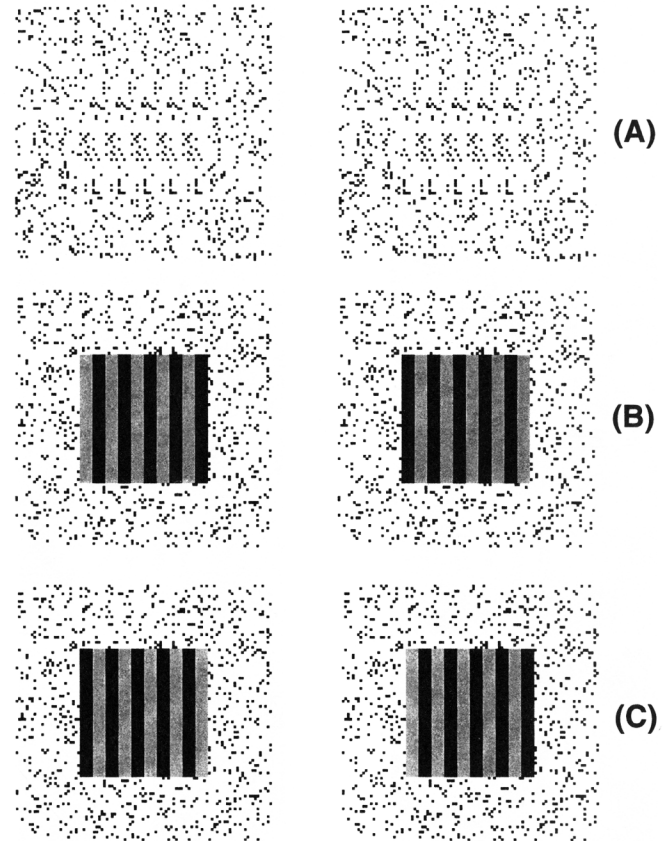


Figure 5. Three different wallpaper patterns. Pattern (A): The wallpaper pattern was created by alternating two columns of random-dot patterns. No particular depth bias is evident, and observers can usually learn to alternate between the front and behind percepts with equal ease. Pattern (B): A wallpaper pattern composed of alternating light and dark stripes, both of which had a lower mean luminance than the background. A definite depth bias may be observed in this pattern (behind, for crossed fusion). Pattern (C): Same as (B), but the half-images have been interchanged. The predominant depth of this pattern is opposite that in (B).

ships that create this bias occur at the wallpaper figure–background borders. To demonstrate this, we present a sequence of stereograms in which the luminance of the background is varied and the luminance values of the light and dark stripes are held constant (see Figure 6). Observers uniformly report that at some intermediate background luminance, the wallpaper pattern becomes bistable, whereas for very dark and light backgrounds, the predominant organization switches from in front of to behind the ground plane. What causes this bias?

To illustrate how contrast acts as a form of depth bias, we have constructed a diagram of the possible matching schemes of the stereograms of Figure 5a in Figure 7. In this figure, a cross section of the corresponding luminance distributions for the left and right eyes is put into register, and small lines depict the possible matches of the edges of the patterns. First, consider the two possible pairings of luminance discontinuities within the wallpaper pattern. Note that the light grey stripes and the black stripes within the pattern share a common edge. These
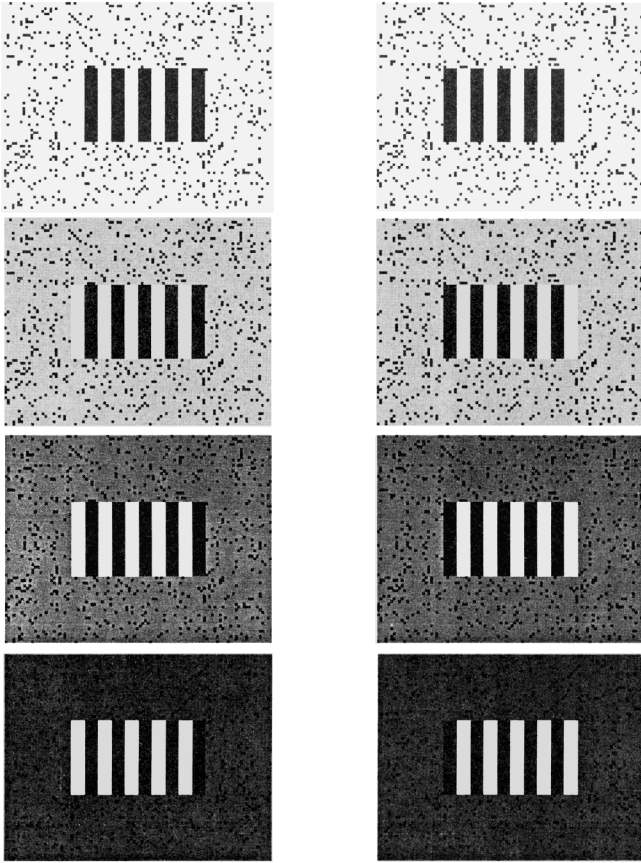
*Figure 6.* A demonstration that the depth biases observed in the patterns in Figure 5 arise from the contrast relationships of the wallpaper pattern relative to the background. The luminance of the background is varied while the luminance relationships within the wallpaper pattern are held constant. A definite shift in bias is observed: The top figure is biased to appear behind (for crossed fusion), whereas the bottom figure is predominantly in front. The middle two figures are relatively bistable.

edges do not bias either depth interpretation of the wallpaper pattern because both depth interpretations are supported by local same-polarity edges (as depicted by the black circles in Figure 7). However, a strong bias is evident when we consider the contrast polarity between the wallpaper pattern and the background. For simplicity, we restrict our description in the text to the left side of the wallpaper–background borders, as a similar analysis holds for the right borders. Consider the matching scheme that results in the left-most black stripe in the left half-image of the wallpaper pattern (i.e., the second stripe from the left) being matched with the left-most black stripe in the right half-image (depicted in Figure 7b). If image features are referenced from left to right, we can see that the direction of contrast is the same for the two half-images (light–dark), and the magnitudes of contrast are also roughly the same (both high contrast). In contradistinction, consider the contrast relations occurring at the borders of the grey stripes (Figure 7a). In the left half-image, the grey stripe has a small contrast relative to the ground plane, and the direction of contrast is light–dark. However, the border of the grey stripe in the right half-image has a large mag-

nitude of contrast and has the opposite direction of contrast (dark–light), depicted by the small open circles. Our visual systems are biased to match contours with similar contrast values, contours with the same contrast polarity, or both.

This analysis can explain the fact that when the two half-images are interchanged in Figures 5b and 5c, the predominant organization of the wallpaper pattern switches from behind to in front of the ground plane. Depth reversals occur when interchanging the half-images of most stereograms because such an interchange reverses the disparities of the two half-images. In our wallpaper pattern, this reversal occurs because the visual system is applying a single rule to both stereopairs: match those contours that have the same contrast polarity, similar magnitudes of contrast, or both. By itself, this finding cannot determine which of these two contrast properties (contrast polarity or magnitude of contrast) was more important in disambiguating the wallpaper pattern, or even if there is a difference in their importance. However, this finding does demonstrate that some property of contrast strongly constrains binocular matching and may be viewed as a predictive achievement of some extant models of stereopsis.

## Demonstration 2: Violation of Contrast Polarity as a Matching Constraint

The previous demonstration establishes direction and/or magnitude of contrast as a potent constraint in binocular matching. It is difficult to know which property was more salient, because these properties were correlated in Demonstration 1. We therefore constructed a new stereogram that uncoupled these two aspects of contrast. The goal was to create a stereogram in which the magnitudes of image contrast were substantially different at the wallpaper–background borders, but the contrast polarity of the edges was the same as Demonstration 1. If contrast polarity was the primary constraint that dictated the matching biases observed in Demonstration 1, then varying the contrast magnitudes at the wallpaper–background borders should have little impact on the pattern of results: The matching schemes preserving contrast polarity should be the organizations that form, causing one of the wallpaper patterns to appear in front of the background and the other behind the background.

In Figure 8, we present a variation of the stereopairs presented in Figures 5b and 5c, but we have reduced the contrast between the dark and light stripes. The qualitative (i.e., ordinal) contrast relations with the background were identical to Figure 5, and, thus, the contrast polarity relations were also identical. However, unlike the patterns in Figure 5b and 5c, observers uniformly report that the wallpaper patterns of both stereopairs in Figure 8 predominantly appear behind the ground plane. As we argue below, no theory that simply maximizes the similarity of binocular matches can account for this result.

To understand the difficulty posed by this result for theories of stereopsis, we present a schematic of the possible matching schemes for these stereopairs in Figure 9. As before, we restrict our description in the text to the left side of the wallpaper patterns to simplify our verbal presentation. The stereogram presented in Figures 9c and 9d depicts the matching alternatives for the stereopair in the top half of Figure 8 (assuming crossed

## Contrast Polarity as a Constraint in
## Binocular Correspondence
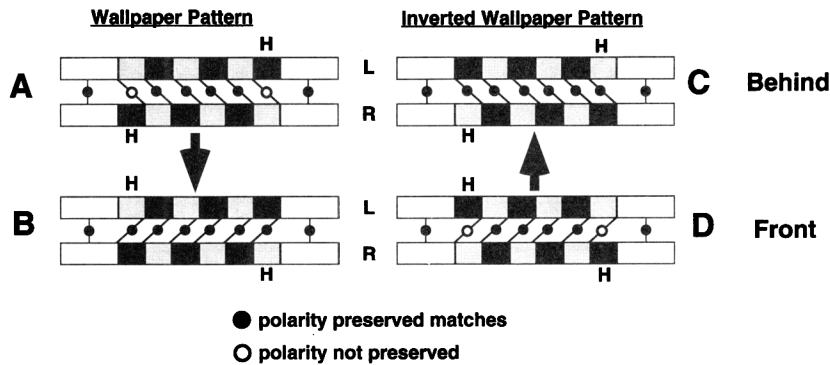


*Figure 7.* A matching explanation of the depth biases observed with the stereopairs depicted in Figure 5. A cross section of the corresponding luminance distributions for the left and right eyes are put into register, and small lines depict the possible matches of the edges of the patterns. The small black circles connect edges with the same polarity, whereas the open circles connect edges that differ in polarity. The large bold arrows indicate the percept that results with each pattern. Note that there is no bias for either matching scheme within the wallpaper pattern. However, a difference is observed between the possible matches at the borders of the wallpaper and the background. The depth biases observed seem to result from a strategy whereby the visual system matches contours that preserve contrast polarity. This is a predictive achievement of many models of stereopsis. H = half-occlusion; L = left eye; R = right eye.

fusion). It is no surprise that this stereopair appears behind the zero-disparity ground, because this organization is consistent with matching contours with the same contrast polarity. However, this is not true for the other stereopair (Figures 9a and 9b).
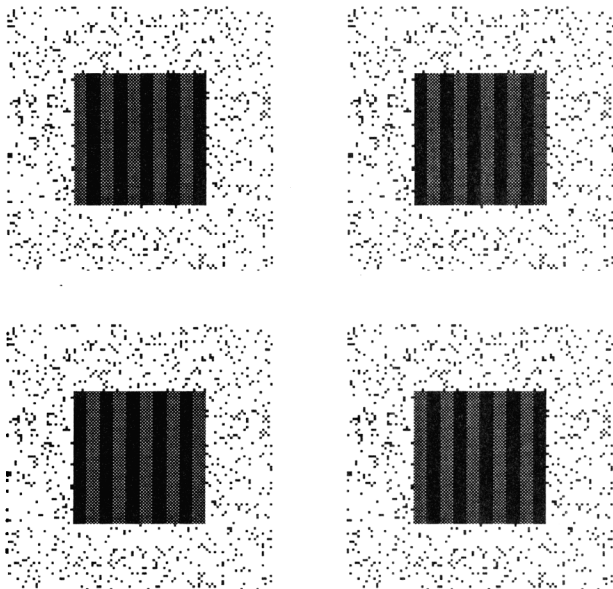


*Figure 8.* Patterns similar to those presented in Figure 5 except that the intensity of the light bars was reduced. The contrast polarity relationships were identical to those in Figures 5b and 5c. Observers report that the predominant depth of both patterns is behind the background. This result challenges all extant theories of binocular matching.

To see why, consider the matching scheme that results in the left-most black stripe in the left half-image (i.e., the second stripe from the left) being matched with the left-most black stripe in the right half-image (Figure 9b). If image features are considered from left to right, we can see that while the magnitudes of contrast are quite different, the direction of contrast is the same for the two half-images (light–dark). Now consider the contrast relations occurring at the borders of the lighter stripes. In the left half-image, the lighter stripe has a high contrast relative to the ground plane, and the direction of contrast is light–dark. However, the border of the grey stripe in the right half-image has a small magnitude of contrast and has the opposite direction of contrast (dark–light). Nonetheless, the visual system selects this matching scheme, resulting in the pattern appearing behind the ground plane. The common feature of both matching schemes is that they result in the wallpaper pattern being organized behind the ground plane. This finding is documented more thoroughly in Experiment 1.

### Experiment 1: Role of Contrast in Determining Adherence and Violation of the Contrast Polarity Constraint

In Demonstration 1, we presented bistable patterns and found that binocular matching and perceived depth were influenced by the contrast relationships at the edges separating the wallpaper pattern from the background. In Demonstration 2, we found that by simply reducing the contrast of the stripes in the wallpaper pattern, both patterns appeared behind the background. In Experiment 1, we made more systematic measurements of the effects of contrast on the depth biases we observed in Demonstrations 1 and 2. Specifically, the contrast separating

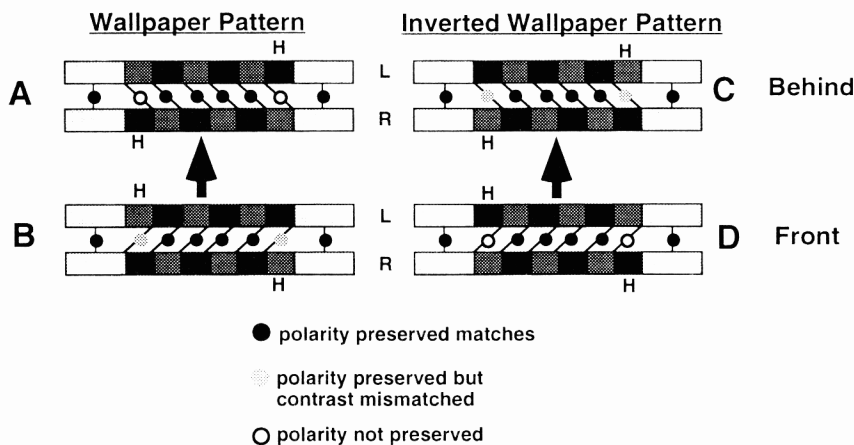## Violation of Contrast Polarity as a
## Matching Constraint



*Figure 9.* Schematic revealing the failure of contrast polarity in predicting the perceived depth of the patterns in Figure 8. Black circles connect edges that preserve contrast polarity, grey circles connect same-polarity edges but edges that differ in their magnitude of contrast, and open circles connect opposite-polarity edges. The visual system has a bias to perceive both patterns behind the background. This occurs even for the patterns depicted in A and B, despite the fact that this seems to require matching edges with opposite contrast polarities, even when the possibility exists to match contours of the same contrast polarity. All models of matching fail to predict this result. See text for details. H = half-occlusion; L = left eye; R = right eye.

the stripes within the wallpaper pattern (which we refer to as the *between-stripe* contrast) was varied along a continuum, and observers reported the perceived depth of the wallpaper pattern relative to the background.

