Depth spreading through empty space induced by sparse disparity cues

Xintong Li Rutgers New Jersey Medical School, Newark, NJ, USA

Abigail E. Huang University of California, San Diego School of Medicine, San Diego, CA, USA

Eric L. Altschuler Rutgers New Jersey Medical School, Newark, NJ, USA

Christopher W. Tyler Smith-Kettlewell Eye Research Institute, San Francisco, CA, USA

A key goal of visual processing is to develop an understanding of the three-dimensional layout of the objects in our immediate vicinity from the variety of incomplete and noisy depth cues available to the eyes. Binocular disparity is one of the dominant depth cues, but it is often sparse, being definable only at the edges of uniform surface regions, and visually resolvable only where the edges have a horizontal disparity component. To understand the full 3D structure of visual objects, our visual system has to perform substantial surface interpolation across unstructured visual space. This interpolation process was studied in an eight-spoke depth spreading configuration corresponding to that used in the neon color spreading effect, which generates a strong percept of a sharp contour extending through empty space from the disparity edges within the spokes. Four hypotheses were developed for the form of the depth surface induced by disparity in the spokes defining an incomplete disk in depth: low-level local (isotropic) depth propagation, mid-level linear (anisotropic) depth-contour interpolation or extrapolation, and high-level (anisotropic) figural depth propagation of a disk figure in depth. Data for both perceived depth and position of the perceived contour clearly rejected the first three hypotheses and were consistent with the high-level figural hypothesis in both uniform disparity and slanted disk configurations. We conclude that depth spreading through empty visual space is an accurately quantifiable perceptual process that propagates depth contours anisotropically along their length and is governed by high-level figural properties of 3D object structure.

Introduction

The primary goal of visual encoding is to determine the nature and motion of the objects in the surrounding environment. In order to plan and coordinate actions, we need a functional representation of the three-dimensional (3D) scene layout and of the spatial and depth configuration of the objects within it. The visual information provided to each eye is, however, two-dimensional (2D), and the 2D configurations of objects in the visual array have an entirely different metric structure from that of the spatial configuration of the visual cues that convey the presence of objects to the brain, or to artificial sensing systems and share none of the physical properties constituting the objects. In general, the visual cues may change in luminance or color, or they may be disrupted by reflections or occlusion by intervening objects. The particular cues such as edge structure, binocular disparity, color, shading, texture, and motion vector fields may carry discordant information about different aspects of an object. Importantly, many of these cues may be sparse, with missing information about the object structure across occlusions or gaps where there are no edge or texture cues to carry information about the object shape.

Thus, a primary requirement of neural or computational representations of the shape of objects is the reconstruction of the 3D configuration and filling-in of its depth surfaces across regions of missing or discrepant information in the local visual cues. Computational approaches to the issue of the structure of objects tend to take either low-level or high-level

Citation: Li, X., Huang, A. E., Altschuler, E. L., & Tyler, C. W. (2013). Depth spreading through empty space induced by sparse disparity cues. *Journal of Vision, 13*(10):7, 1–11, http://www.journalofvision.org/content/13/10/7, doi:10.1167/13.10.7.

doi: 10.1167/13.10.7 Received December 5, 2012; published August 14, 2013 ISSN 1534-7362 © 2013 ARVO

Downloaded from jov.arvojournals.org on 10/20/2019

approaches to the problem. Low-level approaches begin with local feature recognition and attempt to build up the object representation by hierarchical convergence, using primarily feedforward logic with some recurrent feedback tuning of the results (e.g., Marr, 1982; Grossberg, Kuhlmann, & Mingolla, 2007). High-level, or Bayesian, approaches begin with the vocabulary of likely object structures and look for evidence in the visual array as to which object might be there (e.g., Huang & Russell, 1998; Rue & Hurn, 1999; Moghaddam, 2001; Stormont, 2007). Both approaches work well for objects with a stable 2D structure (as in a typical laboratory setup), but are easily confused when viewing a complex 3D scene with poorly-defined depth cues (such as a dinner table with transparent glasses and white plates). To manipulate such objects under visual control, it is therefore of critical importance in full-fledged visual representation to provide reconstructions that complete the 3D structure of the relevant surfaces of objects in the visual world.

Surfaces as a mid-level invariant in visual encoding

Rather than relying on the object templates typical of cognitive investigations, a more fruitful approach to the issue of 3D object structure is to focus the analysis on *mid-level* invariants to the object structure, such as surfaces, symmetry, rigidity, texture invariants, or surface reflectance properties. Each of these cues is invariant under transformations of 3D pose, viewpoint, illumination level, haze contrast, and other variations of environmental conditions. Various computational analyses have incorporated such invariants in their object-recognition schemes, but a neglected aspect of mid-level vision is the 3D surface structure that is an inescapable property of objects in the world.

Surfaces are a key property of our interaction with objects in the world. It is very unusual to experience objects, either tactilely or visually, except through their surfaces. Even transparent objects are experienced through their surfaces, with the material between the surfaces being invisible by virtue of the transparency. Only translucent objects are experienced in an interior manner, as the light passes through them to illuminate the density of the material. Developing a means of representing the proliferation of surfaces before us is, therefore, a key stage in the processing of objects.

The various 3D surface cues such as luminance shading, linear perspective, aspect ratio of square objects, and texture gradient can each specify the slant of a planar surface. Zimmerman, Legge, and Cavanagh (1995), for example, performed experiments to measure the accuracy of surface slant from judgments of the relative lengths of a pair of orthogonal lines embedded in one surface of a full visual scene. Slant judgments were accurate to within 3° for all three cue types, with no evidence of the recession to the frontal plane expected if the pictorial surface was contaminating the estimations. Depth estimates of disconnected surfaces were, however, strongly compressed. Such results emphasize the key role of 3D surface reconstruction in human depth estimation.

