The Role of Occlusion in the Perception of Depth, Lightness, and Opacity

Barton L. Anderson
Massachusetts Institute of Technology

A theory is presented that explains how the visual system infers the lightness, opacity, and depth of surfaces from stereoscopic images. It is shown that the polarity and magnitude of image contrast play distinct roles in surface perception, which can be captured by 2 principles of perceptual inference. First, a contrast depth asymmetry principle articulates how the visual system computes the ordinal depth and lightness relationships from the polarity of local, binocularly matched image contrast. Second, a global transmittance anchoring principle expresses how variations in contrast magnitudes are used to infer the presence of transparent surfaces. It is argued that these principles provide a unified explanation of how the visual system computes the 3-D surface structure of opaque and transparent surfaces.

The primary function of visual processing is to recover the structure of the environment from the images on the retinas. One of the main difficulties encountered in solving this problem stems from the multiplicity of factors that act collectively to generate the images. Illumination sources can vary in intensity, position, number, and spectral content, and surfaces can vary in albedo, color, specularity, texture, depth, opacity, and orientation. This complex array of interacting physical factors act collectively to generate the 2-D images that fall on the retinas. However, despite this complexity, the visual system is remarkably adept at extracting a variety of properties of surfaces in the environment. The visual world appears as surfaces and objects extended in depth that possess properties such as color, lightness, shape, and opacity. A fundamental goal of mid-level vision is to understand how the visual system decomposes the 2-D image data into these different attributes and recovers underlying surface properties.

One of the most basic problems in mid-level vision is recovering the depth of surfaces and objects. It is now well-known that stereoscopic vision provides a powerful source of information about the 3-D structure of the world. One of the most important sources of stereoscopic depth information is provided by binocular disparity (or binocular parallax). Indeed, the vast majority of research on stereopsis has focused on understanding the mechanisms underlying the computation of disparity. These computations are thought to give rise to a disparity map, that is, a map of the disparities of surfaces features that are common in the two eyes. The underlying assumption guiding computational, physiological, and psychophysical investigations of stereoscopic vision has been that stereoscopic surface perception would be explained once it was understood how the visual system computed such disparity maps. The identification of stereoscopic surface structure with a disparity map was reinforced by the invention of the random-dot stereogram (RDS; Ashenbrenner, 1954) and the computer-generated RDS (Julesz, 1960). In these patterns, the relationship between the interocular positions of individual dots and perceived depth seemed straightforward. Once matched, the perception of stereoscopic depth in RDSs seemed to simply be determined by the point-by-point disparities present in these images. The main puzzle generated by RDSs was understanding how the visual system determined which features should be matched, not the relationship between disparity and perceived depth.

The conceptual focus on the correspondence problem makes a strong prediction about the relationship among matching, the disparity map, and the perception of stereoscopic depth. A solution to the correspondence problem requires that the visual system determine which features in the two eyes correspond to a common source in the world. However, because there may be more than one candidate match for a given image feature, the visual system must distinguish correct matches from incorrect matches (the false-target problem). Computational studies have revealed that the ability to determine whether a feature has a unique best match depends critically on the features that are used to solve correspondence, which in turn determines how reliant an algorithm is on assumptions about the properties of the objects that are recovered (such as the smoothness of a matching solution; see, e.g., Marr & Poggio, 1976, 1979; Pollard, Mayhew, & Frisby, 1985; Pradzny, 1985). Nonetheless, a common property of all extant stereo algorithms is that the solution to the correspondence problem is independent of which eye receives a given image. This is because the most similar features in the two eyes (or smoothness of a matching solution) is unchanged when the images in the two eyes are interchanged. Thus, if the two eyes’ views are interchanged, the pattern of matches should be identical; all that should occur is an inversion of the disparities of matched features that, in turn, should simply cause an inversion in perceived depth. This prediction is upheld by all RDSs, providing a critical link among the processes of matching, the disparity map, and the perception of stereoscopic depth.

However, a number of recent findings have revealed that the relationship between binocular disparity and stereoscopic depth may be more complicated than previously thought. It has been shown that dramatic changes in the perceived depth, lightness, and opacity of surfaces can be generated by simply interchanging the images in the two eyes. Consider the transformations in the perceived depth induced when the disparities of the stereo Kanizsa figure depicted in Figure 1 are inverted. When disparity is added...
A potentially related phenomenon was discovered by Takeichi, Watanabe, and Shimojo (1992; see Figure 2). In these stereograms, a few small dots were superimposed on Kanizsa triangles, and disparity was added to the dots in a manner that would place them either in front or behind the Kanizsa figure. When the dots were given a near disparity, they appeared to hover in front of the Kanizsa figure as three isolated points. However, when the disparities in these images were inverted (by interchanging the two eyes’ views), the entire white region within the triangle appeared at the (far) depth of the dots, and the illusory contours of the triangle were enhanced (appearing as a triangular hole). As with the stereo Kanizsa figure described above, the inversion of relative disparity generates a large asymmetry in the perceived surface structure in these images.

In addition to the geometric asymmetries in surface structure induced by inverting near and far disparities, some recent work from my lab has revealed both geometric and photometric asymmetries that arise when binocular images are interchanged (Anderson, 1999b). Consider the stereogram of the grating depicted in Figure 3. In these images, relative disparity is added by interocularly shifting the grating relative to the aperture boundaries. When the direction of this shift is consistent with the grating appearing behind the aperture, the grating is perceived as a coherent surface in a single depth plane visible through an aperture. In this configuration, the perception of depth is independent of the luminance relationships between the grating and surround (i.e., the same percept is obtained for all surround luminances; see Figure 3). However, when the disparities are inverted (by interchanging the two eyes’ views), a dramatically different set of percepts are experienced, which now depend critically on the luminance of the surround. When the surround is black, the grating appears to split into a set of fuzzy black bars that appear to occlude a white diamond lying on a black background. But when the surround is white, the grating appears to split into a set of fuzzy white bars that appear to occlude a black diamond lying on a white background. When the background is gray, the perception of the grating in the near configuration is unstable. Note that these figures contain both geometric and photometric asymmetries: The geometric asymmetry consists of the very different patterns of perceived depth in the near and far depth configuration (i.e., the grating appeared as either a single surface in one depth plane when it is placed behind the aperture boundaries or as two surfaces in two different depth planes when it is placed in front); the photometric asymmetry consists of the very different patterns of lightness that are assigned to the two depth planes when the grating appears to split into two surfaces (i.e., their dependence on the luminance of the surround). Similar phenomena can be observed with 2-D textures such as those depicted in Figure 4.

These phenomena demonstrate that there is something wrong (or at least incomplete) with the conventional view of stereopsis. The conventional view predicts that stereoscopic depth is simply determined by the relative disparity of image features, that is, on the pattern of disparities in a disparity map. There is nothing in this view that captures the asymmetries that arise from simply interchanging the signs of the disparities in the stereograms depicted in Figures 1–4. In this article, I argue that the geometric and photometric asymmetries that arise in these images reflect fundamental constraints on perceived surface structure that arise from the geometry of occlusion. The main insight developed in this article is that the geometry of occlusion imposes restrictions on the structure of local contrast signals that serve as the basis for the
visual system’s computations of surface structure. These restrictions play a critical role in how depth, lightness, and opacity are computed from relative disparity. In essence, then, this article attempts to articulate the constraints that occlusion places on the computation of surface structure from binocular disparity.