*Method.* A random sequence of stereopairs was presented to 3 observers, 2 of which were naive (S. S. and Z. H.), and the third was one of the authors (B. A.). The 3 observers had normal or corrected-to-normal vision. The contrast of the stripes in the wallpaper pattern was varied from trial to trial. The density of the random-dot background was 10%, and the mean luminance was approximately 14 cd/m², both of which were constant for all trials. The luminance of both stripes (light and dark) was higher than the mean background luminance, ensuring that both stripes had the same contrast polarity relative to the background.[1] The displays were viewed from a distance of 1.2 meters, and the entire pattern subtended a visual angle of 3°. The wallpaper targets formed a square and subtended a visual angle of 1.5°. The disparity of the targets was equal to 0.5 of a period of the square wave, which was approximately 9 arc min. The actual contrast values used in these experiments were different for the 3 observers tested, selected on the basis of pilot data to determine the range of contrast values that did not generate 100% front or behind depth organizations. The contrast range for each observer was divided into six (observers Z. H. and S. S.) or eight (B. A.) intervals, presented in random order. Half of the trials contained stereopairs in which the direction of contrast favored matches behind the ground plane, and the other half favored matches in front of the ground plane. The patterns were presented for 1 s, and the task of the observer was to simply report whether the figure was seen in front of or behind the ground plane. In the (rare) instance that the pattern flipped from one depth to the other during a given trial, the observer was instructed to report the first depth that appeared.

*Results.* The results of this experiment are presented in Figure 10. One pattern, depicted by the filled squares, was predicted to appear in front of the background, thus conforming to the top horizontal line. The other pattern, depicted by the open triangles, was predicted to appear behind the background and thus conform to the lower horizontal line. For high values of contrast, this was indeed the pattern that was observed as can be seen in the right-hand sides of each graph. Points here lie on their respectively predicted horizontal lines. However, as the between-stripe contrast was reduced, there was a monotonic decrease in the percentage of patterns seen in front of the ground plane. Note that the only psychometric function that was affected by our manipulation of contrast was the pattern that was predicted to appear in front of the ground plane; the function predicted to appear behind the background followed the expected pattern for all values of contrast.

The amount of contrast at which observers exhibited a departure from the polarity-preserving prediction varied considerably between observers, having values of approximately 20%, 40%, and 80% for observers B. A., S. S., and Z. H., respectively. Presumably, these differences reflect forms of bias specific to a given individual, as has been observed previously for bistable wallpaper patterns (Julesz, 1971; Julesz & Chang, 1976). The important point to underscore here is that the same qualitative pattern of results were observed for all observers, irrespective of

---

[1] The patterns depicted in the figures of this article represent the inverse luminance relations used in the actual experiments we conducted. For example, the random-dot ground planes used in the experimental patterns consisted of white dots on a dark background, whereas the figures used in this article depict black dots on a white background. Observers experience both contrast relations in the same way, reinforcing our use of contrast as the relevant variable constraining the perceived organization of these displays.
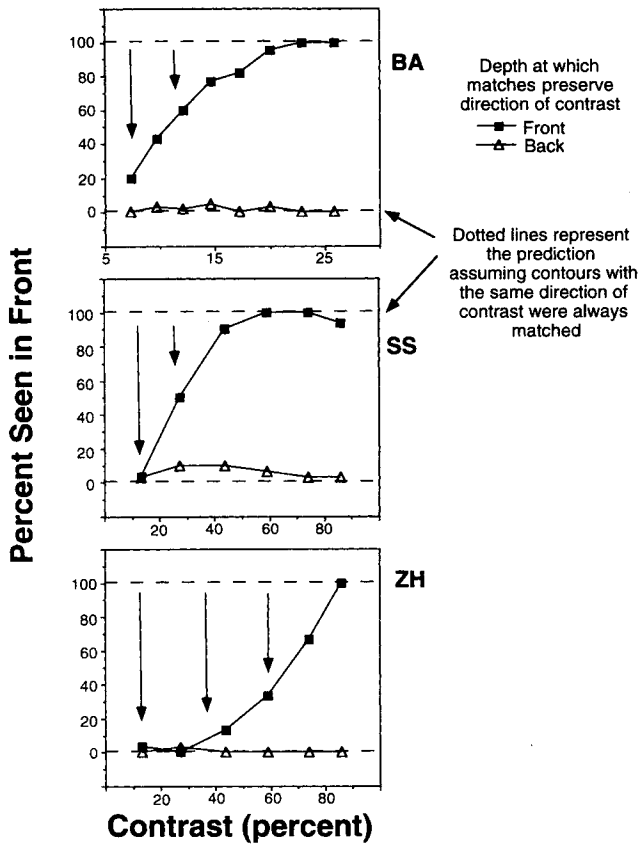
## Contrast Constraints in Binocular Correspondence



*Figure 10.* Results of Experiment 1 for the 3 observers, B. A., S. S., and Z. H. The abscissa represents the contrast of the adjacent light and dark bars within the wallpaper pattern. The dotted lines represent the predicted depth based on the assumption that the contrast polarity of the matched edges will be preserved. Note that only the pattern that is predicted to appear in front departs from the predicted pattern of results. This departure is not understandable by any existing theory of binocular matching.

individual differences in the quantitative values of contrast that led to a bias to perceive both patterns behind the zero-disparity background. Furthermore, the departures from the pattern of results expected on the basis of contrast polarity occurred for contrast values substantially above threshold for all observers. Thus, these results cannot be due to an insensitivity of early visual mechanisms to the contrasts presented in these images, as all of the contrasts used in these experiments were readily visible.

One concern with the use of relatively long display times was the possible contribution of vergence movements on the pattern of results. The use of 1-s display times was for the benefit of the naive observers that were not as experienced as the author (B. A.) in stereoscopic tasks. Such observers found the depth judgments more difficult for brief exposures. However, all of the experimental results have been replicated by the author for very brief display durations (less than 150 ms), and the same pattern

of results were always obtained. We therefore are confident that vergence was not playing a crucial role in any of the results we report in this article, a point that we make more forcefully in some of the experiments that follow.

What, then, caused this departure from the polarity constraint? Why did a monotonic variation in contrast result in such a profound bias in the binocular matches formed? Why were both low-contrast patterns seen behind the ground plane?

### Binocular Matching, Panum's Limiting Case, and Occlusion

The manipulation of contrast had two distinct effects in the patterns used in Experiment 1: First, the contrast between the light and dark stripes varied, and second, the contrasts along the wallpaper–background borders were changed. As the relative contrast between the stripes was reduced, the contrast of the light stripes relative to the background was increased. Indeed, the highest contrasts in the low-contrast patterns were the borders of the central wallpaper pattern (see Figure 8). The contrasts along the wallpaper–background borders of the light and dark stripes became increasingly similar as the between-stripe contrast was reduced. Thus, these borders became the strongest matches in the half-images (where strong refers to their contrast magnitude), which suggests that it was highly likely that they were matched. What impact should this have had on the depth organizations that were formed?

First, consider what should have happened if Panum's (1858/1940) double-matching rule was applied to the monocular contours in these patterns. This rule would match the wallpaper–background borders twice: once to each other and a second time to the contour that serves as the best match within the wallpaper pattern. The second matching problem is identical to the matching possibilities illustrated in Figure 9. Presumably, the double-matching scheme would operate under similar matching constraints as image features with unique matches. If contrast polarity was a necessary condition for binocular matching, then contours with the same contrast polarity should have been matched. Yet this is not what we found: Both patterns revealed a bias to appear behind the ground plane as the between-stripe contrast was reduced. If double matchings did occur, such a result implies that contours of opposite contrast polarity were matched, but only in one of the stereopairs. Thus, even if we allow for the possibility that the edges of the wallpaper pattern were matched twice, there is no coherent explanation of why a double-matching scheme would bias both patterns to appear behind the ground plane.

Our discussion to this point has focused on the edges in the stereopairs. It may be argued that it is not really the edges that are matched in binocular vision, but the *centroids* of the wallpaper stripes (defined as the point between the adjacent zero-crossings around which the first-order moment is zero; see Legge & Gu, 1989; Watt & Morgan, 1985). However, this would not explain our result either. This, too, would predict that the pattern depicted in Figure 9a should appear in front of the background, because the highest contrast regions (relative to the background) are the dark stripes. A general argument could be formulated as to why matching alone cannot explain the fact that both wallpaper patterns appear behind the background. If

a matching bias could be attributed to some spatial asymmetry in the monocular inputs, then this depth bias should reverse when the half-images are interchanged. This, however, was not observed. Hence, something other than matching is needed to understand the perceived depth of this pattern. This result provides a strong counter to the predictive capacity of current models of stereopsis. Whereas Demonstration 1 may be considered a predictive triumph of current models, Demonstration 2 (and the results of Experiment 1) provides direct evidence that there is something missing from current conceptualizations of stereopsis. If we only focus on the process of matching, these results are impossible to comprehend.

To highlight the remarkable nature of the results of Experiment 1, consider again the pattern of Figures 9a and 9b in Figure 11. Here, the neighborhoods of the edges are represented as the input into a generic disparity-detecting mechanism to underscore the failure of such models in accounting for the observed percept. No assumptions are made about the particular structure of the detector, so it is drawn as a circularly symmetric RF for simplicity. Figure 11a illustrates the matches consistent with perceiving the wallpaper pattern behind the background. Note that both the contrast polarity and contrast magnitude are extremely different for this pairing. In contrast, the matches consistent with the pattern appearing predominantly in front of the background have a higher degree of similarity: The contrasts are different, but the polarities are the same. This organization is therefore the better matching scheme available. This analysis also holds even if Panum's (1858/1940) double-matching thesis

is applied to this figure. However, this is not the organization chosen by the visual system. Why? In the section that follows, we suggest that the visual system may be sensing the existence of an occlusion configuration with mechanisms structured to detect a correlation–decorrelation boundary.

### Occlusion Constraints in Binocular Correspondence

Our results clearly indicate that there is something missing from accounts of stereopsis. Apparently, there is more to stereopsis than binocular matching; there is also the problem of detecting the breakdown in correspondence that occurs at occluding contours. Indeed, the perceptual decision about the depth of the wallpaper pattern also involved determining the occlusion relationships in the images. It is here that it is possible to find a bias.

A stereoscopic occlusion relationship is characterized as a breakdown in binocular correspondence in the cyclopean "image" structure. This can occur in a number of ways. Left-eye-only features can appear either to the left or right of a fused region, as can right-eye-only features. However, it is important to note that only two of the four configurations represent occluding contours. These arise when left-eye-only features appear to the left of a (nearer) occluding surface, and right-eye-only features appear to the right of a (nearer) occluding surface (see Figure 1). The other two configurations, right-eye-only features to the left of a disparity or left-eye-only features to the right of a disparity, do not represent occluding contours but,
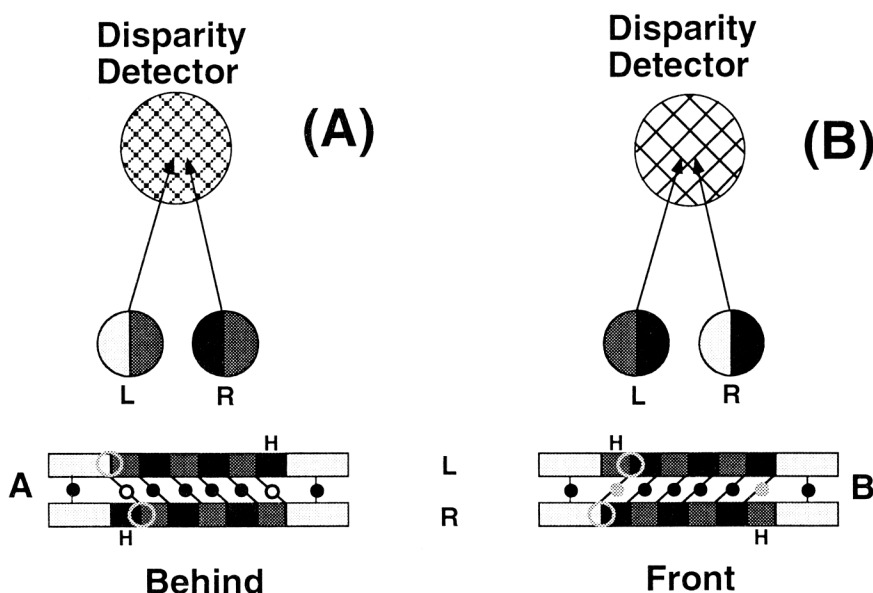


*Figure 11.* A schematic demonstrating the failure of theories to predict the pattern in Figure 9a to appear behind the background (top dotted lines in the graphs of Figure 10). The neighborhoods of the edges from the two eyes are projected into a disparity-sensitive unit structured to detect the similarity of the features. The spacing of the grating pattern within the receptive field of the disparity detector depicts the degree of mismatch between the inputs: The greater the spacing, the smaller the mismatch. Panel (A): The edge matches that would putatively have to form for the pattern to appear behind the background. This matching scheme differs widely in both the magnitude and polarity of contrast. Panel (B): The matches consistent with the pattern appearing in front of the background. This match also has a large difference in the magnitude of contrast, but the polarity is the same, and is therefore the better matching scheme available. However, this is not the organization chosen by the visual system. H = half-occlusion; L = left eye; R = right eye.

rather, portions of the occluded surface. Even though these latter two possibilities generate a breakdown in binocular correspondence, such correlation–decorrelation boundaries are perceptually invisible: In these configurations, the monocular features appear at the same depth as the binocular features, forming an unbroken, continuous surface (see Figure 1). These configurations therefore do not contribute to the depth differences that arise at occluding contours. This fact is crucial in explaining the biases we observed in our experiments.