Surface representation through the attentional shroud

A more vivid representation of the depth reconstruction process is to envisage it as an "attentional shroud" (Tyler & Kontsevich, 1995; Tyler, 2005; Huang, Chen, & Tyler, 2012), wrapping the dense locus of activated disparity detectors as a cloth wraps a structured object. The concept of the attentional shroud can also be thought of as a mechanism that acts like a soap film in minimizing the curvature of the perceived depth surface consistent with the available disparity information. In the original paper, an ambiguous disparity surface paradigm with attentional cueing was used to show that this surface reconstruction mechanism has the attentional property of adhering to the uniqueness constraint of operating on only one surface at a time. To probe the nature of such surface reconstruction for different depth cues, we may measure the localization of objects defined by multiple visual modalities (such as luminance and disparity).

Likova and Tyler (2003) addressed the unitary depth map hypothesis of object localization by using a sparsely sampled image of a Gaussian bulge. The sparse sampling dramatically degraded localization accuracy based on luminance cues, but caused no degradation when based on binocular disparity cues. This dramatic difference suggests that depth surface reconstruction is the key process in the accuracy of the localization process. Furthermore, when depth was nulled by opposing the luminance and disparity cues, localizing the "object" (the Gaussian bulge) became impossible. Only an interpolation mechanism operating at the level of generic *depth* representation can account for the data. Evidently, the full specification of objects in general requires extensive interpolation to take place. The dominance of a depth representation in the performance of such tasks indicates that the depth information is not just an overlay to the 2D sketch of the positional information. Instead, a full 3D depth reconstruction of the surfaces in the scene must be completed by the human visual system before the position of the object is known.

Downloaded from jov.arvojournals.org on 10/20/2019

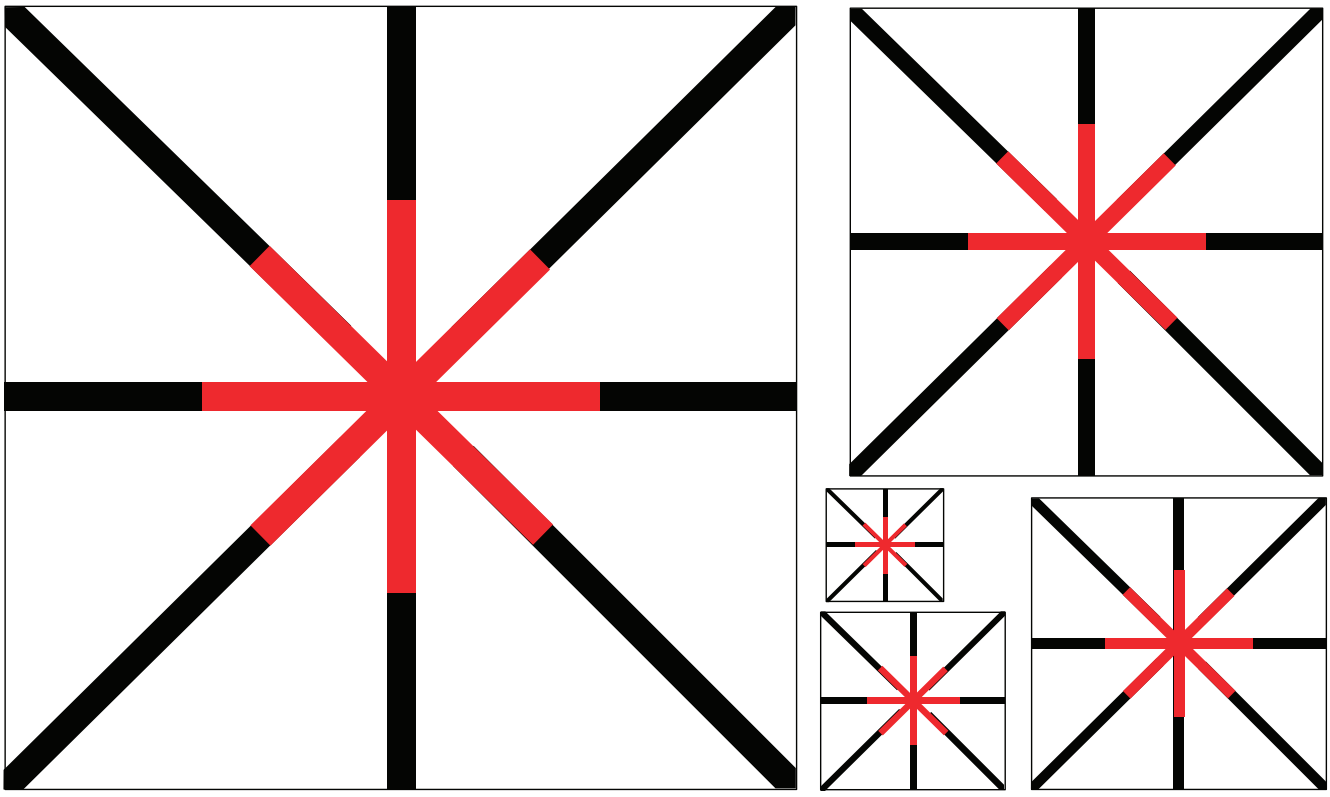

Figure 1. The neon color spreading effect, shown at several different scales. This configuration is derived from one developed by Bach (2002) (http://www.michaelbach.de/ot/col_neon/index.html).

Rationale

To address the question of depth reconstruction in the sparse cue situation, we designed a depth image that is a 3D version of the neon color spreading effect (Figure 1), in which the empty wedge regions between the spokes appear to take on a fainter version of the color of the central segments of the spokes (Varin, 1971; van Tuijl, 1975; Bressan, Mingolla, Spillmann, & Watanabe, 1997). It is generally understood that the colored region has the form of a circular disk (e.g., Gove et al., 1995), although the sharpness of the edges does not appear to have been quantified and our close observation suggests that the edge region is not well defined and cannot be definitively specified as having the sharp edge of a uniform disk.