This article is divided into two parts. I begin by considering the problem of how depth is derived from local contrast signals and then turn to the problem of computing surface lightness and opacity. Although the theoretical arguments developed herein focus on the relationship between local image contrast and stereopsis, it should be noted at the outset that the theoretical principles described below are not restricted to stereoscopic information. Rather, the principles described in this article articulate general relationships between local image contrast and the computation of the depth, lightness, and opacity of surfaces; the role of stereopsis is to provide a means for systematically exploring the relationships between local, low-level computations of image contrast and mid-level computations that transform these signals into representations of surface structure.

Computing Depth From Binocular Image Contrast: The Contrast Depth Asymmetry Principle

To analyze how the visual system infers surface depth from binocular image features, one must begin by understanding what counts as a binocular feature; that is, what counts as the “stuff” that is assigned a disparity. Although there remains considerable uncertainty and debate about the dimensions of the monocular inputs to the matching process (Anderson, 1999a; Dev, 1975; Farell, 1998; Grossberg, 1987; Jones & Malik, 1992; Julesz, 1960, 1971; 1

1 Throughout this article, I use the term local image contrast to refer to any luminance gradient or discontinuity present in an image region.
Marr & Poggio, 1976, 1979; Pollard et al., 1985; Pradzny, 1985; Sperling, 1970), there is psychophysical evidence that local image contrast is the critical image dimension used to compute binocular correspondence (Anderson & Nakayama, 1994; Smallman & McKee, 1995). Indeed, given the nature of early visual processing, it seems almost inevitable that the correspondence process uses image contrast (i.e., local structure) to solve the correspondence problem. Both ganglion and lateral geniculate nucleus cells exhibit antagonistic center-surround organization that highlights information about local image contrast. Because binocular interactions and disparity sensitivity do not occur prior to Area V1, computations of correspondence must rely on measures of contrast at least to the extent that the monocular inputs to stereopsis code for contrast. Thus, the “what” of matching seems to involve some measure of local contrast, a fact that plays a critical role in the theoretical discussion that follows.

The stimulus dimensions used to define correspondence determine not only what is matched but also what kind of image data are assigned a disparity that results from the matching process. If raw pixel intensities are used to establish correspondence, then each individual pixel is assigned a disparity value (if a match for that pixel is found). However, if some local measure of image contrast is assigned a disparity—indeed, if any structured image primitive is used to determine matching—then the relationship

Figure 3. Example stereograms (A) that generate dramatic geometric and photometric transformations (B) when depth is inverted. Identical sinusoidal luminance profiles were viewed through diamond-shaped apertures. Disparity was introduced by shifting the aperture boundaries relative to the gratings. In the top and middle rows of A, when the right two images are cross-fused (or the left two images are fused divergently), the grating appears on a distant surface, visible through a diamond aperture (as depicted in the top row of B; the top left image corresponds to the top row of A, and the top right image corresponds to the middle row of A). However, when disparity relationships are reversed (by cross-fusing the left two images or divergently fusing the right two images), the grating appears to split into two depths. When the surround is black (top row of A), the grating appears as a uniform white diamond visible through hazy black stripes (bottom left image of B). When this occurs, the luminance maxima within the grating appear at the more distant depth layer as part of a uniformly colored white diamond, whereas the minima appear as hazy black stripes in front of the white diamond. When the same grating pattern is viewed on a white background (middle row of A), an entirely different percept emerges. The distant layer within the grating now appears as a black diamond visible through hazy white stripes (bottom right image of B). Note that the depth relationships are the inverse of those in the top row of A, even though the disparity relationships within the grating are identical. The only difference between the top and middle rows in A is that the luminance of the regions neighboring the diamond apertures was changed from black to white. When the luminance of the adjacent background falls within the range of luminances present in the sinusoidal grating (bottom row of A), the perception of multiple layers is absent or greatly reduced. Panel A is adapted from “Stereoscopic Surface Perception,” by B. L. Anderson, 1999, Neuron, 24, p. 920. Copyright 1999 by Cell Press. Adapted with permission.
between disparity and perceived depth is more complicated. To see this, consider the possible ways that local contrast signals can arise in images. Contrast variations can arise from a host of different environmental causes, but these causes can be grouped into two broad categories based on the depth relationships from which they arise. Image contrast can be generated either from luminance differences that occur along a surface (such as changes in reflectance, illumination, surface orientation, or texture) or by viewing an occluding contour against a more distant background. As I argue below, this simple geometric fact imposes constraints on how depth can be inferred from image contrast.

Consider, for example, a simple binocular vertical luminance discontinuity (a contour) that is sensed by a receptive field with a finite extent (for ease of depiction, the boundary of this hypothetical receptive field is depicted as the square window in Figure 5). Assume that this contour has been binocularly matched and has been assigned a disparity value (D). What can be concluded about the depths of the surface(s) that generated the contour?

One class of possible scene events that could account for this contour is some form of reflectance or illumination difference that arose on a surface, which could have been flat, slanted, or bent. In all such cases, the two luminance values that generated the contour abut at a common position in depth, forming part of a continuous manifold. An alternative interpretation of this contour is that it was generated by an occluding edge. In such contexts, the luminance discontinuity in the images arose from a difference in the luminance projected by an occluding surface and a more distant background. Because the regions around the occluding contour are untextured, only a single disparity will be generated in this image region, although two depths are present at this location in the world (namely, the depth of the occluding contour and the depth of the partially occluded surface). To correctly represent scene geometry, the visual system must somehow assign multiple depths to the single value of disparity generated by this contour segment. In addition to determining that two depths are present at an occluding contour, the visual system must also determine which side of the edge “owns” the contour. In one configuration, the right half of the edge is the occluding contour, and the left half is occluded. In the other, the left half of the edge is the occluder, and the right half is occluded. Note that it is impossible to determine the relative depth of the luminances on either side of the edge from purely local disparity computations. Even if it was somehow known a priori that the contour was caused by an occlusion relationship, the local figure–ground (or “border ownership”) relationships cannot be derived from the local disparity of the contour, which implies that the same local disparity can have very different depth interpretations.

The preceding discussion demonstrates that the local geometry of a matched contrast cannot even specify the number of depths occurring at a contrastive edge, even though there is only a single value of disparity present in this image region. What, then, can the visual system infer about surface depth from this local information? In either occlusion configuration, the partially occluded surface could be situated at any distance behind the occluding surface (i.e., at any depth greater than D), and either side of the contour could own the occluding edge. All that can be derived locally from the classification of the edge as an occlusion relationship is that an occluding contour is at the depth specified by the disparity D, whereas the background can be any distance behind the occluder. The occluding surface does not have the same degree of ambiguity, however. By definition, the occluding surface generated the contour in the two eyes, which means that it generated the disparity signal (D), This implies that the disparity of the occluding surface cannot be changed without causing a corresponding change in the disparity of the contour projected to the two eyes (because the occluding surface owns the contour).

Thus, the disparity of a local contour imposes a constraint on the possible depth assignments that can be attributed to the two sides of the contour. This constraint can be captured by the following principle.

Principle 1: Contrast depth asymmetry—The component luminances that generate a local contrast signal must be as-

---

2 In general, a single disparity cannot specify a depth without some additional information about eye posture. This does not affect the argument developed herein.
signed depths that are greater than or equal to the disparity of the contrast signal.

I refer to this geometric constraint on the assignment of depth as the **contrast depth asymmetry principle (CDAP)**.