How might occlusion relationships be detected at an early stage of binocular processing? We suggest that there exist structures that are specialized to sense the local structure of an occluding contour–half-occlusion boundary, such as those depicted in Figure 12. Such units would allow for the detection of occlusion configurations in parallel with mechanisms specialized to detect disparity. In keeping with the current trend of formulating local "detectors" to explain perceptual sensitivities, we postulate the existence of cyclopean RFs that respond optimally to the binocular correlation–decorrelation boundary generated at occluding contours. We contend that only two types of RFs are used for detecting occluding contours: A near surface occludes a surface that appears to the left of a (farther) occluded surface, or a near surface occludes a surface that appears to the right of the occluding surface. No assumptions are made as to how such RFs emerged. For example, it is possible that such RFs were the products of some processes of self-organization during the early development of binocular vision in infancy. Our theory is not affected by such issues. Our hypothesis is simply that the RFs representing the occluding contour configurations can explain the depth biases we observed in Demonstration 1.

The RFs we propose consist of two parts: a disparity-sensitive region that is responsible for detecting the binocular correlation
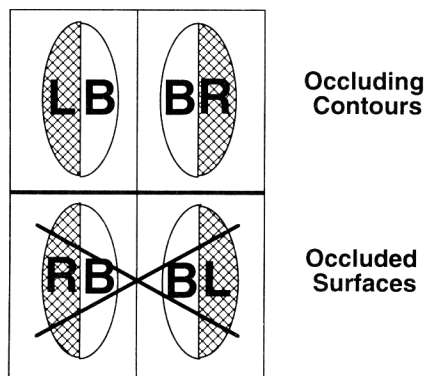


*Figure 12.* The four possible pairings of receptive field (RF) zones that could sense the local cyclopean structure at occluding contours. Open areas denote regions where binocular matching is sensed. Hatched areas denote sensitivity to uncorrelated binocular input, that is, unpaired monocular (right [R] eye or left [L] eye) stimulation. Only the upper combinations signal occluding contours (refer to Figure 1). Pairings below are deemed irrelevant for the perception of occluding contours, forming the invisible boundary between paired and unpaired regions on occluded surfaces. Hence, only the top two RFs should introduce a relative depth bias at a correlation–decorrelation boundary. B = binocular lobe.

of the occluding contour (labeled B in Figure 12) and a region sensitive to a failure of binocular correspondence (i.e., right-eye [R] or left-eye [L] half-occlusions). The depth relationships of the components of the RFs are fixed, at least in terms of the sign of the depth: The monocular regions in which binocular correlation breaks down always belong to a surface behind the occluding contour. For such mechanisms to be sufficiently general, these units would have to cover a range of orientations (as all orientations except those purely horizontal relative to the observer's line of sight generate half-occlusions) and sizes (spanning the range of disparities). Such units could signal the relative (ordinal) depth relationships of occluding contours at a very early stage in processing.

To understand our occlusion hypothesis, consider the stereopair depicted in Figure 9a. The two possible depth organizations of the wallpaper pattern involve choosing between two possible occlusion configurations. When the wallpaper pattern appears in front, two of the stripes in the wallpaper pattern will be the occluding contours. When the wallpaper pattern appears behind the background, the edges of the background with the wallpaper pattern form the occluding contours. Our hypothesis is that the organization that generates the strongest correlation–decorrelation boundary will determine the occlusion relationships in these patterns. To see how, we apply our hypothetical RFs to the wallpaper pattern depicted in Figure 9a. This analysis is depicted in Figure 13. By hypothesis, only two of the RF profiles sensing correlation–decorrelation boundaries correspond to the occluding contours, so only the responses of these two types of RFs need to be considered. In keeping with our emphasis on the importance of the monocular neighborhoods as the input to binocular matching, we depict the monocular inputs into the lobes of our RFs as the entire neighborhood around the monocular edges. At the left side of the wallpaper pattern, the edge formed by the wallpaper pattern–random-dot background generates a high-contrast border at the depth of the background in both patterns. There is a small amount of mismatch in the contrasts of these edges, as depicted by the faint hatched regions in the lobe of the RF sensitive to the correlated input from the occluding contour. To the immediate right of this edge is the lobe sensitive to the binocularly uncorrelated input of the half-occlusions. This lobe is shown as receiving input from both eyes to demonstrate the result of attempting to correlate the features from the two eyes in this region. This corresponds to the features that would have to be matched for a disparity assignment to account for the percept of this contour at this depth. These inputs differ both in the direction of contrast and the magnitude of contrast, generating a strong decorrelated signal. However, an RF of the type depicted, that is, one that senses correlation on the left and decorrelation on the right, would respond vigorously to this pattern of stimulation. Because the left side of the RF is interpreted as the occluder, the response of this RF would bias the regions to the immediate right to appear behind the background, which in turn would bias the wallpaper pattern to this depth. A similar analysis holds for the right side of the wallpaper pattern, except it is now the other asymmetric RF that responds (sensing the left-eye-only features to the left of the fused zero-disparity edge). Again, this would bias the wallpaper pattern to appear behind the background. This is the bias that was observed experimentally.
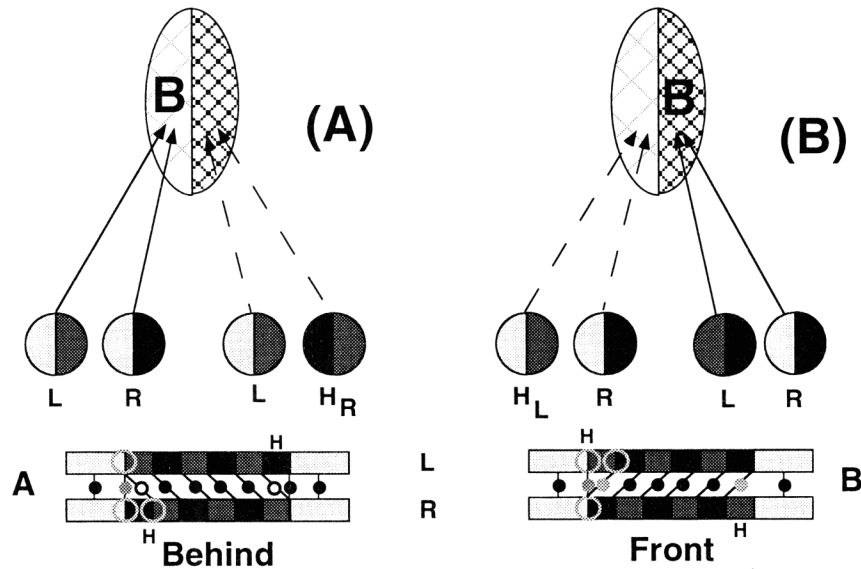
*Figure 13.* Response of our hypothetical receptive field (RF) models for the two possible occlusion configurations. The lobes sensitive to the half-occlusions (H) are shown as receiving input from both eyes to demonstrate the result of attempting to correlate the features from the two eyes in this region. In essence, this representation amounts to attempting to perform the double matches represented in Figure 11 in the two different lobes of our hypothetical RFs sensing occlusions. Panel (A): The edge of the wallpaper pattern creates a strong correlation–decorrelation boundary. The binocular lobe (depicted by the B in the RF pattern) receives inputs that are very similar, differing only slightly in the magnitude of contrast. The neighboring lobe receives input that is strongly uncorrelated. This RF would respond quite vigorously to the pattern depicted, biasing the wallpaper pattern to be part of an occluded surface. Panel (B): The response of the other proposed RF that would sense the occluder in front of the ground plane. Notice that the lobe that would be optimally sensitive to a correlated input would be very weakly activated because there is a strong contrast mismatch of the inputs, and the lobe that "prefers" uncorrelated input would receive very correlated inputs. This pattern of stimulation is of the opposite type preferred by this RF. Hence, the RF in (A) would be much more active, biasing the wallpaper pattern to appear behind the background. A similar analysis holds for the possible matches on the right-hand side of the figure. L = left eye; R = right eye.

Compare this response with the RF that would sense the occlusion configuration for the condition in which the wallpaper pattern appears in front of the background. Because of the large contrast difference between the edges that would need to be matched (in the binocular lobe represented in Figure 13b), this RF would respond very weakly: The lobe sensitive to the correlated input would be receiving stimulation that differed widely in contrast, and hence would not respond. Indeed, the pattern of stimulation in this RF is opposite of its preferred pattern, with a strong correlation on the left (i.e., in the lobe putatively sensitive to decorrelation) and strong decorrelation in the lobe that responds optimally to binocular correlation. Hence, this RF should not be excited, yielding a strong competitive advantage for the pattern depicted in (A).

It is interesting to note that the opposite result would be predicted if we applied a smoothness constraint to the half-occluded features. The high-contrast borders at the zero-disparity background are clearly the strongest disparity signals in the image. If a smoothness constraint was applied and these disparity signals spread into the half-occluded regions, this would create a bias for the half-occlusions to appear at the depth of the background (because this is the depth of the flanks). In turn, this should bias the rest of the wallpaper pattern to appear in front

of the background, as this is the geometric configuration that would be consistent with the half-occlusions appearing at the depth of the background. This was not observed, providing further support for our hypothesis that there exist mechanisms that are tuned to respond to the binocular structure of occlusion configurations.
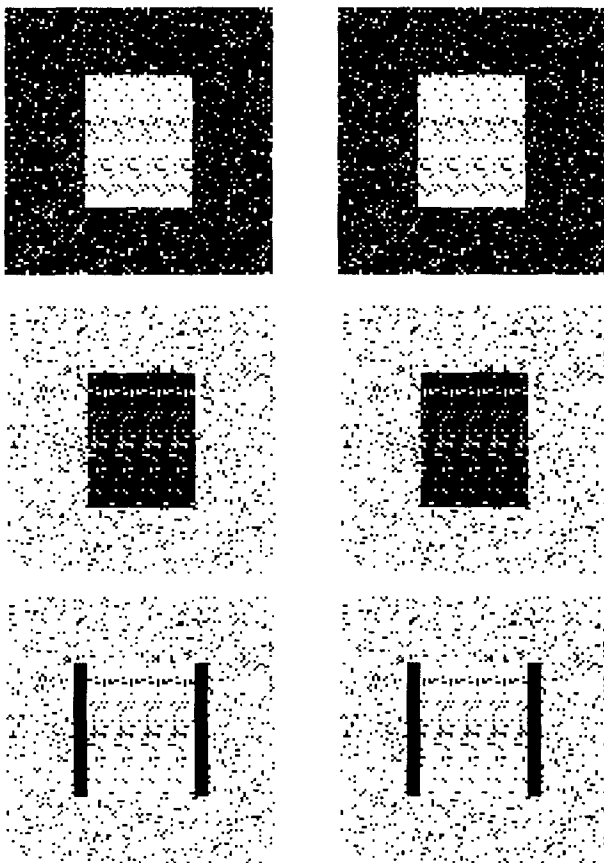
The results of Experiment 1 cannot be understood with a model that only seeks to maximize the degree similarity between binocular features. The explanation of our results that we have forwarded requires that some information about occlusion configurations be integrated with theories of disparity detection. In the experiments that follow, we provide further support for this thesis and eliminate alternative explanations of the depth bias observed in Demonstration 2.

### Demonstration 3: The Lateral Influence of High-Contrast Matches on Bistable Wallpaper Patterns

There is at least one alternative interpretation of Experiment 1 that we must consider before settling on the occlusion hypothesis we have forwarded. A number of observers have spontaneously reported experiencing percepts of transparency when fusing the stereograms of Figure 8, but not with the stereograms of

Figures 5b and 5c. The reader may experience this by comparing the stereograms seen in Figures 5 and 8. In Figure 8, there is a hint of a transparent surface located at the depth of the background, with the stripes appearing visible behind this surface. It is possible that the behind bias observed in Experiment 1 may be related to the perception of transparency. The formation of a transparent surface at the ground plane may bias contours to appear behind this depth, as this would be the most globally consistent interpretation possible: A surface cannot appear transparent without something being visible behind it. We therefore needed to construct a new stereogram to determine whether transparency played a significant role in the results of Experiment 1.

*Method.* In this demonstration, we constructed a random-dot version of the wallpaper pattern and flanked this pattern with either two vertically oriented high-contrast luminance stripes or simply reversed the polarity of the entire background to create a high-contrast border. These stereopairs are illustrated in Figure 14. This figure was constructed by alternating two stripes, say A and B, each filled with a unique random texture. The stripes are repeated ABAB . . . in one eye, whereas the other eye is defined by BABA . . . . Thus, as with the patterns used in Experiment 1, the stimulus may be considered a (random-

dot) square wave, with disparity introduced by a phase shift of one half cycle. The width of the random-dot stripes and number of cycles in the pattern were identical to those used in Experiment 1. Similar effects were observed for all "spatial frequencies" (i.e., stripe widths) and number of cycles we examined, so the restriction to a single stripe width and disparity was done for simplicity of presentation. There was nothing inherent in the random-dot wallpaper pattern that would lead to a bias in organizing the pattern in front of or behind the ground plane. However, note that the RFs we have postulated would respond differentially to these two patterns (see Figure 15). If our occlusion hypothesis is correct, then the most active RF will be the one that senses the zero-disparity background as the occluder, because this creates the strongest correlation–decorrelation boundary. This, in turn, should bias the wallpaper pattern to appear behind the zero-disparity background.

*Results.* Approximately 200 observers have viewed these patterns in our laboratory and public demonstrations, and all observers that could perceive stereoscopic depth reported that the predominant organization of the wallpaper pattern was behind the ground plane. Furthermore, no observer has reported percepts of transparency when viewing these patterns. We may therefore conclude that transparency was not playing a role in the results of Experiment 1, eliminating one alternative explanation of the depth biases we observed.



*Figure 14.* Patterns used to test whether the percept of transparency evident in Figure 8 was crucial for this pattern to appear behind the background. All of the patterns presented have been reported as appearing predominantly behind the background, yet none of these patterns generate percepts of transparency. This eliminates transparency as an explanation of the depth biases observed.

### Experiment 2: Correlation–Decorrelation Occlusion Boundaries or Monocular Contrast?

The crucial link between our occlusion hypothesis and the behind depth bias observed in Demonstration 2 and Experiment 1 is the ability to sense a correlation–decorrelation boundary that may be interpreted as an occlusion configuration. This interpretation predicts an anisotropy in the effects of the contours that we present. Because one's eyes are displaced horizontally on one's head, correlation–decorrelation boundaries, corresponding to occluding contours, will only be generated by contours that have some degree of vertical orientation relative to an observer's line of sight. Thus, if the depth biases we observed were due to mechanisms structured to detect these boundaries, then vertically oriented flanks should have a much stronger influence on the perceived depth of the wallpaper pattern than horizontally oriented flanks. An alternative explanation of our result is that there is simply a bias for high-contrast borders to be interpreted as occluding contours. To test this idea, we compared the effect of flanking a bistable RDS with either vertical stripes or horizontal stripes of equal contrast (see Figure 16). If the depth bias we have discovered was simply the consequence of interpreting high-contrast borders as occluders, then there should be little or no difference between the effects of vertical and horizontal flanks. In contrast, our occlusion hypothesis predicts a strong asymmetry: Vertical flanks should cause a strong bias for the wallpaper pattern to appear behind the background.