The 3D version of the illusion is depicted in Figure 2, in which the color difference in the central region of the spokes is replaced by a binocular disparity difference, generating the impression that this region is recessed in depth. In order to provide equal disparity cues for all spoke angles, the spokes are made of a random-dot texture rather than a uniform black. Note that the white wedges between the spokes appear to have the same depth as the spokes, even though they carry no binocular disparity information per se. In this sense, the visual system is able to use the disparity information in the spoke regions to fill in the perceived depth in the inter-spoke regions. It is noteworthy that the interspoke depth appears continuous with the depth of the spokes, implying 100% strength of the depth filling-in effect, in contrast to the much weaker filling-in of the neon color spreading effect.

Depth spreading through empty space was reported by Julesz (1971, figure 7.7.1), in the form of a white stripe through his classic random-dot stereogram of a depth square. He described the perceived depth as forming a "crisp contour," and showed that a more complex type of interpolation in the form of a white stripe through a random-dot stereogram of a diamond shape (Julesz, 1971, figure 7.7.2) did not give a clear interpolated border. He did not, however, quantify the sharpness of the border, and did not assess the form of a curved border interpolation. Gove et al. (1995) implemented a computational model of the neon color spreading through the LAMINART model of cortical interactions (Grossberg, 1999) that generated a linear contour interpolation. Presumably a more elaborated process would be needed to generate a fully circular completion boundary, either one that was anisotropic, with extended smoothing *along* the contour while allowing propagation of the sharp boundary *across* the contour, or one that incorporates relaxation to simple higher-order figures such as circles.

Downloaded from jov.arvojournals.org on 10/20/2019

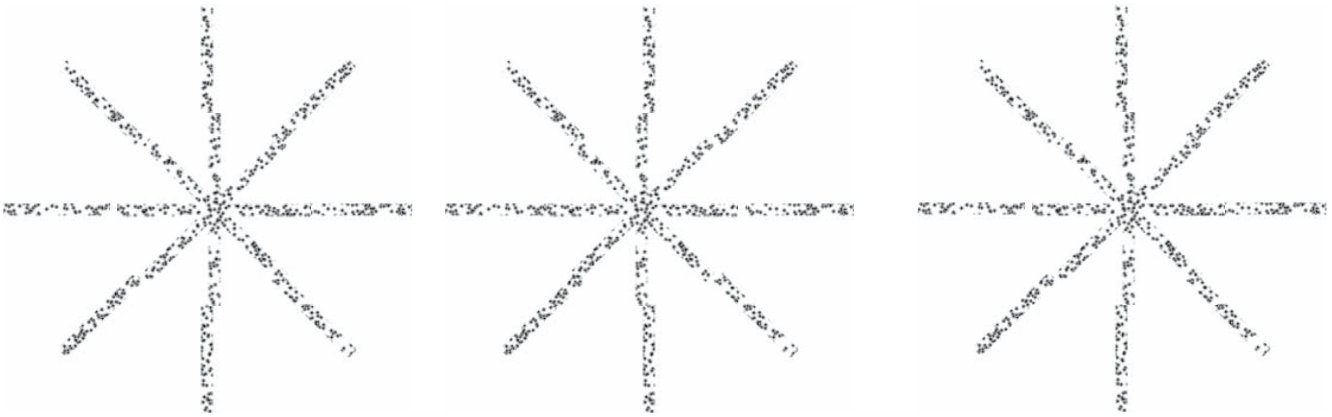

Figure 2. The depth spreading effect, a stereoscopic depth version of the neon color spreading effect. Free fusion across a pair of these images will reveal a pair of stereoscopic images, one with near and the other with far disparity in the central region of spokes, which appears to complete across the white spaces to form a raised or recessed disk, respectively.

These possibilities for the process involved in the completion of the disk, in either the color or depth version of the filling-in, can be formalized into the four hypotheses depicted in Figure 3:

1. Low-level local depth propagation: local diffusion hypothesis that the filling-in proceeds locally in all directions towards the edge in a form predicted by a local 2D diffusion function to form depth gradients in the wedge regions (Figure 3B). (Details of the 2D diffusion hypothesis are provided in Methods.)
2. Mid-level depth-contour propagation: (A) Linear depth-contour *extrapolation* hypothesis that the filling-in propagates the sharp-edged character of the defined disparity terminator edges linearly across the undefined region, such that the shape of the filled-in region extends along the orientation of the terminators to take an octagonal form (Figure 3C, green contour). (B) Linear depth-contour *interpolation* hypothesis that the filling-in propagates the sharp-edged character of the defined disparity terminations linearly across the undefined region, such that the shape of the filled in region takes a truncated octagonal form (Figure 3C, black boundary).
3. High-level figural depth propagation: figural depth-contour interpolation hypothesis that the filled-in area conforms to the disciform figure of a circular disk with a sharp edge, filled-in uniformly up to the edge and falling immediately to the surround level beyond the edge (Figure 3D).

This study is designed to measure the form of the depth profile in the interpolation region in order to gain insight into the neural mechanisms involved in the interpolation process.