One way to understand the perceptual significance of the geometric asymmetry embodied in this principle is to consider what it predicts when the depth of a contour is manipulated. Consider a stereogram in which the two sides of a contour appear in the same plane. If the disparity of the contour is changed in a manner that causes it to move forward in depth, only one side of the contour must be assigned this (new) depth to account for this image data. However, if the contour is moved in a manner that causes it to recede in depth, both sides of the contour must appear to recede to at least this distance, because both sides of the contour are constrained to appear at least this distant in depth. It is this asymmetric “capture” of the two luminances that form the edge that the CDAP is meant to express. It is important to emphasize that this constraint on the possible depth interpretations of local image contrast arises from an asymmetry introduced by the geometry of occlusion. In the following section, I argue that all of the transformations in perceived depth induced by simply changing the sign of disparities that have been reported to date are a consequence of this principle (Anderson, 1998; Ramachandran & Cavanagh, 1985). To see how this transformation is explained by the CDAP, consider the configuration in which the Kanizsa figure appears to float in front of black disks. The two contours to which the CDAP applies intersect in the image and form L junctions; therefore, these junctions are particularly useful for seeing how the CDAP resolves depth ambiguities. Consider first the straight contour segments, which are the nearer of the two contrast signals in this depth configuration. The CDAP restricts the luminance on either side of these straight edges to appear at or more distant to the depth of this contour. The same holds true for the more distant circular arcs, which intersect the straight contours along one of their sides. However, because the CDAP requires both sides of a contrast to appear at least as distant as the contrast signal, the more distant contrast signal must cause the luminance along both sides of the contour to appear at least this distant in depth. Thus, in the regions forming the L junction in which the circular arcs and the straight contour segments intersect, the CDAP requires that both the white surround and the black disks appear in a more distant depth plane.

**Explaining Depth Asymmetries**

The CDAP expresses a constraint imposed by the geometry of occlusion on the pattern of image contrast that forms in the two eyes. By itself, this principle does not appear to provide a particularly powerful constraint for image interpretation, because any single local contrast signal contains a variety of possible depth interpretations. However, when more than one contrast signal is present (i.e., when relative disparity information is available), this principle provides a strong constraint on how depth is assigned.

Indeed, in the following section, I argue that the CDAP provides the explanatory basis for the transformations in perceived surface structure that accompany inversions of relative depth such as those depicted in Figures 1–4. To demonstrate how, I consider each pattern in turn.

**Kanizsa figures.** Consider again the stereoscopic variants of the Kanizsa pattern depicted in Figure 1 (cf. Gregory & Harris, 1974; Lawson & Gulick, 1967; Ramachandran & Cavanagh, 1985). To see how this transformation is explained by the CDAP, consider the configuration in which the Kanizsa figure appears to float in front of black disks. The two contours to which the CDAP applies intersect in the image and form L junctions; therefore, these junctions are particularly useful for seeing how the CDAP resolves depth ambiguities. Consider first the straight contour segments, which are the nearer of the two contrast signals in this depth configuration. The CDAP restricts the luminance on either side of these straight edges to appear at or more distant to the depth of this contour. The same holds true for the more distant circular arcs, which intersect the straight contours along one of their sides. However, because the CDAP requires both sides of a contrast to appear at least as distant as the contrast signal, the more distant contrast signal must cause the luminance along both sides of the contour to appear at least this distant in depth. Thus, in the regions forming the L junction in which the circular arcs and the straight contour segments intersect, the CDAP requires that both the white surround and the black disks appear in a more distant depth plane.
which in turn uniquely determines the border ownership of the (nearer) straight contour segments (see Figure 6A).

When the relative depth of the straight and circular contours are interchanged, the CDAP predicts a very different percept. In this configuration, the straight contour segments are now more distant than the curved segments. The CDAP restricts the luminance on either side of these straight contours to appear at or more distant to the depth of this contour. This means that the luminance along both sides of these edges must recede to (at least) the depth of these straight contours, causing them to appear behind the circular arcs. Because the straight arcs intersect the circular arcs on one of their sides, the CDAP uniquely determines that the other side of the circular arcs must own the contour, causing the arcs to appear as holes in a white surround (see Figure 6B).

Thus, the CDAP explains the striking shift in the perceived surface structure experienced when the relative depths of the contours in these images are interchanged. When the diamond figure appears to complete modally, the black regions appear as isolated disks lying on a white background. But when the diamond figure appears behind the circular inducers, the black regions of the

Figure 6. A schematic illustrating how the contrast depth asymmetry principle (CDAP) explains the shift in surface structure experienced when the disparities of the contours in the Kanizsa figure are interchanged. When the relative disparity of the circular contours and the straight contour segments is inverted, there is a large shift in the perceived surface structure of the Kanizsa figure, particularly of the black portions of the images. When the straight contour segments are in front, the black regions appear as four separated disks (left image of A). When the straight contour segments are behind, the black regions appear as a connected, uniform black surface (or empty space; right image of A). The CDAP requires that both sides of the more distant contour appear at least as distant as this contour, or one side can appear more distant (if it is occluded). Thus, when the circular contour is more distant (left image of B), both sides of that contour are seen at a more distant depth, as depicted in the left image of A. When the straight contour segment is more distant (right image of B), then both sides of the straight contour must recede in depth, as depicted in the right image of A. This allows both the black and white regions adjacent to the straight contour to complete behind the near surface, causing the black regions of the image to appear connected.
inducing figures appear to form a continuous surface that completes amodally behind the circular illusory contours. It should be noted that theories of these two percepts have typically stressed the symmetry in the shape of the interpolated contours that are generated by these images, ignoring the striking transformations in surface structure that arise when the disparity relationships of the contours are inverted (Kellman & Shipley, 1991; but see Anderson, Singh, & Fleming, 2002). Such theories are incapable of providing an understanding of the dramatic shifts in surface structure that accompany the shift in disparity.

Although the preceding discussion was based on the stereoscopic variants of the Kanizsa figures, it should be noted that the CDAP is not restricted to stereoscopic information. The CDAP expresses a constraint on the relationship between image contrast and perceived depth, which means it can be applied to monocular images as well. Indeed, the two percepts of the Kanizsa figure depicted in Figure 1 can be experienced without the presence of stereoscopic information; therefore, it is clear that the principle would be inadequate if it was restricted to stereoscopic vision. The primary difference between the monocular and binocular variants of the Kanizsa figure is that the stereo displays uniquely specify the relative depth of the contours that form the L junctions. When the depths of these contours are ambiguous, the CDAP expresses a constraint on the possible surface structure given a particular depth assigned to the constituent contour segments, which predicts the two percepts that observers report.

The Takeichi et al. (1992) demonstration. The demonstrations of Takeichi et al. (1992) are intimately related to the percepts obtained in the stereo and monocular Kanizsa figures. However, in their variants of this display, there was no depth difference between the contours forming the Kanizsa inducing elements. Rather, a few small dots were added to the display within the boundaries of the illusory figure, and the relative depth of the dots and the Kanizsa figure was manipulated (see Figure 2). When the dots were given a disparity that placed them in front of the Kanizsa figure, they appeared to float in empty space in front of the display. When the dots were placed behind the Kanizsa figure, they appeared to drag the entire region bounded by the illusory contours to their depth, enhancing the strength of the illusory contours in the near depth plane. Takeichi et al. argued that the enhancement of the illusory contours suggested that their phenomena could be related to “occlusion constraints” but did not articulate what those constraints were or how they could explain their phenomena. The theoretical arguments developed herein reinforce the insight that these phenomena are due to occlusion constraints, but the reasoning is quite different.