*Quantifying depth bias.* Our experiments required a means of quantifying the strength of the depth bias we observed in Demonstration 3. We chose a method first used by Julesz and Chang (1976). Julesz and Chang introduced unambiguous dots of a fixed disparity into a random-dot version of the wallpaper pattern and measured the frequency with which a particular amount of bias overcame observers' natural depth biases. *Natural bias* refers to the fact that observers typically have some
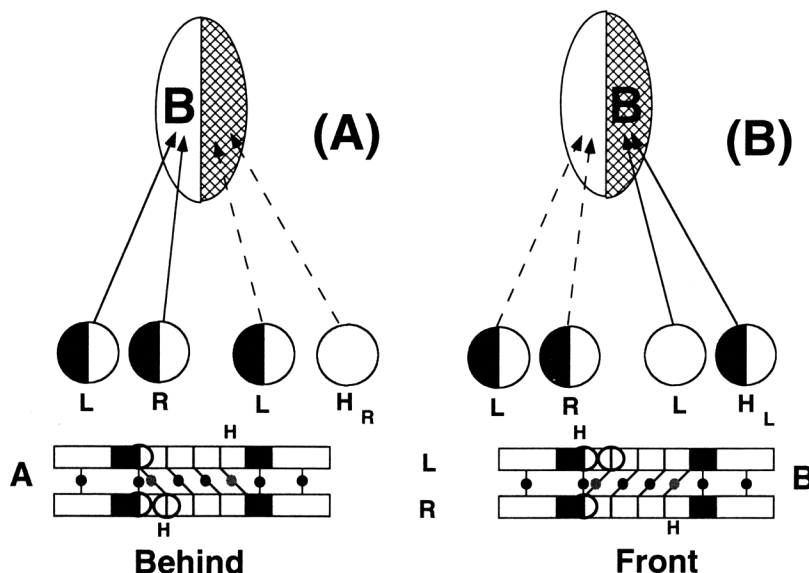
*Figure 15.* Response of our hypothetical receptive field (RF) models for the two possible occlusion configurations of the pattern in the bottom of Figure 14. As in Figure 13, the local neighborhoods of the half-images are used as input for the two halves of the RFs. The lobes sensitive to binocular decorrelation are shown as receiving input from both eyes to demonstrate the result of attempting to correlate the features from the two eyes in this region. Again, the hypothetical RF depicted in (A) receives nearly optimal input, whereas that in (B) has a pattern of stimulation opposite of its preferred input. As before, the wallpaper pattern is biased to appear behind the background. This bias is not predicted by models that do not include mechanisms sensitive to correlation–decorrelation boundaries. B = binocular lobe; L = left eye; R = right eye; H = half-occlusion.

tendency to perceive a bistable random-dot wallpaper pattern either in front of or behind the ground plane. Julesz and Chang found that an observer's natural bias could be overcome with a sufficient number of dots with an unambiguous disparity opposite the observer's natural bias. One of the attractive features of this method was that the psychometric functions they obtained were monotonic functions of the number (or percentage) of bias dots introduced, which served as a metric of bias "strength." We have adapted this method to quantify the relative strengths with which the high-contrast flanks biased the wallpaper pattern to appear behind the zero-disparity background.

Two observers (1 naive and 1 of the authors [B. A.]) were presented with 12 stereograms that differed in the number of bias dots (i.e., dots with an unambiguous disparity) introduced in the "forward" plane (0, 10, 20, . . . , 100, 110). The forward plane refers to 9 arc min crossed disparity, corresponding to the disparity of the wallpaper pattern when it appeared in front of the ground plane. Each stereogram was presented 30 times in random order for 500 ms, preceded and followed by a zero-disparity random-dot pattern. The random-dot pattern served as a fixation pattern and backward mask. The observer's task was simply to report whether the wallpaper pattern appeared in front of or behind the ground plane. Three types of patterns were compared: a random-dot wallpaper pattern, a random-dot wallpaper pattern flanked by vertical zero-disparity stripes, and a random-dot wallpaper pattern flanked by horizontal zero-disparity stripes (see Figure 16). The luminance and spatial separation of the horizontal and vertical flanks were equated.

*Results and discussion.* The data from this experiment are presented in Figure 17 for the two observers. These graphs reveal a strong anisotropic effect of horizontal versus vertical flanks in biasing the perceived depth of the wallpaper pattern. The vertical flanks (open squares) clearly biased the wallpaper pattern to appear behind the ground plane, whereas the horizontal stripes (filled circles) had virtually no influence on the perceived depth of the pattern. Note that the psychometric functions of the horizontal stripes and the unbiased wallpaper pattern (open triangles) were essentially identical, whereas the vertical stripes introduced a strong bias for the wallpaper pattern to be seen behind the ground plane.

This finding supports our hypothesis that the depth bias we observed in our previous experiments and demonstrations was linked to mechanisms sensitive to the structure of occlusion relationships: Only those contours that generated cyclopean correlation–decorrelation boundaries exhibited a bias to appear behind the background. The horizontal stripes did not generate unmatchable regions that could be interpreted as half-occlusions. Consider the response of our hypothetical RF model to the vertical stripes. On the left side of the wallpaper pattern, the high-contrast flank would strongly excite the matched portion of the RF depicted in Figure 15a. Furthermore, the failure to match regions neighboring this match would also excite the lobe sensitive to decorrelation. The configuration in which the wallpaper pattern would appear in front of the background requires matching regions that differ widely in contrast in the binocular lobe and receives a highly correlated input in the lobe
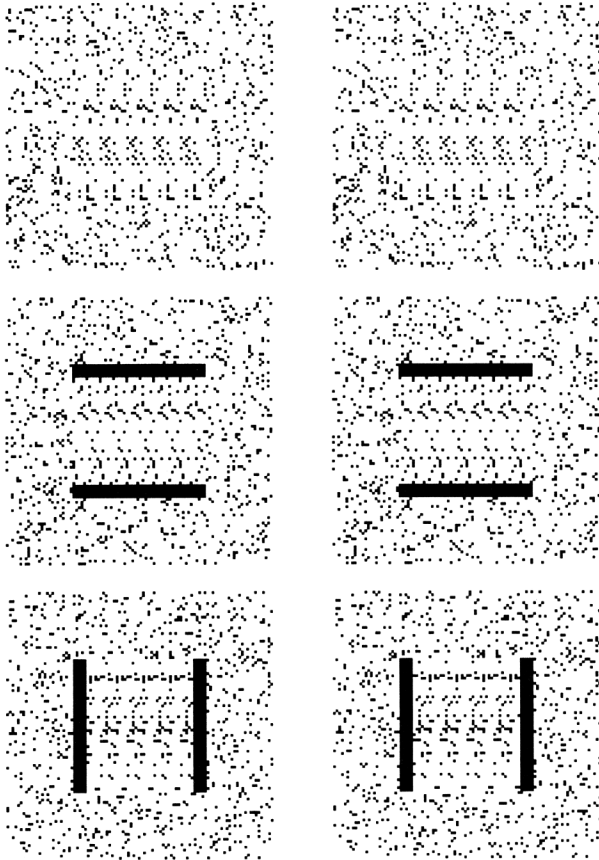
*Figure 16.* Patterns used to determine whether the depth biases were related to the correlation–decorrelation boundaries created at occluding contours or whether it was simply a bias to treat high-contrast borders as occluders. See text for details.

sensitive to decorrelated input. Hence, this RF should not respond. As before, this gives a competitive advantage to the RF depicted in Figure 15a, biasing the wallpaper patterns to appear behind the background.

It is also worth mentioning that the results we obtained in this experiment cannot be understood as the consequence of vergence biases. Such biases should be uniform across the different stimulus conditions and could not account for the differential effect observed for vertically oriented high-contrast flanks.

*Phenomenological observations.* A number of comments are warranted about the phenomenology of the perceived depth in the wallpaper patterns used in Experiment 2. For the (random-dot) wallpaper patterns flanked by vertical stripes, there was a strong tendency for the wallpaper pattern to appear initially behind the ground plane and then to be "pulled up" to the forward plane by the bias dots after durations of about 500 ms. Indeed, the display duration of 500 ms was chosen to reduce the frequency with which the wallpaper pattern perceptually flipped from behind to in front of the ground plane during a given trial. However, even if the pattern did change its perceived depth within a trial, observers were instructed to report the initial depth of the target. We would like to emphasize the implications of this fact, the tendency for the wallpa-

per pattern to initially appear behind the ground plane, for models of when occlusion information is processed stereoscopically. If the interpretation of our results is correct, then this phenomenology implies that occlusion relationships are processed at the earliest stages of stereoscopic processing, that is, during the binocular matching phase. Such an interpretation provides support for our contention that occlusion is processed in parallel with binocular disparity.

## Experiment 3: The Dependence of Depth Bias on Contrast

In Experiment 3, we more thoroughly explored the sensitivity of our hypothetical RFs to the contrast relationships at occlusion borders. We used patterns identical to the vertical flank pattern used in Experiment 2 but systematically varied the contrast of the flanks relative to the random-dot background. At least two effects are created by this manipulation of contrast. Increasing the contrast of the vertical flanks simultaneously boosts the strength of the match at the depth of the flanks and increases the interocular contrast difference between the borders of the wallpaper pattern and the boundaries of the random-dot stripes (see Figure 16). The construction of RF models for sensing occlusion relationships implies that the response of these mechanisms should be scaled by the strength of the correlation–decorrelation cyclopean boundary. Our hypothetical mechanisms predict that the tendency for the wallpaper pattern to appear behind the ground plane should increase when the contrast of the flanks is increased and diminish when the contrast of the flanks is diminished. Experiment 3 was conducted to test this hypothesis.

*Method.* We measured the tendency for wallpaper patterns to be seen behind vertical flanks for four values of flank contrast (0%, 33%, 82%, and 94%). These contrast values were chosen on the basis of pilot observations as values that did not lead to 100% behind responses. Two observers (1 naive and 1 of the authors [B. A.]) were presented with 12 stereograms that differed in the number of bias dots introduced in the forward plane (0, 10, 20, . . ., 100, 110). Within a block of trials, flank contrast was held constant, and observers were tested with the 12 amounts of bias dots presented in random order. Each data point represents 30 responses to a single amount of bias and a specified flank contrast. Observers performed four blocks of trials, one for each value of contrast. The blocks were presented in a pseudorandom order. The viewing distance, target size, and disparities were identical to those used in the first two experiments (1.2 m, 3°, and 9 arc min, respectively).

*Results.* The results of Experiment 3 are presented in Figure 18 for the two observers. These graphs reveal a strong relationship between the contrast of the vertical stripes and the tendency for the wallpaper pattern to be seen behind the zero-disparity background: The bias for the wallpaper pattern to appear behind the background diminished systematically as the contrast of the vertical stripes was reduced.

The results of this experiment provide converging support for the existence of local mechanisms structured to sense the strength of cyclopean correlation–decorrelation boundaries. The contrast dependence documented here lends credence to our suggestion that the depth biases we observed are created by early visual mechanisms, as it would be unlikely for such a dependence to occur for higher level cognitive processes. This is supported by the phenomenology of the perceived depth of these patterns as well: The depth bias was observed as soon as stereoscopic depth was resolvable. Finally, the observed suprathreshold contrast dependence greatly
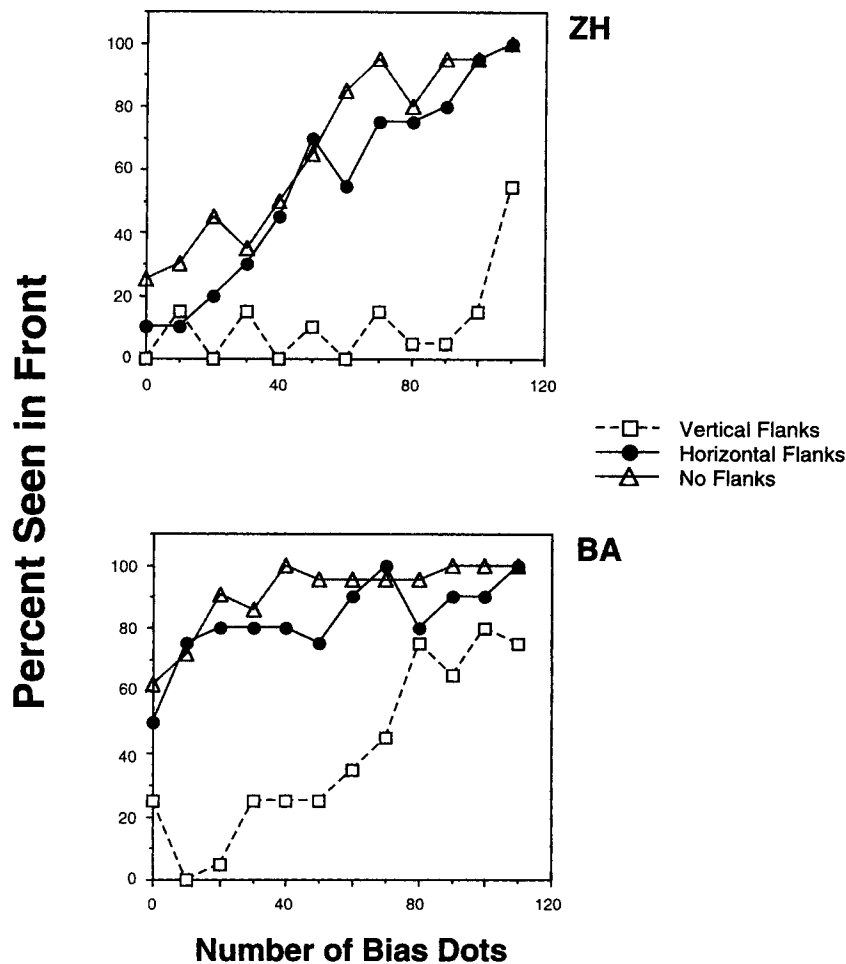
## Vertical/Horizontal Anisotropy



*Figure 17.* Results of Experiment 2. Both observers (Z. H. and B. A.) reveal a strong anisotropy between the role of vertical versus horizontal high-contrast contours in biasing the perceived depth of the wallpaper patterns. The unbiased random-dot wallpaper pattern is virtually indistinguishable from the pattern with horizontal flanks, whereas the vertically oriented flanks induced a strong bias for the wallpaper pattern to appear behind the background. This provides support for our theoretical link of these results to mechanisms sensitive to the breakdown in correlation that occurs at occluding contours with some degree of vertical orientation.

diminishes the possibility that vergence biases had an impact on our results, as vergence biases should be uniform over all values of contrast tested.