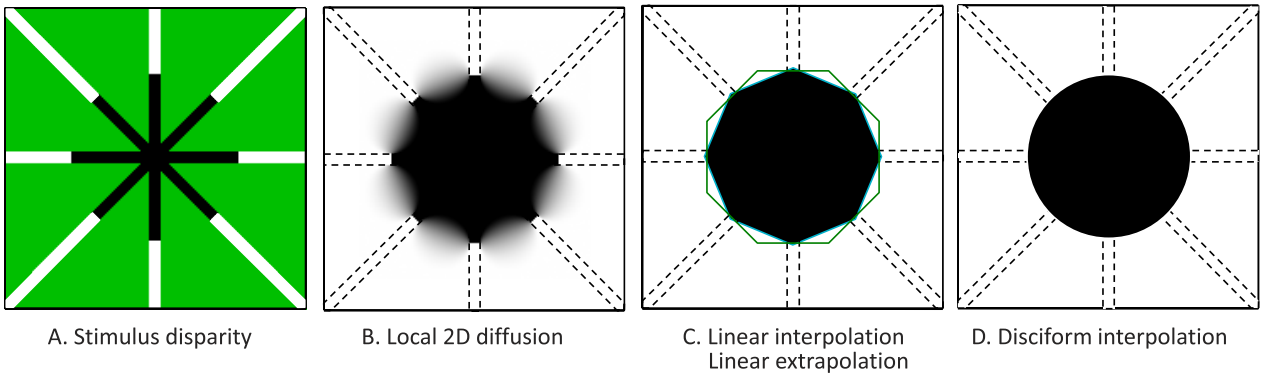

Figure 3. Depiction of four forms of depth surface interpolation. (A) Stimulus disparity arrangement (black to white: far to near disparities, green: undefined). (B) Depth interpolation by diffusion from the spokes, forming depth gradients in the wedge regions. (Black to white: far to near perceived depth, dashed lines: region of defined disparity). (C) Two forms of depth propagation with straight contours in forming a sharp-edged octagon between the spokes. The green outline depicts linear extrapolation of the depth contours in the spokes; the cyan outline around the black octagon indicates linear interpolation between the endpoints of the depth contours. (D) Depth propagation along curved contours in the form of a sharp-edged disk.

Downloaded from jov.arvojournals.org on 10/20/2019

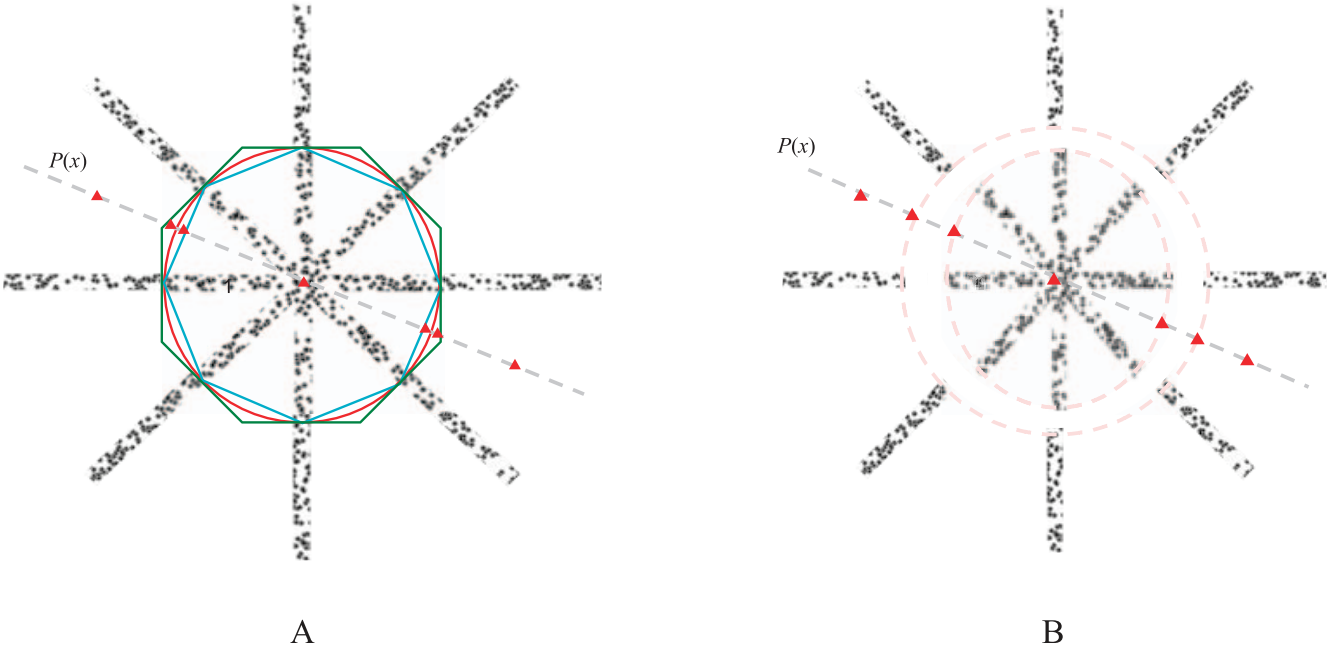

Figure 4. (A) Diagram of nonslanted test image with red triangles indicating tested locations and blue, green, and red lines delineating, respectively, the perceived depth boundaries predicted by the linear interpolation, linear extrapolation, and disciform hypotheses. (B) Diagram of the slanted test image with red triangles indicating tested locations. The white gap was included to minimize the effect of the vernier shifts of the inducing spokes in each eye's image.

Methods

Subjects and experimental setup

Four naïve subjects, all with normal or corrected-to-normal vision and able to free-fuse the adjacent pair of (nonanaglyphic) binocular stimuli comfortably, each performed binocular observations at a viewing distance of 86 cm of the stereoscopic patterns presented on a computer screen with a horizontal resolution of 1024 pixels. The study was approved by the UMDNJ IRB, and all subjects gave written informed consent.