The main insight needed to understand the phenomena reported by Takeichi et al. (1992) is captured in the CDAP. Although the previous discussion of this principle has focused on the depth assigned to contours, the principle applies to any local contrast signal, such as the contrastive edges between a dot and its surround. In the Takeichi et al. demonstrations, there are two kinds of local contrast signals: those generated by the edges of the Kanizsa figure and those generated by the dots. The CDAP requires that both sides of any local contrast signal are constrained to appear at or more distant to the depth of the contrast signal. Thus, when the contours of the Kanizsa figure are more distant than the dots, both sides of these contours appear in a more distant depth plane, whereas only the interior regions of the dots appear in the near depth plane. When the relative depths of these regions are reversed, the CDAP requires that both the interior of the dots and the regions surrounding the dots must appear as far in depth as that specified by the disparity of the dots’ contrastive edges. This causes both the dots and the regions surrounding the dots to recede to this (more distant) depth. The CDAP was derived from the relationship between local image contrast and the geometry of occlusion; therefore, the preceding analysis is in agreement with Takeichi et al.’s assertion that their phenomena are related to occlusion constraints. Note, however, that the occlusion constraint driving their phenomena is the asymmetric capture of surrounding regions by more distant contrast signals, not by the enhancement of the illusory contours. Indeed, in the present analysis, the enhancement of the illusory contours in their displays is an epiphenomenon of the asymmetric capture of the regions surrounding the dots when the dots appear behind the Kanizsa figure.

Transparency: Decomposing Images Into Stratified Layers of Lightness and Opacity

The preceding analysis focused on understanding how the visual system derives surface structure from stereoscopic images that give rise to percepts of occluding and occluded surfaces. I have shown that the CDAP explains the transformations in perceived surface structure that arise when the relative depths of local contrast signals are inverted. The central insight behind the CDAP is that occlusion geometry introduces a fundamental asymmetry between near and far, which is responsible for the shifts in the way depth is assigned to contrast signals. However, in addition to transformations in perceived depth, recent studies (Anderson, 1999b; Anderson et al., 2002) have shown that the inversion of relative depth can lead to striking transformations in the perceived opacity and lightness of surfaces (see Figures 3 and 4). What is needed to explain these transformations?

To answer this question, I must reconsider the possible causes of a local, binocularly matched contour segment (or more generally, a local contrast signal). In the previous discussion, the only interpretations considered were those depicting opaque surfaces (see Figure 5). More generally, however, a local contrast signal can be generated in one of two ways: by a surface in plain view (that projects a contrast equal to that in the image data) or by a higher contrast signal partially obscured by a (contrast-reducing) transparent surface (see Figure 7). This introduces an additional dimension of ambiguity that cannot be resolved by local computations. How does the visual system determine the number of surfaces present along a given line of sight? What conditions must be met for the visual system to infer the presence of multiple surfaces rather than a single surface in plain view?

For nearly three decades, the answers to these questions were thought to be given by Metelli’s (1970, 1974a, 1974b) seminal work. Metelli developed a quantitative model of transparency that was based on images generated by viewing a two-toned background viewed through a rapidly rotating disk with a missing sector (generating images similar to those depicted in Figure 8). Metelli derived equations for the transmittance and reflectance of a transparent filter and showed that a number of ordinal inequalities had to be satisfied for these equations to be consistent with the presence of a transparent surface. These inequalities were derived by recognizing that the filter’s transmittance and the reflectance
are both proportions and therefore must be positive and fall on the interval (0, 1). It has been recently argued that there are two basic constraints that can be derived from Metelli’s model (Anderson, 1997; cf. Metelli, 1974a, 1974b). The restriction that the transmittance of the transparent filter is positive implies that a transparent layer cannot reverse the polarity of an underlying contour (i.e., whether it is a dark–light or light–dark edge). Second, the restriction that the transmittance lies on the interval (0, 1) implies that the transparent layer must reduce the luminance difference of the underlying contour.

What impact does the preceding discussion have on the CDAP? Above, it was argued that a contour that has a disparity $D_0$ (or more generally, a depth $D_0$) implies that the luminances on both sides of the contour must be positioned at a depth equal to this depth or that one side of the contour can be more distant (if it is occluded). However, the possibility that a transparent surface is present means that not all of the luminance present along the two sides of a contour must obey this constraint. In regions of transparency, some of the luminance reaching the eyes is (typically) due to the transparent surface as well as the underlying surface. However, because transparent surfaces cannot reverse the polarity of underlying contours, the CDAP expresses an inviolable constraint on the ordinal lightness relationships of the underlying surfaces in regions of transparency. Stated differently, the CDAP expresses a constraint on how depth is assigned to the polarity of image contrast, but it does not uniquely determine how depth is assigned to contrast magnitudes.

To understand the meaning of this assertion, consider a contour that has a contrast polarity that is dark–light and has some depth associated with it (such as by generating a local disparity signal). The CDAP requires that there must be a dark–light transition at the depth of this edge or that one side of the contour is more distant. However, this dark–light contrast could have been formed by either a moderately contrastive contour in plain view or by a higher contrast marking overlaid by a contrast-reducing filter. Indeed, there are an infinite family of possible combinations of transparent and underlying surfaces that could give rise to any given image contrast; therefore, the CDAP is not sufficient to determine whether visible contrast is seen in plain view or is partially obscured by a transparent surface. How, then, does the visual

---

3 The designation of the polarity of an edge requires the imposition of a coordinate system. The arguments developed herein are independent of the coordinate system chosen; therefore, the reader is free to impose a coordinate system of his or her choosing.
Another shortcoming of Metelli’s theory is that it requires that a transparent surface (or medium) be homogeneous in both transmittance and reflectance, which means that it cannot account for perceptions of transparency observed when viewing scenes through smoke or fog. Unlike the uniform filters introduced by Metelli (1970, 1974a, 1974b), smoke and fog can generate a continuous range of transmittance values (although they are roughly uniform in reflectance). Thus, a more general set of theoretical principles is needed for a theory of transparency that is capable of predicting when transparency is and is not perceived and that accounts for the perception of transparency when the transmittance of the two layers is not uniform.

**Beyond Metelli: The Transmittance Anchoring Principle (TAP)**

To provide insight into the complexity of the problem that must be solved for the visual system to compute transparency, assume the visual system has determined that a transparent surface is present in some portion of the image and the problem is simply one of localizing the image region containing it. For simplicity, also assume that the reflectance of the transparent material is constant (which is approximately true for smoke or fog). In such viewing conditions, the visual system must determine how the pattern of luminance values in the image is mapped onto a set of transmittance (or opacity) values of a transparent surface layer. In regions in which the transparent surface is not present, a transmittance value of 100% must be assigned, whereas some lower value(s) of transmittance must be assigned in regions occupied by a transparent surface. In general, the visual system has to map the intensity distributions on the retinas onto a set of surface properties of the transparent layer and the underlying surface. A conceptually related mapping problem arises in the lightness domain, which has been studied more extensively than the related problem in the transparency domain. To motivate the discussion that follows, I begin by considering how such mapping problems have been addressed in the perception of surface lightness. In so doing, I show that there are strong parallels between the conceptual issues that must be addressed in both lightness and transparency perception.