### Experiment 4: Half-Occlusions and Eye-of-Origin Information

The experiments and demonstrations described above have revealed depth biases for a bistable wallpaper pattern to appear behind the background. In our final experiment, we attempted to construct a stereogram in which the occlusion relationships would bias a wallpaper pattern to appear in front of a zero-disparity ground. In this experiment, we again used random-dot wallpaper patterns flanked by vertically oriented stripes. How-

ever, we varied the contrast of the flanks differentially in the two eyes. In one stereopair, the left eye had a larger contrast along the left side of the central figure, whereas the right eye had a larger contrast along the right side of the central figure. The contrast relations were reversed in the second stereopair. Thus, the relevant difference between the two stereopairs was which eye contained the higher contrast flanks on a given side of the wallpaper pattern, that is, the eye-of-origin of the greater contrast. These stereopairs are shown in Figure 19. The rationale for this experiment can be understood by considering Figure 20, where we depict the response of our hypothetical RFs to the organization in which the wallpaper pattern would appear in front of the background. Note that our RF models predict that the front depth organization should be easier to achieve with the pattern
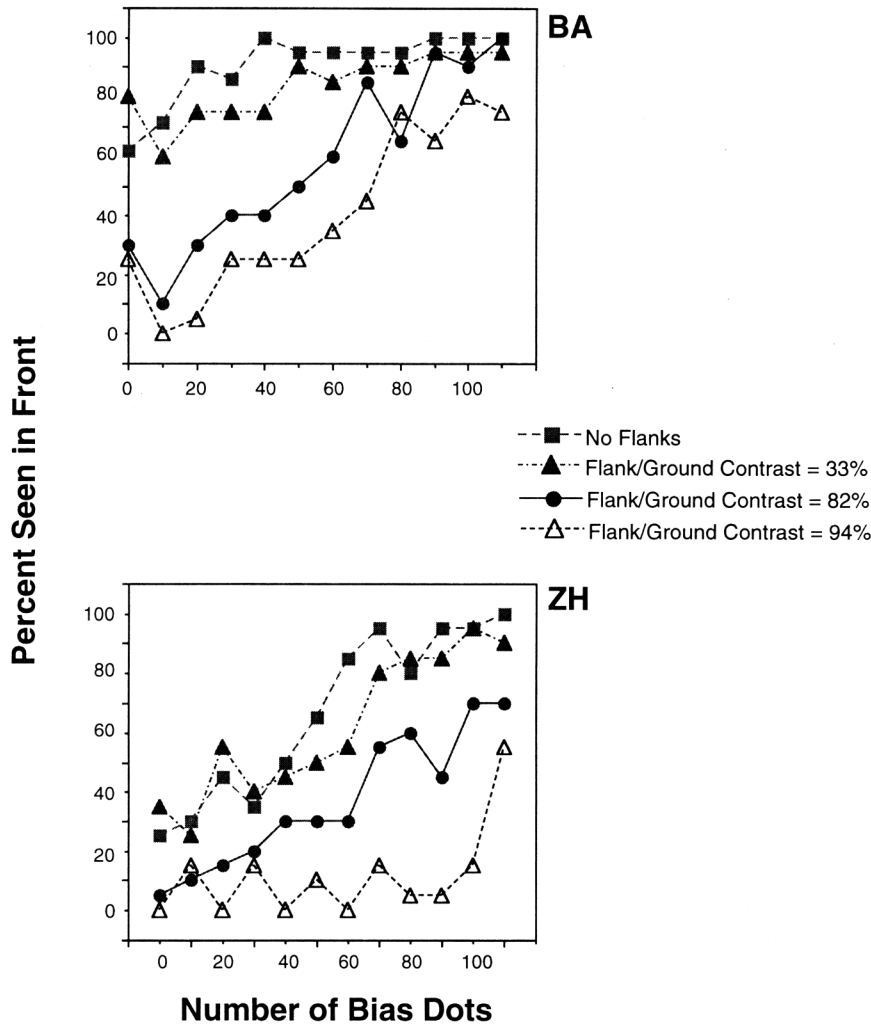
## Effect of Flank Contrast on Binocular Matching



*Figure 18.* Results of Experiment 3 for 2 of the observers, B. A. and Z. H. The strength with which the wallpaper pattern appears behind the background depends on the contrast of the vertical flanks relative to the (mean) contrast of the wallpaper pattern. Note that all values of contrast are well above threshold, so the observed dependence does not simply document a transition of the flanks from invisible to visible. The contrast dependence of this bias provides converging support that the biases we have observed represent the operation of early visual mechanisms sensitive to properties such as contrast.

that had the strongest contrast flank on the left side of the left stereogram and the right side of the right stereopair (Figure 20b). The correlation–decorrelation boundary created by this pattern of stimulation would more optimally stimulate the RF that would sense the near configuration of the wallpaper pattern. Note, however, that there is no bias within the wallpaper pattern for either organization to appear with any degree of greater frequency than the other.

*Method.* Observers were presented both stereopairs simultaneously, and in a two-alternative forced-choice task, they judged which of the two stereopairs was more likely to be seen in front of the zero-disparity background. The occlusion correct answer was varied randomly from top to bottom of

the display, such that for half the trials it appeared on the top and the other half it appeared on the bottom. A correct response corresponded to the stereopair in which the left eye had a larger contrast along the left side of the central figure, whereas the right eye had a larger contrast along the right side of the central figure (see Figure 20b). There was no time limit for each judgment. Three observers served as subjects, 2 of which were naive (Z. H. and P. C.), whereas the third was 1 of the authors (B. A.).

*Results.* The results of this experiment for one value of interocular contrast difference are presented in Figure 21 for the 3 subjects. Observers are remarkably consistent in choosing the "occlusion appropriate" pattern as the one seen predominantly in front, that is, the stereogram in which the left eye has the stronger contrast along
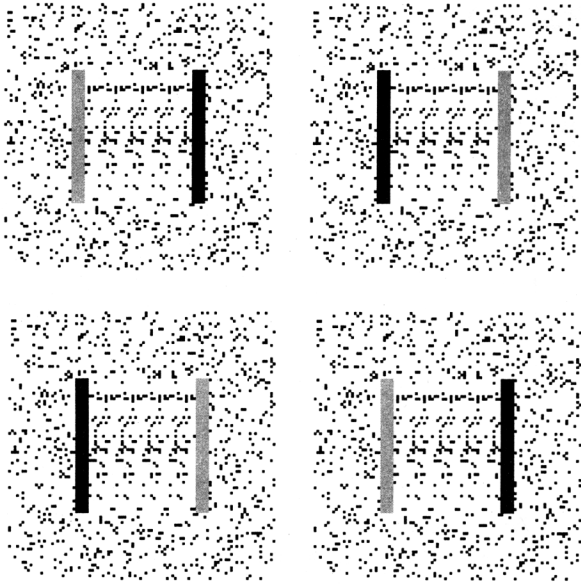
*Figure 19.* Patterns used in Experiment 4. See text and Figure 20 for rationale behind this stimulus construction.

the left side of the wallpaper pattern and the right eye has the strongest contrast along the right side of the wallpaper pattern. This choice is consistent with the prediction of our hypothetical RF models, as shown in Figure 20. Remarkably, observers can exploit interocular contrast differences as a source of information specifying occlusion relationships in the absence of any differences in the disparities of the image features.

One concern with these displays was the possibility that the differential contrast of the stripes in the half-images initiated vergence movements to bring the two high-contrast stripes into binocular correspondence. Note, however, that such movements are inversely correlated with the preferred depth organization of the two patterns: The pattern that is selected as the pattern most likely to appear in front of the ground plane requires a divergence eye movement to fuse the two high-contrast stripes. A divergence eye movement would bias the wallpaper pattern to appear behind the ground plane, as a divergence of the eyes would bring the far depth of the wallpaper pattern closer to the horopter. Thus, although we did not measure vergence movements, it again seems highly unlikely that vergence could account for the findings of this experiment.

## Section 3: Summary of Experiments and Theoretical Implications

The results reported in the previous section provide new challenges for models of stereopsis. The use of ambiguous wallpaper patterns allowed us to assess the kinds of constraints that influence binocular matching. Our first result (Demonstration 1) may be viewed as a triumph of extant models in predicting how matching is achieved. However, the remainder of our results reveal the inadequacy of current models in capturing the kinds of information that can influence binocular matching.

We have forwarded the hypothesis that the inability of current models to predict our results stems from their failure to incorporate the information contained in occlusion relationships into the "machinery" involved in establishing binocular correspondence. If our
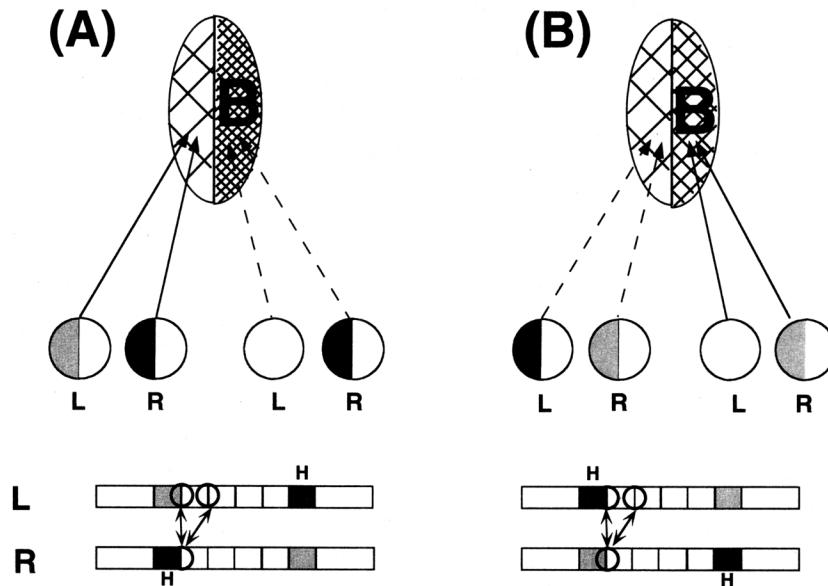


*Figure 20.* This figure depicts the response of the receptive field (RF) structured to sense a "near" occluder to the two patterns presented in Figure 19. Panel (A): This pattern would create a pattern of stimulation opposite of that preferred by this RF. Note that the binocular lobe would receive inputs that differ widely in their local contrast. Panel (B): The result of interchanging the half-images in (A). Note that the contrasts that feed into the binocular lobe of the RF are more similar than that of the pattern depicted in (A). Our RFs predict that observers would more readily perceive this pattern in front of the ground plane than in (A). B = binocular lobe; H = half-occlusion; L = left eye; R = right eye.
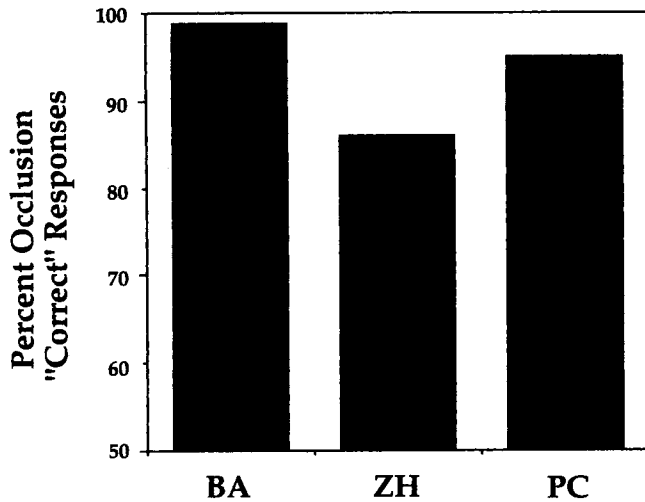
*Figure 21.* Results of Experiment 4 for the 3 observers, B. A., Z. H., and P. C. These results are consistent with the predictions of our receptive fields as depicted in Figure 20.

interpretation of the depth biases we observed is correct, we have demonstrated that information about occlusion influences the earliest stages of stereoscopic processing. The phenomenology associated with our experiments reinforces this conclusion. Observers uniformly reported that the bias for the wallpaper pattern to appear behind the background occurred as soon as stereoscopic depth was resolvable for all patterns exhibiting this bias. Thus, there was no phenomenological evidence to support the hypothesis that the depth biases we observed, putatively a reflection of occlusion constraints, operated through a form of feedback, favoring our contention that information about occlusion is processed in parallel with binocular disparity.

How is it possible for occlusion relationships to influence binocular matching? The great puzzle involved in this question is understanding how regions characterized by a breakdown in binocular correspondence can influence regions that are matchable. Part of the solution to this puzzle arises from the fact that only two of the four possible combinations of correlation–decorrelation boundaries correspond to occluding contours. Sensitivity to the breakdown in binocular correspondence occurring at occluding contours can provide an early indication of the ordinal depth of surfaces in a scene, contributing both to image segmentation and perceived depth. The need to identify occlusion relationships early in binocular processing demands a conception of stereopsis in which such regions can contribute positive information. This seems mandated by the results of the previous section as well as a previous report (Gillam & Borsting, 1988). Indeed, many of the procedures advocated to facilitate disparity processing interfere with the resolution of occluding contours. The application of smoothness constraints is particularly problematic in occluded regions, as we have already discussed. If these processes are retained, some separate process is needed to understand the contribution of occluding contours to stereoscopic depth.

The importance of discontinuities, or edges, for processes such as image segmentation is widely recognized. However, the effective detection of discontinuities in domains defined by

multiple views, such as stereopsis and motion, has yet to be solved. The difficulty emerges from the fact that what is typically held as the quantity of interest, disparity or motion, is not defined at the edges of occluding contours. Consequently, models that focus on these quantities necessarily perform the worst at these edges. The fact that such edges do seem vivid and clear to human observers implies that there must be some mechanisms that sense this breakdown in either disparity or motion.

We have described some general implementation strategies for sensing the information contained in occlusion configurations when using complex primitives as the input to stereopsis. Our contention has been that a rapid determination of features as matchable or unmatchable requires the use of complex primitives as the input to stereoscopic mechanisms and was explicit in our representation of the input to our RFs. This insight is also sustained by reflecting on the richness of the image properties contained in V1. The hypothetical stereoscopic RFs that we have advocated require such input to sense the breakdown in correspondence that occurs at a very early stage in binocular processing. However, to fully support a claim that such RFs exist, physiological evidence would be required. At present, no such evidence exists. The value of the proposed mechanisms is that they provide heuristic insight into how the computation of stereoscopic occlusion relationships may be performed at a very early stage of stereoscopic processing. Our hope is that our formulation of these RFs is sufficiently interesting to motivate physiological experiments to study the nature of the mechanisms supporting these percepts.