Stimuli

Each stimulus in any given binocular pair had eight disconnected, equidistant radial inducing bars composed of a nonoverlapping random dot pattern at zero disparity, serving as the "outer spokes" (Figure 2). Eight equidistant radial "inner spokes," joined at the center, were of equal length and composition, with a disparity offset from that of the outer spokes. In the frontoparallel stimulus set, these inner spokes were given a uniform disparity of −23 arcmin when viewed binocularly, with the random dots forming a continuous pattern through the region of the disparity edges within the spokes. In the horizontally slanted configuration, the spokes were arranged to lie in a slanted plane with the left edge at a disparity of −28 arcmin and the right edge at a disparity of +11 arcmin when viewed binocularly. To control for any interference from the monocular vernier-type shifts present in the uniform disparity condition (Figure 2), the slanted condition also included a gap between the inner and outer spokes (Figure 4B).

Procedures

The stimuli were displayed to subjects in a series of nine pairs of frontal and slanted stimuli for each radial position tested, to provide nine steps of 2 arcmin of disparity around the fixed disparity of the disk (Figure 4). A binocular blue dot (Figure 5) was placed at varying disparities in empty space along the imaginary oblique line through the center of the gaps. A red comparison line was shifted to align with the blue dot in the right eye's image. A cyan disparity line was aligned with the blue dot in the left eye's image, so that upon free-fusion, the two lines would form a Nonius pair to ensure binocular alignment. (Note that, although Tsirlin, Wilcox & Allison, 2011, reported that a binocular depth probe could influence the percept of depth in an illusory interpolated region, there is no suggestion that it could generate the kinds of sharp depth edge implied by the hypotheses depicted in Figure 3C and D, and validated in the Results below).

Downloaded from jov.arvojournals.org on 10/20/2019

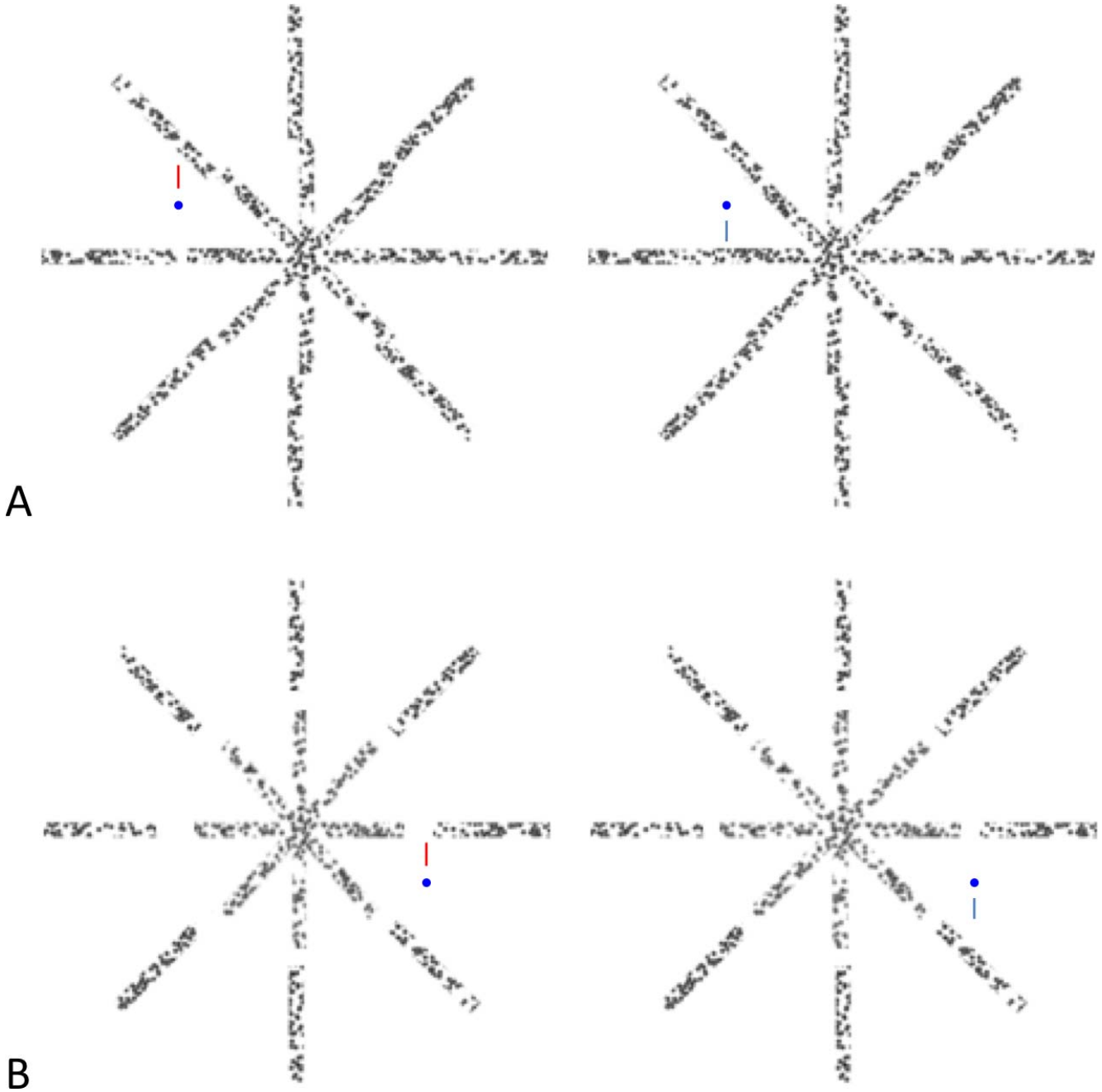

Figure 5. (A) Example of the nonslanted stimulus configuration that subjects were asked to free-fuse to test the sharpness of the depth boundaries. Blue test dots are set at the upper radial position just outside the circular boundary of the inner disc. The red fiducial line is aligned with the blue dots in the right eye, and the cyan line in the left eye, to form a Nonius pair. (B) Example of the slanted stimulus configuration, with the blue test dot and Nonius lines set at the lower radial position just outside the circular boundary.