In lightness perception, the visual system must infer surface reflectance from the luminance values on the eyes. The luminance values on the retinas are products of surface reflectance and the illumination; therefore, the visual system must somehow “undo” this multiplicative relationship to infer surface lightness. The problem is that there is no simple or unique way to do this decomposition because there are an infinite number of combinations of illumination and reflectance that could have given rise to any image luminance. Suppose, for example, that an observer was able to determine that a surface patch x has twice the reflectance of patch y. Although this information solves the problem of determining the relative lightness of the two patches, it does not determine the absolute reflectances because this information is consistent with an infinite family of possible reflectance pairs. For example, the patch reflectances could be 5% and 10% (black and dark gray), 15% and 30%, or 35% and 70% (mid-gray and nearly white). All such pairs satisfy the 2:1 reflectance ratio. To resolve this ambiguity, one needs some additional principle to understand

---

**Figure 8.** Three displays similar to those studied by Metelli (1970, 1974a, 1974b) to evaluate his episcotister model of transparency. **A:** The regions p and q appear as a mixture of two surfaces—a transparent filter and the background regions a and b. **B:** When the contrast of the central contour where the regions a and b meet is lower than the contrast of the central contour where the regions p and q meet, the regions p and q no longer appear transparent. Rather, the surrounding regions a and b can appear as containing a transparent surface with a hole cut out of its center. **C:** The polarity of the central contour reverses, and no perception of transparency is experienced.

---

system determine when more than one layer is present? How is luminance partitioned between multiple depth planes when multiple surfaces are present? What image conditions must be present for multiple surfaces to be perceived, and what does this reveal about the mechanisms responsible for generating a representation of surfaces stratified in depth?

The answers that have been proposed to these questions have typically focused on the conditions in which Metelli’s equations could be applied (e.g., Beck, Prazdny, & Ivry, 1984; Gerbino, Stultiens, Troost, & de Weert, 1989). This imposes a number of restrictions on the kinds of transparent surfaces that have been studied experimentally. Metelli’s equations can only be solved if the transparent surface is balanced (uniform) in both transmittance and color (Metelli, 1970, 1974a, 1974b; Metelli, da Pos, & Cave don, 1985). There is also an assumption that a bipartite background surface is present and seen in plain view, but there is no explana-
how these relative measures of lightness are anchored to an absolute scale of reflectance values. This has been described as the anchoring problem in lightness perception, and a number of experiments have shown that the visual system exhibits a bias to anchor the highest luminance to white (as opposed to, for instance, anchoring the mean luminance to mid-gray, or the darkest luminance to black; cf. Land & McCann, 1971; Gilchrist et al., 1999).

In addition to this anchoring problem, there is also another problem that must be solved in lightness perception. The highest-luminance-is-white anchoring rule explains only how one point of the lightness scale is determined. An additional principle is needed to understand how all of the other luminance values present in the image are mapped onto a set of perceived reflectance values. This is known as the scaling problem in lightness perception. One of the most well-known scaling rules was described by Wallach (1948), who proposed that there is a one-to-one mapping between ratios of image luminance and ratios of perceived lightness. In this theory, two stimuli that have the same luminance ratios are predicted to have the same perceived lightness. In support of this thesis, Wallach (1948) found that two disk–annulus displays appeared equal in lightness when the disk–annulus luminance ratios were equal.

Thus, a theory of lightness perception requires both a scaling rule and an anchoring rule to explain how the luminance values on the eyes are transformed into a set of perceived lightness values. A related set of ambiguities arises when attempting to infer the opacity of transparent media from the distribution of luminance values on the retinas. First, as in lightness perception, there is an anchoring problem. In the context of transparency, the problem is one of anchoring the perceived transmittance of a transparent layer. Suppose, for example, that it was (somehow) known that two transparent surfaces were present in a scene and that one of the surfaces was twice as transmissive as another. Just as in the lightness domain, this only provides relative information about the opacity of the transparent surfaces; it does not specify their absolute degrees of opacity. Here, too, an anchoring principle is needed to resolve this ambiguity and tie these relative measures of opacity onto an absolute scale. I have recently proposed that the visual system resolves this ambiguity by treating the highest contrast segment of a uniform polarity contour as unobscured, that is, devoid of a transparent layer (Anderson, 1999b). More specifically, this principle can be stated as follows.

**Principle 2:** Transmittance anchoring—The highest contrast region along a continuous contour that undergoes changes in contrast magnitude, while preserving contrast polarity, will appear as a surface in plain view, whereas lower values of contrast along such contours are decomposed into multiple layers.

Intuitively, this principle asserts that the highest contrast segment along a continuous, smooth contour serves as a reference—or anchor—that is used to scale the extent to which lower contrast regions along the contour are obscured by partially transmissive surfaces or media.

To understand how this anchoring principle constrains the perception of transparency, consider the variants of Metelli’s displays depicted in Figure 8. In Figure 8A, the central regions appear to be composed of two layers, one of which is transparent. In this figure, the higher contrast portion of central contour lies outside the central circle. The TAP asserts that this region should appear in plain view (i.e., unobscured by a transparent haze), which is what observers report. However, in Figure 8B, the same two shades of gray are now the highest contrast contour segment in this display. The TAP therefore predicts that the central region of this display should appear as surfaces in plain view, whereas the surrounding region should now appear to contain multiple layers. This is what observers report (cf. Metelli, 1974a, 1974b). Indeed, in all variants of Metelli’s displays, the highest contrast contour segment along the central contour appears as a pair of surfaces in plain view, whereas the lower contrast segment appears to contain multiple layers, the nearer of which is transparent.

It should also be noted that the TAP can also explain why no percept of transparency is experienced if the contrast polarity of the central contour reverses. The TAP states that the perception of transparency is caused by changes in the contrast magnitude of a contour that preserves contrast polarity. Thus, in Metelli displays that contain polarity reversals, all of the contour segments are the highest contrast of that polarity. This implies that transparency should not be experienced when this occurs, consistent with observers’ responses. This occurs for the straight contours in Figure 8C and for all of the circular contours in Figure 8. When multiple polarity preserving contours are present, and they both undergo a reduction in contrast (such as occurs along the vertical and horizontal contours forming the X junctions in Figure 9), both surfaces can appear transparent, as predicted by the TAP.

However, just as an anchoring theory does not provide a complete account of the perception of surface lightness, an anchoring theory of transparency does not provide a complete account of the perception of transparent surfaces. The TAP only describes how one point along the scale of possible transmittance values is computed. As in the lightness domain, an additional principle is needed to understand how quantitative estimates of surface opacity are assigned to all other transmittance values. Given that the TAP is grounded on measures of contrast along continuous contours, a conceptually consistent theory would require that the scaling of

---

**Figure 8** shows the central regions appear transparent.

**Figure 9** Multiple contours forming X junctions that both preserve polarity and undergo a reduction in contrast. Note that both rectangles can appear transparent.
perceived opacity be some function of the contrast change along contours. Singh and Anderson (2002) recently performed quantitative experimental and theoretical investigations into this scaling problem and found that the perceived opacity of a transparent layer is determined by the contrast ratio of the lower contrast contour segments (the region of transparency) relative to the higher contrast contour segments (i.e., the region seen in plain view). This scaling rule is similar to Wallach’s (1948) ratio rule in the lightness domain, except that it involves the ratios of Michelson contrast rather than of luminances. The experimental results on the scaling of perceived transmittance and their implication for extant models of transparency are discussed at length elsewhere (Singh & Anderson, 2002). In this article, I focus on how the CDAP and the TAP determine the perception of surface lightness and opacity in stereoscopic images.