As a final note, the RF profiles we have suggested may also begin to provide an account of how half-occluded features can generate illusory contours, a phenomenon that has received no explanation to date. These RFs would be weakly stimulated by placing monocular features from the occlusion correct eye in the monocular lobe and nothing in the binocular lobe. The illusory contours form either to the left or to the right of an unpaired feature, depending on whether it is a right-eye-only or left-eye-only feature (respectively). This property is captured by our restriction of our RF profiles to only two types. Additionally, one of us (Anderson, 1994) has recently generated displays demonstrating that these illusory contours can appear not just in depth but can have a definite orientation that only exists in the cyclopean image. In part, this property is retained by our use of oriented RFs for the detection of the half-occlusions. These results will be described in greater detail in a subsequent article.

One shortcoming of the RF models proposed here pertains to the perceived depth of the half-occlusions. In its current form, our RFs only predict the sign of the depth bias correctly; it does not assign a specific depth to the half-occlusions. There is currently very little data that allow us to examine this question in any detail. Presumably, the fact that half-occlusions appear at the depth of the occluded surface suggests that the disparities of neighboring features somehow spread into the half-occluded regions. Little is currently known about these interactions and remains as an important unsolved problem of stereopsis. We return to this problem in our discussion of the impact of half-occlusions on theories of binocular fusion.

Whether or not the RF structures we have proposed actually exist is unclear. What does seem clear is that there do exist mechanisms that exploit the structure of occlusion configura-

tions as a source of information during the earliest stages of binocular processing. In this regard, we view our hypothetical RFs as a novel effort to suggest some "front end" mechanisms that could use the information generated at occluding contours rather than treating this breakdown in disparity as a source of computational difficulty.

Some purely formal models have recently emerged that recognize the need to handle half-occlusions explicitly at the earliest stages of processing. The majority of our article has focused on the shortcomings of stereo models that fail to incorporate the information provided by occlusion configurations, so it seems useful to consider how our insights may be expressed in a formal mathematical context. The purpose of this section is twofold: First, it provides some examples of models that recognize the relevance of including information about occlusion at the initial stages of stereoscopic processing, and second, in contrast to the mechanistic RF model that we have forwarded, these models are more general because they focus on the problem of occlusion at a more abstract level of analysis, namely, what Marr (1982) called the computational level. However, while the probabilistic (Bayesian) framework used by these models may be unfamiliar to many readers, we emphasize that the way these models integrate the structure contained in occlusion configurations with disparity detection is similar to our integration of a multiple-channel disparity representation with RFs structured to detect occlusion relationships. To see how, we now turn to a brief discussion of these models.

## Algorithmic Models Exploiting Half-Occlusions

The preceding sections of this article have focused on the shortcomings of various models of stereopsis in accounting for the early processing of disparity discontinuities, especially those that contain half-occlusions. Recently, some computational models have been developed that have had some success in incorporating the structure present in occlusion configurations into the earliest phase of stereoscopic processing (Belhumeur & Mumford, 1992; Geiger et al., 1993). These models exploit Bayes's theorem to reconstruct a 3D surface layout. The general approach is to treat the binocular images as "data," and, on the basis of certain prior expectations, to reconstruct a 3D surface layout that provides the best account of the data. The Bayesian approach has become very widespread within the artificial intelligence community and has been offered as a general framework in which different models may be compared (Yuille et al., 1991). It is therefore instructive to review the common elements of these models here. Bayes's theorem may be written:

$$P(D(x)|I_l, I_r) = \frac{P(I_l, I_r | D(x))P(D(x))}{P(I_l, I_r)} \qquad (1)$$

where $P(I_l, I_r | D(x))$ is known as the data term, a measure of how well the distance function $D(x)$ (the distance from the vantage point to a point in space) agrees with the data. One of the major differences between models of stereopsis is the way in which the left and right image functions $I_l$ and $I_r$ (the image data) are characterized. This was discussed in the introduction of this article, where we reviewed the different matching primitives that have been used as the input to stereopsis. The second term, $P(D(x))$, is known as the prior and is a measure of the a priori likelihood

of $D(x)$. This term describes the types of surface biases that are built into the network, the most familiar of which is a smoothness constraint. However, any number of such constraints may be captured in this term, and we shall see that it is here that occlusion may be incorporated into the computational problem of stereopsis. The denominator $P(I_l, I_r)$ is simply a normalizing constant. $P(D(x)|I_l, I_r)$ is therefore the posterior probability of $D(x)$ given the image data $I_l, I_r$.

The major differences between models of stereopsis arise in two ways: through different characterizations of what is considered as the image data $I_l$ and $I_r$ and the form of the prior probability term $P(D(x))$. The most successful Bayesian models have recognized the inadequacy of grey-scale correlation and have used some form of preprocessing (or at least some "windows" of pixels) as the input for their models. Here, we focus on how these models have incorporated occlusion relationships into the prior term $P(D)$. In the Bayesian approach, the computational problem of stereopsis is usually transformed into a cost function (or energy) that is to be minimized, which may then be given a probabilistic interpretation. No generality is lost in this transformation and is done primarily so that various methods of minimization may be applied to these functions (e.g., dynamic programming). Minimizing the cost function is tantamount to finding the maximum a posteriori estimator of $P(D)$. The basic idea is to treat this minimization as a "relaxation" onto a 3D representation that has the lowest energy, that is, the 3D representation that may be interpreted as the most probable cause of the half-images.

To understand the role of the prior term $P(D)$, let us start by considering models biased to detect smooth surfaces (which was also the type of model that emerged first historically). This term may assume a number of forms. One common form for this term uses an approximation to a membrane surface using an operator of the form $\delta/\delta x$, which may be written in discrete form $\Sigma_k (d_k - d_{k+1})^2$ (see Yuille et al., 1991). Note that this, operator is defined on the equipolar lines, allowing for the problem of stereopsis to be treated one dimensionally. The operator $\delta/\delta x$ is less restrictive than models that simply reinforced same-disparity neighbors (Marr & Poggio, 1976; Julesz & Chang, 1976), which created a bias to reconstruct surfaces that were strictly fronto-parallel (or more accurately, surfaces that fell on iso-disparity contours). Not surprisingly, when smoothness is the only contribution to the prior term $P(D)$, then the model will perform very poorly at depth discontinuities. Thus, a new constraint is needed, one that allows for breaks in the smoothness term at discontinuities but, at the same time, preserves the putative benefits of smoothness within regions bounded by discontinuities. This is typically achieved by allowing for regions (contours) that are removed from the smoothness process. However, a cost or penalty must be imposed for categorizing a region as a discontinuity. Otherwise, there would be nothing to prevent all regions of the image from being treated as discontinuities, and any possible benefits of smoothing would be lost. The overall cost function then becomes a combination of smoothness constraints and discontinuity fields.

This general strategy of decomposing the prior term, $P(D)$, into a number of different contributions can be extended to include half-occluded features generated at occluding contours. As we have mentioned, two types of occlusions may be distin-

guished. If we traverse a surface from left to right, a surface may discontinuously end by becoming an occluding contour. The second case involves a surface that discontinuously ends by becoming occluded. These two types of occlusion are expected to occur with equal probability. A number of forms for this probability may be given, and the interested reader should consult Belhumeur and Mumford (1992) and Geiger et al. (1993) for some examples of this derivation.

The important aspect of this computational theory is that all terms simultaneously contribute to the overall energy function that is minimized (or equivalently, the probability that is maximized). There is nothing in these algorithms that gives precedence to features that can be matched over half-occluded regions, in general agreement with the experiments we have reported in this article. Furthermore, the need to explicitly distinguish the possibility of smooth regions from regions containing discontinuities is similar to our proposal that there exist specialized RFs that underlie the detection of depth edges. In a sense, the fact that each form of occlusion must be represented by a distinct prior is equivalent to assuming that distinct "mechanisms" detect these regions.

One common objection that is often raised about the application of Bayesian methods to biological systems is their explicit use of prior probabilities. The problem is one of understanding how this prior "knowledge" can be inserted into the neural machinery so that it may then perform the necessary computations. To a large extent, this objection simply reflects semantic prejudices, as many models that are not explicitly Bayesian invoke some set of priors to constrain their search for the best matching solution. The only difference is that in Bayesian models, these constraints are explicitly called priors. For example, the suggestion that disparity-sensitive cells spread excitatory signals between neighboring cells tuned to the same disparity invokes a bias for detecting regions of constant disparity. This bias may be thought of as a form of prior knowledge about the kind of surfaces "expected" by the visual system. Indeed, such models have been shown to be formally equivalent to some simpler variants of extant Bayesian models (see Yuille et al., 1991). The advantage of an explicit construction of priors is that it becomes clear just what assumptions are being imposed by the model on the types of solutions that are sought.

For our purposes, the essential element of the Bayesian approaches described here is that the goal of the computation is to generate a 3D representation of surfaces as "solutions." From this perspective, disparity plays no privileged role but is just one type of information generated in the projection of a 3D scene to our two eyes. The specification of a general energy function motivated by the consequences of projecting a 3D layout of surfaces onto our two eyes shifts the emphasis from disparity detection to surface reconstruction. Indeed, the recognition that many regions of the images will actually represent occlusions implies that disparity computations will not capture a large number of regions of the 3D scene. This necessitates a theory of stereopsis that explicitly recognizes the incompleteness of disparity computations for representing stereoscopic depth. This is the main point we have tried to develop in this article to this point. Indeed, to a large extent, our suggestion that RFs may exist for detecting occluding contours may be construed as a

mechanistic implementation of the Bayesian priors that explicitly acknowledge the possibility of their occurrence.

## But What About Cooperativity?

Our discussion of the problems of stereopsis have focused on two properties: (a) the relationship between the primitives used for matching and the false target problem and (b) the need to develop computational strategies for processing occlusion configurations. We have noted that cooperativity was typically invoked to solve the false target problem. However, by using complex matching primitives, it seems that such interactions may be unnecessary. Indeed, in Section 1 of this article, we suggested that the false target problem is most likely a false problem. However, this thesis was forwarded on the basis of theoretical arguments; we did not consider the data that have been interpreted as evidence for cooperativity. Lest the reader feel that we have ignored data that support the cooperative thesis, we review these data here.

The two lines of evidence that have been used to uphold the cooperative thesis are the phenomena of fusional hysteresis (Diner, 1978; Diner & Fender, 1987; Erkelens, 1988; Hyson, Julesz, & Fender, 1983; Piantanida, 1986) and the pulling effect (Julesz & Chang, 1976; cf. Marr, 1982). Fusional hysteresis has been interpreted as evidence that cooperative interactions maintain the stability of a fused stereopair. The pulling effect has been interpreted as providing evidence that cooperative interactions are needed to explain how correspondence is achieved. In the following sections, we develop alternative explanations of these phenomena that do not entail the use of cooperative interactions. We then briefly describe some phenomena that do seem to provide evidence for the existence of cooperative interactions. However, the phenomena that seem to reflect cooperative interactions suggest that the role of cooperativity is markedly different than previously believed.

*Hysteresis.* One of the earliest experiments that seemed to indicate the existence of cooperative interactions was fusional hysteresis, initially reported by Fender and Julesz (1967). In general terms, hysteresis is defined as a lag between a cause and an effect. Fusional hysteresis refers to the fact that the diplopia threshold for an unfused stereopair that is slowly brought into retinal correspondence is substantially smaller than the diplopia threshold of an initially fused stereopair that is slowly pulled apart. This result has been replicated numerous times (Diner, 1978; Diner & Fender, 1987; Hyson et al., 1983; Piantanida, 1986) and, until very recently (Erkelens, 1988), was interpreted as evidence that there were neural interactions that preserved the state of fusion by some form of cooperative "locking" or a cortical shifting of fusional zones. If it could be demonstrated that the fusional zone had been enlarged once fusion was obtained, then this result would indeed provide evidence for the existence of cooperative mechanisms in fusion. However, all of these studies lacked a crucial control experiment: assessing the static fusional limit for the stereopairs used in the temporally changing displays. The previous studies simply assumed that this static value corresponded to the value established by Panum (1858/1940), that is, approximately 6 arc min. This control experiment was performed by Erkelens (1988) for large RDSs, as well as the two dynamic conditions that had been eval-

uated in previous studies. The surprising result of this study was that the fusional limit for an RDS presented without a previous history of fusion or diplopia was larger than the target that was initially fused and slowly pulled apart! This implies that the hysteresis effects previously observed were not due to a locking mechanism that preserved the state of fusion but, rather, revealed that fusional hysteresis was due to a difficulty in obtaining fusion when preceded by a state of diplopia. Erkelens argued that the state of diplopia initiated binocular rivalry and that the interocular suppression characteristic of rivalry interfered with the process of fusion. Thus, hysteresis was not attributed to cooperative interactions that preserved the state of fusion but, rather, to interocular suppression that made fusion more difficult to obtain.

A similar argument was suggested by Blake (1989) to account for studies performed by Julesz and Tyler (1976) and Tyler and Julesz (1976). In these studies, Julesz and Tyler found that it was much easier for observers to detect transitions from a correlated RDS to an uncorrelated RDS than vice versa. Julesz and Tyler called this phenomenon neurontropy, in recognition of the fact that this result seemed to be a neural analogy of entropy. The cooperative reinforcement of local matches would predict the opposite result: Because the correlated RDS would have a lot of neighborhood support, it should be difficult to "break" this state once it was formed. Yet this was not what was observed. Blake suggested that the difficulty in detecting changes from decorrelation to correlation could be explained by assuming that the decorrelated state generated interocular suppression. Again, the matching process was more seriously effected by the putative suppression generated by decorrelation than by hypothetical cooperative interactions that act to maintain the organization of the correlated RDS.

More recently, one of us (Anderson, 1992) performed similar neurontropy experiments with dynamic displays depicting transparent surfaces or volumes. Recent models that exploit neighborhood support in the disparity domain do so only for features that can be matched, not for decorrelated stimuli, so it may be argued that the Julesz and Tyler (1976) and Tyler and Julesz (1976) studies did not provide a fair test of cooperative interactions. Using sparse RDSs, Anderson presented targets in which all features could be assigned matches. If local matches were reinforced through interactions with their neighbors, and this support was strongest for similarly tuned mechanisms, then the pattern with two disparity values should have been more stable than disparity volumes. This was not observed; in fact, the opposite result was obtained. As with the interpretation of fusional hysteresis forwarded by Erkelens (1988) and the neurontropy studies of Julesz and Tyler, the results of Anderson's experiments do not support the thesis that matching is either achieved or maintained through the reinforcement of local matching successes.