For each radial position, subjects stepped the binocular blue dot through the nine available disparity levels in sequence to find a perceived depth match, with the capability of going back and forth along the sequence to refine their choice when they were close to their perceived match. Both sets of frontoparallel and horizontally slanted depth structures were run for each of two experiments to evaluate the perceived depth in the intervening white spaces and the sharpness of depth boundaries.

Method of adjustment

Perceived depth of the white space interpolated from the inducing spokes was first quantified to test the observation that depth did spread into the empty space

Downloaded from jov.arvojournals.org on 10/20/2019

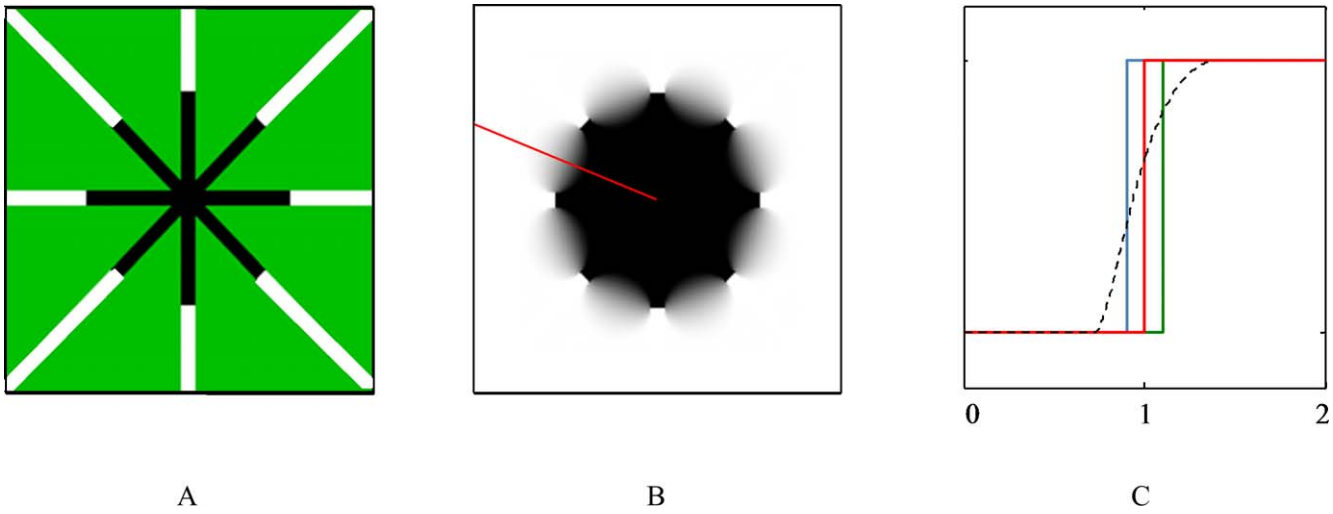

Figure 6. Perceived depth predictions of the four hypotheses. (A) Illustration of the disparity configuration of the stimulus of Figure 4A, with white and black designating the two disparities and green designating indeterminate white regions. (B) Surface interpolation in radial distance ($\theta = \pi/8$) calculated using MATLAB's TriScatteredInterp function. The red line in B indicates the radial position of the depth of the interpolated surface in panel C. (C) Hypothetical perceived depth functions. The red line in panel C portrays the hypothesis of a sharp border on a curved (disciform) contour at the same radial distance as the disparity edges, in contrast to the linear extrapolation of the disparity contours (green line), linear interpolation of the sharp border between the disparity contour endpoints (cyan line), and the surface interpolation hypothesis of panel B (dashed line).

of the gaps between spokes. Subjects selected the stimulus in which the disparity of the blue dot appeared to lie at a depth that matched their perceived depth of that space (Figure 4). This process was repeated for each of seven positions along the radial line per set of frontoparallel and horizontally slanted inner disks: in the empty space between the outer spokes, depth boundary of the outer spokes, depth boundary of the inner spokes, middle of the illusory depth structure, and three corresponding positions on the opposite side of the structure (Figure 4).

Radial position was the distance in pixels from the outer radial boundary of the outer spokes to the point of interest. This was measured as viewing angle in arcmin by multiplying the angle in radians by 60 × (180°/π). Disparity was measured in pixels and converted into arcmin, as shown in Equation 1:

$$\begin{aligned}\text{disparity} &= \text{pixel shift} \\ &\quad \times \frac{\text{screen width}}{\text{horizontal resolution} \times \text{viewing distance}} \\ &= \text{pixel shift} \times 34\text{cm} \frac{180^\circ/\pi}{1024\,\text{px} \times 86\,\text{cm}} \\ &\quad \times 60\,\text{arcmin} \end{aligned} \quad (1)$$

Predictions

Depth interpolations in the empty white space were calculated based on the actual disparity of our test points and the local depth interpolation hypothesis was implemented using the TriScatteredInterp function in MATLAB, while the predictions of the depth-contour propagation hypotheses were evaluated geometrically (Figure 6). The disciform (figural depth propagation) hypothesis predicts that a sharp depth contour in the empty white space is at the same radial distance as the disparity edge in the spokes (i.e., 0.5). (Only the disciform hypothesis was evaluated for the slanted configuration because the gaps between the inner and outer spokes made the others difficult to implement). Radial distance from the center was specified in arcmin (Figure 6C). Depth value outputs with the same range were rescaled to the respective actual disparities of the outer boundary of the inner illusory disk and the inner boundary of the outer spokes.