In summary, I have argued that there is a strong analogy between the computation of transparent surface properties and the computation of surface lightness. First, in the lightness domain, the visual system appears to adopt a strategy whereby the highest luminance in an image is anchored to white. In the transparency domain, the contrast of underlying contours serves as an anchor that is regarded as a surface in plain view. This is tantamount to asserting that the visual system anchors the transmittance scale to 100%, analogous to anchoring the most reflective surface to white in the lightness domain. Second, in the lightness domain, evidence suggests that the visual system uses the ratio of luminances to scale perceived lightness (Wallach, 1948). In the transparency domain, it has been shown that the scaling of perceived opacity is determined by the ratio of Michelson contrasts along continuous contours (Singh & Anderson, 2002). In the following section, I argue that the TAP and CDAP provide a unified explanation of the perceived depth, lightness, and opacity experienced when viewing stereoscopic and monocular images generating percepts of opaque and transparent surfaces.

The Role of the CDAP and TAP in Determining Surface Depth and Lightness

Two main principles shape the theory described in this article. The CDAP embodies the projective asymmetries caused by occlusion and imposes a local constraint on the ordinal luminance values that are assigned a particular depth. In contradistinction, the TAP can be viewed as a perceptual Occam’s razor that assures that transparency is not a ubiquitous perceptual quality experienced when viewing all regions of all scenes. This principle states that transparent surfaces (or media) are only inferred when there is a perturbation in contrast that must be in some sense explained. Note that the CDAP is inherently local because it can be applied to any local contrast value, whereas the TAP is inherently global because it requires comparing the magnitudes of contrast along a contour. I argue that these two principles provide a unified explanation of all stereoscopic occlusion and transparency displays that have been reported to date. In the previous section, I demonstrated how the TAP could be used to explain Metelli displays. In the following section, I demonstrate how this principle works together with the CDAP to provide a unified explanation of illusory transparency percepts elicited by simple stereograms containing stereoscopic T junctions (see Figure 9) as well as illusory percepts of inhomogeneous transparency depicted in Figures 3 and 4 that are beyond the scope of Metelli’s model.

Illusory Transparency Induced by Stereoscopic T Junctions

A number of authors have recently reported that illusory percepts of transparency can occur in displays containing only three luminance values and that such percepts can be strengthened through the introduction of stereoscopic depth to the T junctions in these images (see Figure 10; see Anderson, 1997; Nakayama, Shimojo, & Ramachandran, 1990; Watanabe & Cavanagh, 1993).

Figure 10. The three qualitative kinds of transparency elicited by stereoscopic T junctions and their corresponding junction types. Cross-fusers should fuse the left two columns in A, B, and C, and divergent fusers should fuse the right two columns in A, B, and C. When the top of the T junction is the darkest luminance (A), the intermediate luminance adjacent to the stem of the T appears as a transparent filter that appears to darken the surfaces it overlies (bottom panel, left image). When the top of the T junction is the lightest luminance (B), the intermediate luminance adjacent to the stem of the T appears as a transparent filter that appears to lighten the surfaces it overlies (bottom panel, middle image). When the top of the T junction is the intermediate luminance (C), the near contour appears to float in empty space as a disembodied contour (bottom panel, right image).
Previous accounts of these phenomena have noted that transparency was experienced when the luminance relationships of these images were consistent with a (degenerate) version of Metelli’s equations (Nakayama et al., 1990; Watanabe & Cavanagh, 1993). Although Metelli’s equations do correctly predict when transparency will and will not be experienced in these patterns, Singh and Anderson (2002) recently showed that Metelli’s model completely fails to predict the perception of surface opacity and lightness of transparent surfaces. It is therefore instructive to see how the CDAP and TAP explain transparency percepts in these simple displays before demonstrating their success in explaining more complex patterns of transparency experienced in Figures 3 and 4.

It has been empirically shown that two conditions must be satisfied to elicit percepts of illusory transparency in displays containing stereoscopic T junctions: The continuous contour segment forming the top of the T junction must be more distant than the T junction stem, and the contrast along this distant contour must vary in magnitude without changing polarity (Anderson, 1997). When this occurs, illusory percepts of transparency arise along the side of the T junction that generate the lower contrast. How do the CDAP and TAP explain these percepts? The CDAP requires that there must be a luminance relationship at the far depth that is responsible for the polarity of the more distant contour (the top of the T junction in these images). The TAP states that the highest contrast region along this polarity preserving contour (the triangle and background borders in Figures 10A and 10B) should appear as a contour in plain view—that is, unobscured by a transparent layer—whereas the lower contrast region should appear as multiple surface layers, one of which is transparent (the triangle and disk borders in Figures 10A and 10B). This is what observers report (Anderson, 1997; cf. Nakayama et al., 1990; Watanabe & Cavanagh, 1993).

In addition to correctly predicting when transparency should be experienced in these images, these principles also account for similar displays that do not generate percepts of transparent surfaces (see Figure 10C). This occurs when the contrast polarity reverses along the top of the T junction (i.e., when the top of the T junction is the intermediate luminance). In such conditions, the CDAP requires that there must be a luminance relationship at or more distant than the depth of the top of the T junction. Note that for this display, the more distant contour reverses contrast polarity, causing a shift in the lightness polarity of the surfaces that abut in the distant depth plane. The TAP states that the highest contrast contour segment of a continuous contour that preserves polarity should appear in plain view (note that the TAP only applies to contours that have the same polarity). When the polarity of the more distant contour reverses, the anchoring principle must be applied to each contour segment individually. Thus, the TAP requires that both contour segments should appear as surface regions in plain view, causing the near contour (the T junction stem) to appear disembodied, that is, without any visible surface lightness attached to the contour. This is what observers report (cf. Anderson, 1997; Anderson & Julesz, 1995).

In summary, the constraints imposed by the CDAP and the TAP provide a unified account of both stereoscopic and nonstereoscopic images that generate percepts of transparency in images composed of contours and contour junctions. In the following section, I argue that these principles are sufficiently general to predict the pattern of perceived depth, lightness, and opacity of stereoscopic displays that elicit inhomogeneous percepts of transparency.

**Beyond Junctions: Generalized Computations of Surface Opacity and Lightness**

To test the generality of these principles, I constructed a new class of stereoscopic stimuli (see Figures 3 and 4). The design of these stimuli was motivated by the intuition that the perception of transparency observed in stereoscopic displays containing contour junctions was driven by two image properties: the presence of a depth discontinuity between intersecting contours (to which the CDAP would apply) and the preservation of polarity of the more distant contour (to which the TAP would apply). One of the simplest ways to satisfy these conditions without generating explicit contour junctions is to replace one of the contours with some form of luminance gradient. An example of one such pattern is presented in Figure 3, which depicts a sinusoidal luminance grating abutting a diamond shaped contour. Note that there are no explicit contour junctions in these images. Rather, the continuous luminance gradients of the grating generate continuous contrast fluctuations along the length of the aperture boundary. Depth is introduced by interocularly shifting the grating relative to the aperture boundary. When the aperture boundary appears in front of the grating, observers report that the grating appears as a surface that lies in a single depth plane behind the aperture boundary, as all models of stereopsis would predict. However, when the depth relationships are inverted, observers report that the grating appears to split into two layers: a near layer containing a series of transparent bands that vary in opacity and a homogeneously colored, diamond-shaped surface at the depth of the aperture edges. Remarkably, the appearance of the two layers depends dramatically on the luminance of the surround. When the surround is black, the far layer within the texture appears white, and the transparent layer appears as bands of black fog that vary in transmittance. However, when the surround is white, the far layer appears black, and the transparent layer appears as bands of white fog that vary in transmittance. Note that in both cases, there are regions in which the near layer appears to be completely transmissive, providing an unobscured view of the more distant layer.