*The pulling effect.* One of the strongest psychophysical results used to sustain the cooperative thesis was described by Julesz and Chang (1976). Julesz and Chang studied the impact of placing unambiguous "bias" dots in a bistable, random-dot wallpaper pattern flanked by a zero-disparity random-dot background (a method exploited in the experiments of this article). In its unbiased form, observers perceived the wallpaper either in front of or behind the background, but exhibited some ten-

dency to see one organization more frequently than the other. Julesz and Chang studied the effects of introducing unambiguous bias dots to the depth plane opposite the observer's natural bias. They found that as little as 2% bias will "pull" the wallpaper pattern to the depth of the unambiguous points. They argued that the remarkable aspect of this result was that the visual system "chose" between two organizations that differed by only a very small percentage of matches. Julesz and Chang called this the *pulling effect*, suggesting that this choice was made by a facilitative spread of information between disparity-sensitive units tuned to the disparity of the unambiguous points (i.e., a smoothness constraint).

Note, however, that this interpretation assumes that the input to stereopsis is individual dots. If more complex primitives are used during binocular matching, then there is no need to describe this finding as a pulling effect. Note that this description assumes that it is the interactions of disparity units sensitive to individual dots that pull the ambiguous features to the biased depth. An alternative explanation of this result is to assume that the matching primitives are so rich that the addition of unambiguous points renders the unbiased interpretation unstable. If the visual system is matching regions, and the size of the region or the density of bias dots is large enough, then the features that will be matched will contain a mixture of ambiguous and unambiguous dots. When considered as regions, no ambiguity exists: The presence of the unambiguous features breaks the symmetry of the two organizations, "forcing" matches to form at the depth of the bias dots. Thus, no cooperative interactions are necessary to explain the pulling effect. Indeed, the notion that the pulling effect reflects cooperative interactions is a classic example of the relationship between the nature of the primitives that input to stereopsis and the presumed ambiguity of matching.

## The Role of Cooperativity

The preceding discussion presents arguments that cooperativity is not needed to understand the processes involved in achieving or maintaining binocular correspondence. Is cooperativity just a popular myth of stereoscopic processing? This question is difficult to answer, especially because the cooperative construct has survived even in the absence of a clear definition. The core insight behind this idea seems to be the notion that neighboring detectors interact in such a way to generate global structures (Julesz, personal communication, 1993). Indeed, in physical systems, the term cooperativity has been used to describe systems that exhibit transitions from a locally disordered state to a coherent macroscopic pattern (e.g., Prigogine, 1980). However, this term has also been adopted by the neural network community with a different meaning. Here, cooperative interactions are contrasted with competitive interactions, corresponding to the neural properties of excitation and inhibition (respectively). Is there evidence for either meaning of cooperativity in stereopsis?

There seem to be at least two classes of phenomena that might entail cooperative interactions. One such phenomenon was crucial for all of the experiments we reported in this article. This is the fact that the bistable wallpaper patterns can only be seen in one organization at a time. Consider, for example, the

random-dot wallpaper pattern presented in Figure 3a. Why are the two percepts unique at a given point in time? Both depth organizations have equal matching energy, and there is no bias for either of our hypothetical RFs to be more active than the other. Observers can learn to flip between the two organizations, but they cannot be trained to see both patterns at the same time. Why? What prohibits the simultaneous percept of both depths?

Two types of explanation seem possible, and at least one involves interactions that could be properly termed cooperative. One explanation is that one depth inhibits the responses of the disparity-sensitive units along the same line of sight. Such inhibition was initially postulated to help resolve the false target problem and more recently was offered as an explanation of depth repulsions observed in dynamic RDS displays depicting transparent surfaces (Stevenson, Cormack, & Schor, 1991). A second possible explanation is that some form of uniqueness constraint prohibits (or at least reduces) the probability of multiple matches for a given feature. It is unclear whether cooperative interactions are necessary to implement uniqueness constraints, so we can only say that it is possible that the uniqueness of multi-stable patterns involves cooperative interactions.

Perhaps the most striking example of cooperative interactions is the enhancement of illusory contours that occurs when portions of the inducing elements are portrayed in stereoscopic depth (Anderson, 1994; Anderson & Nakayama, 1992; Lawson & Gulick, 1967; Nakayama, Shimojo, & Silverman, 1989; Ramachandran & Cavanagh, 1985). The tendency for contours to perceptually "complete" cannot be understood simply on the basis of the disparities present in the images. It would be somewhat ironic if the most compelling examples of cooperativity arise in stereograms that are relatively sparse in detail, rather than the rich textures provided in RDSs (where the concept was originally applied). Rather than using cooperative interactions to solve the false target dilemma, cooperativity may be most useful when there are very few targets in the image. Perhaps this should not be surprising. In an RDS, the rich, local structure allows the disparity of each region to be highly specified, at least if the matching primitives are assumed to be more than individual pixel intensities. However, with sparse targets, phenomena such as illusory contours play a more significant role in image segmentation. It is here that cooperative interactions may be most useful. Indeed, cooperativity of this type has been used in the models of Grossberg (1987) and Grossberg and Mingolla (1985a, 1985b). However, as with fusional hysteresis, illusory contours do not appear to exhibit hysteretic stability, as these models originally suggested (Anderson, 1990). Thus, the role of cooperativity in visual processing has yet to be fully understood and remains to receive sufficient theoretical and empirical evaluation.

## Section 4: Implications for Theories of Binocular Combination: Fusion Versus Suppression

Our preceding discussion of stereopsis has been couched as an abstract process of establishing matches for mutually visible binocular features, or "detecting" or "sensing" occlusion configurations. One fundamental problem that has been almost completely ignored since the emergence of computational models is the relationship between binocular correspondence and binocular fusion (but see Grossberg, 1987; Sperling, 1970). Historically, the process of fusion has been dissociated from stereopsis. Binocular fusion has come to be identified as single vision, which only occurs for a very restricted range of disparity values. In contrast, stereopsis occurs over a much broader range of disparities, well into regions dominated by percepts of diplopia (double vision). Thus, there is a range of disparities for which relative depth can be reliably perceived, even though the images appear diplopic. If fusion is interpreted solely as a process that leads to binocular single vision, then it cannot be strictly tied to stereopsis, as stereopsis occurs for a much broader range of disparities than single vision. However, if every binocular feature is diplopic, then stereopsis fails. What, then, is the relationship between binocular fusion and stereopsis?

The most popular answer to this question is that stereopsis and fusion represent the operation of distinct mechanisms (see, e.g., Regan, Frisby, Poggio, Schor, & Tyler, 1990). Binocular fusion refers to the processes by which our perceptual world appears single, despite the fact that it is viewed by two eyes that receive disparate inputs. In contrast, stereopsis uses disparate inputs to recover depth. These definitions have forced the conclusion that stereopsis and fusion represent the operation of distinct mechanisms that respond quite differently to disparate inputs. Here, we briefly outline a perspective in which fusion and stereopsis may be seen as a common process.

Two theories of binocular single vision have been forwarded. Fusion theories claim that single vision arises from a deformation (or "fusion") of the two half-images into a common representation, with both eyes contributing to the localization of a given surface feature. This localization has two components: depth and visual direction. Suppression theories contend that we see a single world because at any given instant one eye is suppressed. Note that suppression theories say nothing about stereopsis, at least disparity-based stereopsis. By definition, disparity computations require that differences be formed between features in the two half-images; both eyes must contribute to this computation. Thus, the only phenomenon that suppression theories attempt to account for is binocular single vision and operates under a strained premise: namely, that we do see a single world.

A simple experiment demonstrates that we do not generally perceive a single world. Simply hold a finger up between yourself and this page, and fixate the text on the page (i.e., continue reading). Note that your finger appears double (and in depth!) at virtually every distance between you and the page. This is also true whether the finger is slightly off your line of sight or directly in front of you. Suppression theories have great difficulty accounting for this simple phenomenological fact.

More critically, suppression theories are incapable of accounting for the simultaneous stability of half-occlusions from both the left and right eyes. Consider, for example, any of the stereograms used in Experiments 1 and 2. When fused, both left-eye-only and right-eye-only half-occlusions coexist in the same percept. These regions do not undergo temporal fluctuations in visibility. Therefore, it is not possible to claim that their simultaneous visibility is due to the suppression of a given eye. Rather, both eyes are clearly contributing to a stable percept of a surface in depth. We believe that this should provide the last

argument needed to reject suppression theories of binocular fusion.

A counter to this argument would be to assert that interocular suppression acts locally, not globally over the entire eye. However, this would require the articulation of a set of complex rules that could predict the local patterns of suppression. No such rules currently exist.

It seems that we are left with some variant of a fusion theory. The problem, then, is to decide what we mean by fusion. Informally, vision scientists will speak of fusing stereograms in the sense of achieving stereoscopic depth. However, if pressed, most would acknowledge that the link between stereopsis and fusion is tenuous at best. For fusion and stereopsis to be seen as the same process, a number of phenomena must be explained. This includes an account of single vision (for the range over which single vision is observed), the coexistence of diplopia and stereoscopic depth (see, e.g., Kulikowski, 1978), and the stability of half-occlusions.

Implicit in the identification of fusion with single vision is the notion that the diplopia threshold represents the limit of a single mechanism, a mechanism that can only exist in one of two states: fused or nonfused. However, consider what fusion would mean if disparity is represented by a set of "channels" tuned to different magnitudes of disparity. In this case, it would not make sense to speak of a single fusional limit. Rather, each disparity-sensitive scale (or channel) would have a peak response for a disparity of a given magnitude. Units sensitive to smaller disparities would have smaller diplopia thresholds, such that relatively small values of disparity would not be fusible and therefore generate diplopia. Units tuned to larger disparities would require larger disparity values before diplopia would be perceived. A model of this kind has been suggested previously to account for disparity increment thresholds (McKee, Levi, & Bowne, 1990), and we used this model in the previous section to provide an account of binocular matching. In a representational medium of this kind, depth with diplopia emerges from the fact that when the disparity of a feature is large, scales tuned to smaller disparities will not be able to fuse this disparity, but the larger scales will, at least up to some limit. This selective response will generate an activity peak in the distribution of disparity-tuned mechanisms at the scale optimally sensitive to the projected disparity, coding the relative depth of the specified image feature. However, at the same time, the small scales will not be able to fuse such a large disparity, generating a percept of diplopia.

It is important to underscore the fact that we are emphasizing a link between the size of the RFs of disparity-sensitive units and their respective diplopia thresholds. This may or may not be related to the spatial scale of the half-image features that input to disparity-sensitive units (Schor & Badcock, 1985; Schor & Wood, 1983; Yang & Blake, 1991). Such questions are beyond the scope of this article.

The important theoretical consequence of linking a multiple-scale representation of disparity with binocular fusion is the conclusion that diplopia is actually a poor index of fusion. It is possible to conceive of fusion occurring in a disparity-specific manner. From this perspective, there would be many diplopia thresholds physiologically, even though the psychological decision is limited to categorizing a stimulus as fused or nonfused.

Our arguments may be supported by considering what this multiple-scale representation would predict about the perceived geometry of patterns that were fused. There are two geometric consequences of binocular fusion. First, disparity and occlusions generate percepts of relative depth (i.e., the two half-images are transformed into relative depth signals). Second, the half-images that are projected on the retina are deformed in the X-Y plane, a phenomenon referred to as *allelotropia*, or *fusional displacement*. In a multiple-channel disparity representation, these two transformations are not independent. Rather, they both contribute to a 3D representation of binocular space, implicit in the projective representation originally described by Kepler. This can be seen by considering the Keplerian diagram depicted in Figures 2–4. Note that the two transformations of the half-images we have mentioned, depth and allelotropia, are both implied in this representation (assuming the existence of a cyclopean "eye" situated between the two eyes). The significance of this representation for depicting the relative depth of objects has always been appreciated. However, the transformation of allelotropia is also present, although this fact is typically overlooked.

If allelotropia and stereopsis are both created by embedding the two half-images in a multiple-scale representation of disparity, and this embedding is what is meant by fusion, then the range of disparities over which stereopsis and allelotropia occur should be the same. This follows directly from the assumption that these two transformations, allelotropia and depth ("stereopsis"), reflect a common representational scheme. Furthermore, because stereopsis can be achieved even for diplopic targets, allelotropia should also occur for images that appear diplopic but in depth, and allelotropia should cease to exist at the same point that stereopsis fails. From this perspective, allelotropia should occur up to the point where at least the largest disparity scales can fuse a visual feature. Recently, Rose and Blake (1988) discovered just this: Allelotropia was observed even when the images appeared diplopic but failed to occur for disparities that did not also support stereoscopic depth perception. Our interpretation of these results differs from theirs, however. Because they identified fusion with binocular single vision, Rose and Blake were led to interpret their results as evidence for two mechanisms of binocular combination: one responsible for binocular single vision and the other for stereoscopic depth perception. We are arguing that this dichotomy is unnecessary. By identifying fusion with a deformation of the half-images into a projective space, the phenomenon of allelotropia, together with a percept of relative depth, becomes a "signature" of binocular fusion. If the disparities are small, then no diplopia is experienced; but if the disparities are sufficiently far from the horopter (but still fusible), then diplopia, allelotropia, and depth will be experienced.

## Implications of Allelotropia and Depth Transformations for the Perceptual Stability and Perceived Depth of Half-Occlusions

The identification of fusion with transformations of allelotropia and depth only provides an account of features for which disparity can be computed. We have suggested that binocular fusion should not be identified with single vision but, rather, to

the process of embedding the half-images into a multiple-scale disparity representation, which also applies to half-occlusions. To respect our own emphasis on the importance of occlusion in stereopsis, we need to generalize the definition of fusion to account for the perceptual stability of half-occluded features.