Sharpness of depth boundaries

The sharpness of the depth boundaries delineated by the inner and outer spokes was also quantified by determining their radial positions, using a similar setup as in the previous experiment. Here, the blue binocular test dot was aligned with the radial position of the circular boundary of either the inner spokes or that of the outer spokes. Subjects selected the stimulus in which the position of the blue dot matched their perceived edge of the boundary in question. They repeated this process four times along the same radial line used to test perceived depth of empty space: once for the circular depth boundary of the outer spokes,

Downloaded from jov.arvojournals.org on 10/20/2019

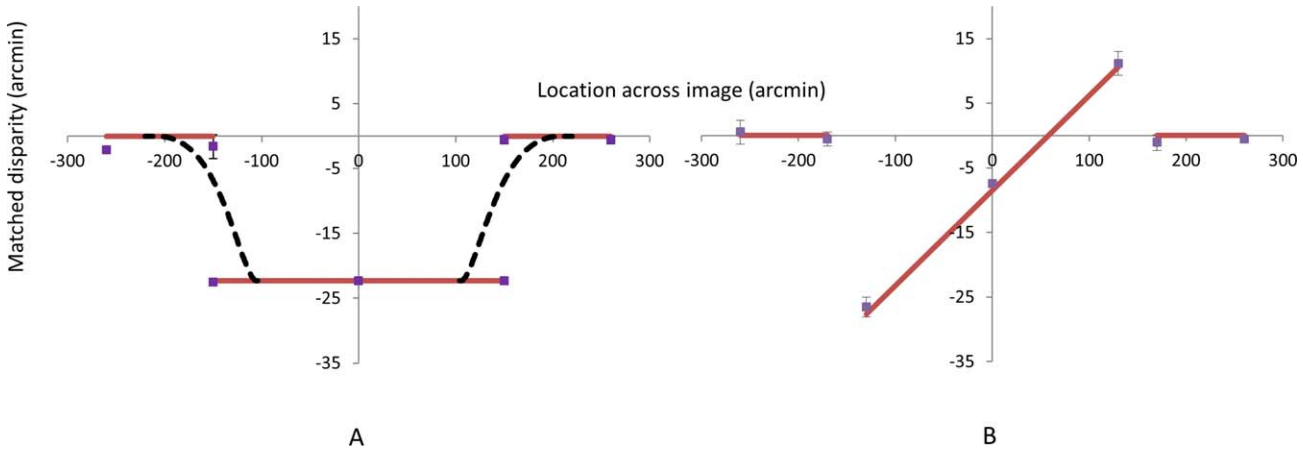

Figure 7. Graphical representation of the results and predictions for each set of frontoparallel (A) and horizontally slanted (B) depth structures. Actual results (purple dots), depth interpolations (red solid line), and surface interpolation (black dashed line); $n = 4$.

once for the boundary of the inner spokes, and finally once each at the two corresponding positions on the opposite half of the structure. Pixel distances were converted into arcmin for further statistical analysis using Equation 1. The distance from the center of the inner spokes to the midpoint of any edge of an interpolated octagonal depth boundary was determined as shown in Equation 2, where r is the length of any inner spoke and the radius of an interpolated circular boundary:

$$\cos\frac{45^\circ}{2}\times r. \quad (2)$$

Thus, the octagonal extrapolation predicts a distance ratio of 1.082 times that of the circular interpolation while the octagonal interpolation predicts a distance ratio of 0.924.

Results

The actual depth of the white space adhered closely to the depth interpolated from the disparities in the inducing spokes for both the frontoparallel and horizontally slanted disk stimuli, in accordance with the observation that perceived depth spreads into empty space (Figure 7; Table 1). The form of the perceived depth, however, failed to match to the graded surface prediction of the local depth propagation hypothesis (Figure 7A). The sharp borders of the linear depth-contour extrapolation and interpolation hypotheses are also significantly different from the actual results for the frontoparallel disk sets ($p < 0.05$, corrected for multiple applications in Table 1, $n = 4$). The predictions of the disciform hypothesis were not

A. Perceived depth predictions and results for the frontoparallel figure							B. Perceived depth predictions and results for the slanted figure			
	Depth predictions									
Radial position	Linear interpolation	Linear extrapolation	Disciform hypothesis	Local depth propagation	Actual results	SD	Radial position	Disciform hypothesis	Actual results	SD
−260	0	0	0	0	−2.1	0.5	−260	0	0.5	1.9
−150	0	−23*	0	−5.6	−1.6	2	−170	0	−0.5	1.2
−150	0*	−23	−23	−7*	−22.6	0.7	−130	−28	−26.5	1.6
0	−23	−23	−23	−23	−22.3	0.5	0	−7	−7.4	0.5
150	0*	−23	−23	−7*	−22.3	0.5	130	11	11.1	1.9
150	0	−23*	0	−5.6	−0.5	0.8	170	0	−1.1	1.3
260	0	0	0	0	−0.5	0.8	260	0	−0.5	0.8

Table 1. Depth predictions for several hypothetical forms of depth spreading compared with results from naïve subjects ($n = 4$) with standard deviation (SD) at each of seven radial positions (bolded) for the frontoparallel (A) and horizontally slanted (B) depth structures at radial positions designed to discriminate among the hypotheses of Figure 6. All measurements in arcmin of the visual angle of subtense. Asterisks in A (*) indicate significant differences ($p < 0.05$, corrected for multiple applications across the tables, $n = 4$) between observed and predicted depth values for a particular model (see Figure 7A). Values in B were only evaluated for the disciform hypothesis, for which the data showed no significant differences from the predicted depths (see Figure 7B).