The decomposition of textures into multiple surface layers is not restricted to the one-dimensional luminance variations of the grating depicted in Figure 3 and can be observed with a broad class of 2-D textures. Figure 4 depicts one example. A uniform disparity texture is viewed through three apertures, and the aperture boundaries are shifted horizontally relative to the texture. When the aperture boundaries appear in front, the textures appear in a single depth plane visible through three holes. However, when the two eyes’ views are interchanged (inverting the relative disparities), the figure appears as three light disks visible through dark colored mist. Throughout the texture, the mist appears to be approximately uniform in color but varies in perceived opacity. When the background luminance is changed to light gray, the apparent lightness of the disks shift to dark gray, and the cloudy texture takes on an appearance of light smoke. As with the examples in Figure 3, the regions of the texture that appear unobscured by the transparent layer reverse when the luminance of the surround is changed from light to dark: The lightest regions in the top stereopair appear as
portions of light disks that are unobscured by the dark transparent clouds, but these same image regions appear in the front of the disk in the bottom stereopair as the densest portions of the transparent layer. Note again that there are regions in which the near layer appears completely transmissive for both surround luminances, providing an unobscured view of the more distant layer.

In the following section, I argue that the CDAP and TAP can explain the dramatic transformations in perceived depth, lightness, and opacity that occur when the depth and luminance polarity of the surround are inverted in Figures 3 and 4. There are two primary intuitions that guide this explanation. First, the CDAP is needed to explain why the apparent lightness of the near and far layers inverts when the luminance of the surround is switched from light to dark. The inversion of the surround luminance causes the polarity of the aperture boundary to reverse. When this contour is more distant than the texture, the CDAP requires that there must be something at or more distant than the polarity of the aperture boundary, which must reverse when the polarity of the contour reverses. Second, the TAP provides the explanatory principle needed to understand why there are regions of the more distant surfaces that appear unobscured by a transparent layer.

I begin by considering what the CDAP and the TAP predict for the grating pattern. In the depth configuration that generates a percept of transparency, the contrastive border of the aperture boundary is more distant than the luminance modulations of the grating. The CDAP requires the presence of a luminance relationship at or more distant than the depth of the aperture boundary to account for the polarity of this contour. In this example, the surround is black, and the polarity of the surround–grating border is dark–light (respectively) along the entire boundary (see Figure 3A, top row). This means that there must be a surface that is lighter than the surround within the grating. The TAP asserts that the highest contrast contour segment along this more distant contour corresponds to a region of the boundary that is in plain view. In this example, the highest contrast contour segments arise where the surround abuts the maxima of the grating (the white portions of the grating). Thus, the maxima within the grating are predicted to appear as portions of a distant surface in plain view.

When the contrast polarity of the aperture boundary is inverted, the TAP and CDAP predict a dramatic shift in the depth and lightness of this stimuli. The simplest way to transform the polarity of this contour is to lighten the (dark) surround so that it becomes lighter than all of the luminance values within the grating (see Figure 3A, middle row). The CDAP now requires that there must be something darker within the grating at or more distant than the depth of the aperture boundary to account for the light–dark polarity of the surround–grating boundary. The TAP states that the highest contrast contour segment should (again) appear in plain view, which now occurs where the surround intersects the minima within the grating. Consistent with this prediction, observers report that the grating appears to split into a set of misty light stripes hovering in front of a homogeneous, dark diamond. Note that the perceived depth of the minima and maxima, the apparent lightness of the two layers, and the relative opacity of the transparent layers were all inverted by simply changing the luminance of the surround. A similar transformation in depth and surface lightness is experienced in the 2-D textures depicted in Figure 4.

The dramatic transformation in surface structure obtained when viewing these figures is completely outside the scope of current theories of stereopsis. However, the qualitative characteristics of these phenomena can be understood to follow directly from the CDAP and TAP. The question remains whether these principles can also be used to derive quantitative predictions about the pattern of perceived depth and lightness of the surfaces experienced in these patterns. There are three aspects of the textured patterns in Figures 3 and 4 that are dramatically affected by manipulating the luminance of the surround. First, the pattern of perceived lightness of the two layers within the texture is inverted. Second, the perceived depth of the maximal and minimal luminance values within the texture is inverted. Third, the relative opacity of the near (transparent) layer is transformed. The CDAP and the TAP make explicit predictions about the shift in perceived depth and the transformation in apparent lightness of the two layers. The transformation in the perceived opacity of these figures requires a theory of how the transmittance of transparent surfaces are scaled, which is described at length elsewhere (Singh & Anderson, 2002). Here, I focus on the role played by the CDAP and the TAP in determining the perceived depth and surface lightness of these patterns.

What quantitative information about surface structure can be derived from these two principles? First, the CDAP restricts the ordinal lightness relationships that must exist at a given position in depth. Thus, in these images, it predicts that the far layer within the texture must appear light when the polarity of the surround–texture boundary is dark–light (respectively) and, conversely, that the far layer within the texture must appear dark when the polarity of the surround–texture boundary is light–dark (respectively). Second, the TAP states that the highest contrast contour segment of the far contour (the surround–texture border) should appear in plain view. This implies that the perceived lightness of the far layer within the texture should be identical to the perceived lightness of the textural components that generate the highest contrast contour segments along the aperture boundary. The perceived lightness of the underlying layer should therefore correspond to the perceived lightness of the grating’s minimal or maximal values, depending on whether the surround luminance was higher or lower than the grating, respectively. (Note that the apparent lightness of the more distant surface could not be predicted with these principles if there was no region that was predicted to appear in plain view because the lightness of the transparent layer and the lightness of the underlying surface would be conflated.) A recent series of psychophysical experiments have confirmed these predictions (Anderson, 1999b).

In addition to predicting the apparent lightness of the layer underlying the transparent surface, the CDAP and TAP also predict a shift in the perceived depth of the minimal and maximal luminance values within the textures. The CDAP determines whether the far layer within the texture should appear light or dark, and the TAP states that either the highest or lowest luminance in the texture should appear as portions of the far surface in plain view (depending on whether the surround is dark or light, respectively). This implies that the perceived depth of the maximal and minimal regions within the texture should appear to shift when the polarity of the surround–texture boundary is inverted, which is what observers report.
General Discussion

Since the invention of the RDS (Ashenbrenner, 1954; Julesz, 1960), the study of stereoscopic vision has focused on the correspondence problem. The absence of familiar shapes and objects in these patterns provided incontrovertible evidence that higher level form recognition was not a prerequisite for binocular matching (Ogle, 1959), situating stereoscopic processing within the province of early visual processing. But with this achievement came a new puzzle: How does the visual system determine corresponding features when there are so many candidates to choose from? This question has become the conceptual focus of virtually all stereoscopic models and has remained as the primary theoretical issue addressed by these models since the invention of the RDS. Indeed, it has been forcefully argued that the only interesting theoretical problem that arises in stereopsis is understanding how correspondence is achieved (Julesz, 1971; Marr, 1982). Once correct matches are determined, a disparity map can be generated, which is thought to contain all of the necessary information needed to derive stereoscopic surface structure. The transformation from a disparity to perceived depth seemed to involve a simple trigonometric transformation of the disparity map; it did not seem to introduce any theoretically interesting difficulties of its own.