Our analysis of this problem consists of two parts. One part is a geometric description of the problem; the other relates our hypothetical RF models for detecting occluding contours to this geometric description. The great puzzle of half-occlusions with respect to fusion is understanding why they do not initiate binocular rivalry. Rivalry has usually been addressed two dimensionally, in the sense that the conditions thought to generate rivalry were related to the structure of the half-images (Blake, 1989). In accordance with our multiple-channel model described above, we suggest that the stability of half-occlusions can only be understood by considering the conditions for rivalry as occurring in a 3D representation. From this perspective, half-occlusions are perceptually stable because the transformations that arise when fusing the half-images shift the retinal regions that are suppressed. Consider, for example, the simple stereo-pair illustrated in Figure 1. How do the half-occlusions escape interocular suppression? If we just consider the half-images, it is difficult to understand this fact. There are no matches for these features, and the corresponding regions in the complementary eye are occupied by features that are matched with features on disparate regions in the other eye. However, consider the effects that the two transformations (allelotropia and depth) would have on the half-images. Allelotropia would displace the central disparate zone to an intermediate position between the half-occlusions. This would create a vacant "space" in the images in which the half-occlusions may be placed. If this space was exactly the width of the half-occluded regions, then this could provide the explanation of the nonrivalry of these regions: The conditions for rivalry would be determined by the positions of the displaced features in the $X$-$Y$ plane after fusion has occurred (i.e., the transformation induced through allelotropia), not by the (nontransformed) locations of features on the retinas. This argument implies two facts: First, that allelotropia remaps the zones that would be subjected to suppression on the two retinas, and second, that the conditions for rivalry may still be expressed in terms of the positions of features on two-dimensional retinas. However, the astute reader may have noticed that although this argument has some qualitative appeal, the space created by allelotropia is only half of that needed to create a space equal to the angular width of the half-occluded zones. This is because allelotropia essentially averages the two visual directions, dividing the displacement of matched features within a given eye in half. The width of the half-occlusions, however, is equal to the magnitude of the disparity, not half the disparity. Therefore, the shift in the $X$-$Y$ plane induced by allelotropia cannot create a space large enough in which to place the half-occluded features, and therefore cannot by itself provide a complete account of their stability. But allelotropia is only one of the transformations that accompanies binocular fusion. The other transformation is depth.

It is a relatively simple matter to show that when the transformations of allelotropia and depth are combined, that the angular space created alongside an occluding contour is exactly equal to the angular size of the half-occluded features. This is the geo-

metric fact from which the problem of half-occlusions arises. Although this may appear as a trivial restatement of the problem, our point is that the conditions for binocular rivalry are determined by the geometric transformations present in the 3D representation, not the two-dimensional half-images.

Even this, however, is still not sufficient to understand how half-occlusions escape binocular rivalry. The multiple-scale disparity representation is, after all, just a representation of disparity. Some mechanisms must exist to detect half-occlusions. In Section 3 of this article, we suggested that half-occlusions are detected by asymmetric RFs that sense a correlation–decorrelation boundary. We argued that only two general classes of RFs were needed, corresponding to the configurations that represented occluding contours: left-eye-only features to the left of an occluder and right-eye-only features to the right of an occluder. In part, this restriction of our RFs to just two types was inspired by some recent experiments on binocular rivalry. Shimojo and Nakayama (1990) recently demonstrated that the monocular features in these configurations did not generate binocular rivalry. However, when the half-occlusions could not be organized as part of an occluded surface, binocular rivalry ensued. Yet by itself, this mechanistic explanation also seems incomplete. Without the transformations of allelotropia and depth to shift the locus of the matchable features, there would still be the problem of where to put the half-occlusions in the 3D representation. In other words, our hypothesis that there must be mechanisms sensitive to the structure of half-occlusions does not constitute a theory of how these features are placed in the 3D representation that emerges, it only postulates the existence of a mechanism that responds to their occurrence. To answer this question, we need to recognize the effects of the transformations that accompany stereopsis, namely, depth and allelotropia. When we do, we see that the half-occlusions are situated in the 3D gaps that are created by the transformations that accompany fusion. By integrating the mechanisms that detect disparity and occlusion configurations with a theory of binocular combination that recognizes the geometric transformations that accompany stereopsis, we have at least the beginnings of a theory on how stereopsis and binocular fusion may be construed as a single process.

## Conclusions

In this article, we have presented data demonstrating that occlusion configurations may influence the earliest stages of stereoscopic matching. We have argued that this indicates that stereopsis must exploit relatively complex matching primitives to achieve a rapid classification of features as matchable or half-occluded. Stereopsis is more than just binocular matching; surface properties such as occlusion relationships also contribute significant information to the 3D representation generated by binocular viewing. Models of disparity detection perform extremely poorly at occluding contours, suggesting that some other process must be responsible for representing these regions. We have constructed some hypothetical RF models that could be used to detect stereo occlusions and noted the relationship of these RFs to recent Bayesian models of stereopsis. It remains to be seen whether such units will be discovered by neurophysiologists. Finally, we noted that the simultaneous inte-

gration of half-occlusions from both eyes provides a strong counter argument against suppression theories of binocular single vision. This led us to reconsider the relationship between binocular fusion and binocular matching. We argued that fusion and stereopsis may be understood to be the same process if fusion is defined as transformations that embed the half-images in a 3D representation, rather than as single vision. This framework can explain the coexistence of stereopsis and diplopia without assuming these phenomena are handled by distinct mechanisms. We hope that our attempts to unify the stereoscopic phenomena described herein will motivate psychophysical and physiological experiments to discover both the merits and the shortcomings of these ideas.

## References

Anderson, B. L. (1990). Hysteresis, cooperativity, and scale-space models of stereopsis. Unpublished doctoral dissertation, Vanderbilt University, Nashville, TN.

Anderson, B. L. (1992). Hysteresis, cooperativity, and depth averaging in dynamic random-dot stereograms. Perception & Psychophysics, 51, 511–528.

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

Anderson, B. L., & Nakayama, K. (1992). Binocular versus monocular junctions in 3D surface perception. Investigative Ophthalmology and Visual Science Supplement, 33(Suppl. 4), 3197.

Belhumeur, P. N., & Mumford, D. (1992). A Bayesian treatment of the stereo correspondence problem using half-occluded regions. In Proceedings of the Institute of Electrical and Electronics Engineers Conference on Computer Vision and Pattern Recognition (pp. 506–512). Los Alamitos, CA: IEEE.

Bulthoff, H. H., & Mallot, H. A. (1988). Interaction of depth modules: Stereo and shading. Journal of the Optical Society of America, 5, 1749–1758.

Blake, R. (1989). A neural theory of binocular rivalry. Psychological Review, 96, 145–167.

Brewster, D. (1844). On the law of visible position in single and binocular vision and on the representation of solid figures by the union of dissimilar plane pictures on the retina. Transactions of the Royal Society of Edinburgh, 15, 349–368.

Cormack, L. K., Stevenson, S. B., & Schor, C. M. (1993). Disparity tuned channels of the human visual system. Visual Neuroscience, 10, 585–596.

Dev, P. (1975). Perception of depth surfaces in random-dot stereograms: A neural model. International Journal of Man–Machine Studies, 7, 511–528.

Diner, D. B. (1978). Hysteresis in human binocular fusion: A second look. Unpublished doctoral dissertation, California Institute of Technology, Pasadena.

Diner, D. B., & Fender, D. H. (1987). Hysteresis in human binocular fusion: Temporalward and nasalward ranges. Journal of the Optical Society of America, 4, 1814–1819.

Erkelens, C. (1988). Fusional limits for a larger random-dot stereogram. Vision Research, 28, 345–353.

Fender, D., & Julesz, B. (1967). Extension of Panum's fusional area in binocularly stabilized vision. Journal of the Optical Society of America, 57, 819–830.

Geiger, D., Ladendorf, B., & Yuille, A. L. (1993). Occlusions and binocular stereo (Tech. Rep. No. 93–1). Harvard Robotics Laboratory, Harvard University, Cambridge, MA.

Gettys, C. F., & Harker, G. S. (1967). Some observations and measurements of the Panum phenomenon. Perception & Psychophysics, 2, 387–395.

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

Grossberg, S. (1987). Cortical dynamics of three-dimensional form, color, and brightness perception: II. Binocular theory. Perception & Psychophysics, 41, 117–158.

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

Grossberg, S., & Mingolla, E. (1985b). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. Perception & Psychophysics, 38, 141–171.

Gruen, A. W. (1985). Adaptive least squares correlation: A powerful image matching technique. South African Journal of Photogrammetry, Remote Sensing and Cartography, 3(14), 175–187.

Halpern, D. L., & Blake, R. R. (1988). How contrast effects stereoacuity. Perception, 17, 483–495.

Helmholtz, H. V. (1925). Treatise on physiological optics (J. P. Southall, Trans.). New York: Dover. (Original work published 1910)

Howard, I. P., & Ohmi, M. (1992). A new interpretation of the role of dichoptic occlusion in stereopsis. Investigative Ophthalmology and Visual Science Supplement, 33(Suppl. 4), 3382.

Hyson, M., Julesz, B., & Fender, D. H. (1983). Eye movements and neural remapping during fusion of misaligned random-dot stereograms. Journal of the Optical Society of America, 73, 1665–1673.

Jones, J., & Malik, J. (1992a). A computational framework for determining stereo correspondence from a set of linear spatial filters (Tech. Rep. No. UCB/CSD 91/655). Berkeley: University of California, Computer Science Division (EECS).

Jones, J., & Malik, J. (1992b). Determining three-dimensional shape from orientation and spatial frequency disparities. Proceedings ECCV, Santa Margherita Ligure, Italy.

Julesz, B. (1960). Binocular depth perception of computer generated patterns. Bell System Technical Journal, 39, 1125–1162.

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

Julesz, B., & Chang, J. (1976). Interaction between pools of binocular disparity detectors tuned to different disparities. Biological Cybernetics, 22, 107–119.

Julesz, B., & Tyler, C. (1976). Neurontropy, an entropy-like measure of neural correlation, in binocular fusion and rivalry. Biological Cybernetics, 23, 25–32.

Kanade, T., & Okutomi, M. (1990). A stereo matching algorithm with an adaptive window: Theory and experiments. In Proceedings of the Image Understanding Workshop. Washington, DC: DARPA, Science Application, Inc.

Kulikowski, J. J. (1978). Limit of single vision in stereopsis depends on contour sharpness. Nature, 275, 126–127.

Lawson, R. B., & Gulick, W. L. (1967). Stereopsis and anomalous contour. Vision Research, 7, 271–297.

Legge, G. E., & Gu, Y. (1989). Stereopsis and contrast. Vision Research, 29, 989–1004.

LeVay, S., & Voight, T. (1989). Ocular dominance and disparity coding in cat visual cortex. Visual Neuroscience, 1, 395–414.

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

Marr, D., & Poggio, T. (1976). Cooperative computation of stereo disparity. Science, 194, 283–287.

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. Proceedings of the Royal Society of London, 204(B), 301–328.

McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereo judgments. Vision Research, 30, 1763–1780.

Mowforth, P., Mayhew, J. E. W., & Frisby, J. P. (1981). Vergence eye movements made in response to spatial frequency filtered random-dot stereograms. Perception, 10, 299–304.

Nakayama, K., & Shimojo, S. (1990). DaVinci stereopsis: Depth and

subjective occluding contours from unpaired image points. *Vision Research, 30,* 1811–1825.

Nakayama, K., Shimojo, S., & Silverman, G. (1989). Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception, 18,* 55–68.

Nelson, J. I. (1975). Globality and stereoscopic fusion in binocular vision. *Journal of Theoretical Biology, 49,* 1–88.

Nishihara, H. K. (1987). Practical real time imaging stereo matcher. In M. A. Fischler & O. Firschein (Eds.), *Readings in computer vision* (pp. 63–72). Los Altos, CA: Kauffman.

Ono, H., Shimono, K., & Shibuta, K. (1992). Occlusion as a depth cue in the Wheatstone-Panum limiting case. *Perception & Psychophysics, 51,* 3–13.

Ono, H., & Wade, N. (1985). Resolving discrepant results of the Wheatstone experiment. *Psychological Research, 47,* 135–142.

Panum, P. L. (1940). *Physiological investigations concerning vision with two eyes* (C. Hubscher, Trans.). Hanover, NH: Dartmouth Eye Institute. (Original work published 1858)

Piantanida, T. P. (1986). Stereo hysteresis revisited. *Vision Research, 26,* 431–437.

Poggio, G. F., Gonzalez, F., & Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: Binocular correlation and disparity selectivity. *Journal of Neurophysiology, 40,* 1392–1405.

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

Pollard, S. B., Mayhew, J. E. W., & Frisby, J. P. (1985). A stereo correspondence algorithm using a disparity gradient limit. *Perception, 14,* 449–470.

Prazdny, K. (1985). Detection of binocular disparities. *Biological Cybernetics, 52,* 93–99.

Prigogine, I. (1980). *From being to becoming.* New York: Freeman.

Quam, L. H. (1987). Hierarchical warp stereo. In M. A. Fischler & O. Firschein (Eds.), *Readings in computer vision* (pp. 80–86). Los Altos, CA: Kauffman.

Ramachandran, V. S., & Cavanagh, P. (1985). Subjective contours capture stereopsis. *Nature, 317,* 527–530.

Regan, D., Frisby, J. P., Poggio, G. F., Schor, C. M., & Tyler, C. W. (1990). The perception of stereodepth and stereomotion: Cortical mechanisms. In L. Spillman & J. S. Werner (Eds.), *Visual perception: The neurophysiological foundations* (pp. 317–347). San Diego, CA: Academic Press.

Richards, W. (1970). Stereopsis and stereoblindness. *Experimental Brain Research, 10,* 380–388.

Richards, W. (1971). Anomalous stereoscopic depth perception. *Journal of the Optical Society of America, 61,* 410–414.

Richards, W. (1972). Disparity masking. *Vision Research, 12,* 1113–1124.

Rose, D., & Blake, R. (1988). Mislocalization of diplopic images. *Journal of the Optical Society of America, 5,* 512–521.

Schor, C. M., & Badcock, D. R. (1985). A comparison of stereo and vernier acuity within spatial channels as a function of distance from fixation. *Vision Research, 25,* 1113–1119.

Schor, C. M., & Wood, I. (1983). Disparity range for local stereopsis as a function of luminance spatial frequency. *Vision Research, 23,* 1649–1654.

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

Sperling, G. (1970). Binocular vision: A physiological and neural theory. *American Journal of Psychology, 83,* 461–534.

Stevenson, S. B., Cormack, L. K, & Schor, C. M. (1991). Depth attraction and repulsion in random dot stereograms. *Vision Research, 31,* 805–813.

Tyler, C. W., & Julesz, B. (1976). The neural transfer characteristic (neurontropy) for binocular stochastic stimulation. *Biological Cybernetics, 23,* 33–37.

Watt, R. J., & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research, 25,* 1661–1674.

Westheimer, G. (1986). Panum's phenomenon and the confluence of signals from the two eyes in stereoscopy. *Proceedings of the Royal Society of London, 228(B),* 289–305.

Wheatstone, C. (1838). Some remarkable phenomena of binocular vision. *Philosophical Transactions of the Royal Society, 128,* 371–394.

Wilson, H. R., Blake, R., & Halpern, D. L. (1991). Coarse spatial scales constrain the range of binocular fusion on fine scales. *Journal of the Optical Society of America, 8,* 229–236.

Yang, Y., & Blake, R. (1991). Spatial frequency tuning of human stereopsis. *Vision Research, 31,* 1177–1189.

Yuille, A. L., Geiger, D., & Bulthoff, H. H. (1991). Stereo integration, mean field theory and psychophysics. *Network, 2,* 423–442.