In this article, I have argued that this view of stereopsis is incorrect and demonstrated that local disparity estimates are often not sufficient to derive a unique representation of surface structure. More specifically, I have argued that an understanding of stereoscopic surface perception requires understanding the relationship between local image contrast and the computation of depth, lightness, and surface opacity. Given the physiology of the early visual pathway, it now seems almost self-evident that the visual system uses some measures of local image contrast as the primitives for stereoscopic matching. Indeed, the very concept of disparity is predicated on the availability of monocular positional signals that can be compared in the two eyes. However, even though disparity is computed from the positional shifts of local image contrast, the visual system must assign depth to the component luminances that generated these contrast signals. Perhaps the simplest way to appreciate this fact is to consider the case in which a locally matched (untextured) luminance discontinuity arises from an occluding edge, as discussed above. When this occurs, only a single disparity signal will be generated by the edge, but the visual system nonetheless must assign two depth values to this single value of disparity to correctly compute the 3-D structure of a scene. This implies that there are common situations in which regions in the disparity map do not contain sufficient local information to derive 3-D surface structure.

The analysis presented above revealed that the polarity and magnitude of image contrast play distinct roles in constraining the perception of stereoscopic surface structure. The most fundamental constraint articulated in this article is imposed by the polarity of image contrast and was dubbed the contrast depth asymmetry principle. This principle expresses the asymmetric role of near and far image contrasts in specifying stereoscopic surface structure. The asymmetry embodied in this principle arises from the geometry of occlusion. The fact that nearer surfaces occlude more distant surfaces, and not conversely, has significant implications for the pattern of contrasts that fall on the two eyes. At the core of this principle is the recognition that image contrast can arise from the projection of surface structure onto the eyes in only two general ways: from luminance changes that occur within the visible boundaries of surfaces (i.e., reflectance changes, changes in surface orientation [shading], illumination differences, and texture) or by the projection of an occluding surface against a more distant background. This, in turn, imposes a constraint on the ordinal lightness relationships—the relative polarity of surface lightness—that can be attributed to a contrast signal. More specifically, this constraint states that a dark–light transition at a given position in depth implies that there must be some surface(s) either at or more distant to this position in depth that is responsible for the dark–light polarity of this contrast signal. The caveat that one side of a contrast signal can appear more distant in depth arises from the possibility that the contrast signal was generated by an occluding edge viewed against a more distant background.

Note that the CDAP applies to all local image contrast. Thus, not only is this principle capable of explaining the perceived depth of occluding and transparent surfaces, but it also can be used to derive the perceived depth in all other stereograms as well.

The CDAP expresses a geometric constraint on the ordinal lightness and depth relationships that can be attributed to binocular image contrast. However, by itself, it was shown that this principle cannot uniquely specify the magnitude of the contrast signal that must be present at the depths consistent with this principle; it only embodies a constraint on the polarity of that image contrast, and hence, it does not determine the absolute lightness values of the surfaces in a scene (or the number of surfaces present along a given line of sight). This is because any given image could have been generated by surfaces in plain view or by a higher contrast scene viewed through a contrast-reducing haze (or transparent surface). This ambiguity implies that an additional principle is needed to understand how the visual system distinguishes between viewing a surface in plain view and viewing a surface through a partially transmissive haze (or filter). It is always possible for a transparent surface to be present over the entire scene; therefore, there is never any way for the visual system to veridically determine whether it is viewing a surface in plain view. Thus, any constraints articulated on how the visual system infers transparency can only reflect heuristics imposed by the visual system that are used to guess how likely it is that transparent surfaces are or are not present. Such principles do not have the geometric force embodied in the CDAP and, hence, can only be derived empirically.

In this article, I have proposed that the visual system applies a visual version of Occam’s razor to compute transparency. I argued that the visual system treats the highest contrast portion of a contour as a region in plain view—an anchor—against which perturbations in contrast magnitude are scaled to infer the presence of transparency (cf. Anderson, 1999b). Not only does this anchoring principle account for all versions of homogeneous transparency that were reported by Metelli (1970, 1974a, 1974b), but it also successfully predicts all reported percepts of inhomogeneous transparency—that is, percepts of transparency that are inhomogeneous in transmittance (such as fog and smoke). The concept of anchoring has been previously used in the domain of lightness perception to understand how the visual system attaches reflectance values to patterns of image luminance. In the lightness domain, the visual system appears to have a bias to anchor the highest luminance to white. The anchoring rule in lightness is also...
an empirically derived rule, not an ecologically imposed constraint. Indeed, to this date, no principled reason has been articulated that explains why the visual system should anchor the highest luminance to white; the fact that the visual system apparently invokes this particular anchor is currently just a description of an empirically observed bias. Thus, in some sense, the anchoring principle in lightness is less principled than the TAP, because the highest-luminance-is-white anchoring principle has not been ascribed any descriptive economy. In contradistinction, in the transparency domain, the TAP can be viewed as a form of Occam’s razor in perceptual inference: It states that the visual system assumes the minimal number of surfaces needed to account for the data and only infers transparency when there are perturbations in contrast magnitudes that provide evidence for its presence.

The theoretical framework articulated in this article focused on a particular class of images, namely, those containing luminance discontinuities and/or gradients that intersect in two different depth planes. In part, this focus was inspired because these properties have been present in the vast majority of articles that have evaluated the mechanisms of stereoscopically induced transparency and occlusion (Anderson, 1997; Anderson & Julesz, 1995; Nakayama & Shimojo, 1992; Nakayama et al., 1990; Ramachandran & Cavanagh, 1985; Watanabe & Cavanagh, 1993). However, there is also a more principled reason for this focus. The CDAP forms the theoretical heart of this article, and as I have shown, it arises from the geometry of occlusion. The study of depth discontinuities is a natural place to observe the consequence of occlusion in surface perception and therefore provides a natural starting point to observe constraints imposed by occlusion geometry. Nonetheless, it should be noted that depth discontinuities are not necessary to observe the asymmetries embodied in this principle. The demonstrations reported by Takeichi et al. (1992) do not contain any explicit depth discontinuities but nonetheless generate similar geometric asymmetries of the kind experienced with patterns that do (such as the stereo Kanizsa figures).

In this article, I have articulated two theoretical principles that are needed to understand how the visual system resolves occlusion and transparency relationships. I have argued that these principles provide the theoretical foundation for understanding the geometric and photometric asymmetries captured in Figures 1–4, which are completely outside the scope of extant theories of stereopsis. However, there are a number of remaining aspects of these phenomena that need to be solved to establish a complete theory of these percepts. Many of the displays considered herein contain large image regions that are untextured (such as Figures 1 and 2), which means that there are no local contrast signals in these regions to which the CDAP can be applied. This implies that the constraints on depth and lightness imposed by the CDAP must be interpolated from image regions that contain local contrast signals. The theory developed in this article does not address the nature of the mechanisms and computations underlying this interpolation process. It should also be noted that this interpolation problem not only is restricted to image regions devoid of texture but also occurs even in images that contain relatively dense contrast variations (such as Figures 3 and 4). In these patterns, the contrast relationships that occur at the depth discontinuities caused the nearer textures to be decomposed into multiple layers throughout the texture, not just along the positions of the discontinuities. One of the main problems that remains for further theoretical and empirical investigation is the mechanisms and computations underlying these interpolation processes.

References


Metelli, F. (1974a). Achromatic color conditions in the perception of