Downloaded from jov.arvojournals.org on 10/20/2019

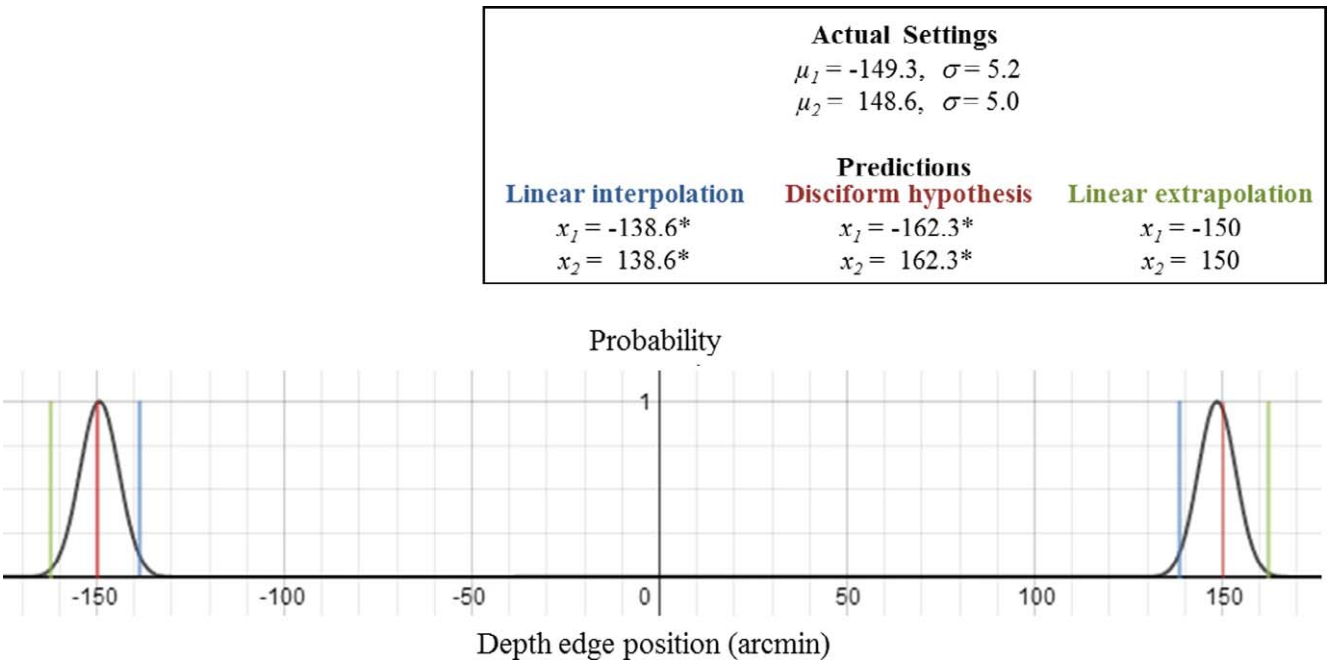

Figure 8. Results of the position estimation of the perceived border for the frontoparallel disk. Gaussian curve represents the mean and standard deviations (σ) of the results. Vertical bars represent the predicted locations of the interpolated (blue), extrapolated (green), and disciform (red) hypotheses as to the position of the perceived depth border. Asterisks (*) in the inset panel indicate predicted values for which the actual depth differed significantly ($p < 0.05$, $n = 4$).

significantly different from our results for either the frontoparallel or slanted disc configurations (Figure 7A, B).

Sharpness of depth boundaries

From the data, Gaussian curves depict the error distributions at the inner depth boundaries to quantify sharpness and boundary position relative to the three hypotheses (Figure 7). The data from boundary sharpness are consistent with the prediction of a circular and not either of the octagonal depth structures. Gaussian distributions of lateral shift data for the frontoparallel set indicate that the location of the border is remarkably precise (Figure 7), with σ of the order of 5 arcmin, or only about 2% of the radius of the disk. At a criterion of $p < 0.05$, the disciform hypothesis is the only one that is not significantly different from observed values of depth, while being significantly different from the two locations predicted by the other two hypotheses (see Figure 8).

Discussion

This study had two main outcomes. First, there is perceptual completion of the disparity border specified in the spokes across the empty white space between the spokes, forming a perceived depth border as sharp as could be measured by our probe technique. This form of diffusion is anisotropic in that the border is propagated lengthwise between the spokes rather than spreading out isotropically from the points where the disparity information stops. To quantify the alternative, we formalized the isotropic diffusion concept as a local diffusion process of the 3D surface information provided by the disparity in the spoke regions, predicting the 2D surface function depicted in Figure 3B and graphed as a 1D function through the line halfway between the spokes in Figures 4 and 6B. This diffusive form of interpolation is strongly repudiated by the quantitative results, which show that the perceived border is much sharper than the isotropic 2D diffusion prediction, requiring some form of anisotropic diffusion in which the border information is propagated only along a 1D linear extension of its length and not in any other directions.

This raises the issue of the second outcome, which shows what form of line the 1D propagation takes. The two forms depicted in Figure 3C are incompatible with the results in Figure 7, which shows that the location of the sharp depth border conforms only to the fourth hypothesis of anisotropic curvilinear extrapolation of the disparity edges to form a homogeneous flat disk in depth. Whereas the isotropic 2D propagation and the anisotropic 1D propagation hypotheses are compatible with local processes, the anisotropic disciform propagation hypothesis seems to be compatible only with a

Downloaded from jov.arvojournals.org on 10/20/2019

global process that conforms the interpolated depth contours into a figural Gestalt with global simplicity. Admittedly, a disk shape is a very familiar form of global simplicity, but it is nevertheless one that has to be applied to the perceived depth structure, which is itself a higher-order construct relative to the local luminance edges that are considered the usual constituents of form processing. The disciform outcome of the interpolation process is therefore a substantial challenge to an understanding of the underlying perceptual mechanisms in terms of local neural processes.

Explanatory mechanisms for the anisotropic disciform propagation hypothesis

Two mechanisms that could be invoked to account for the disciform interpolation of the depth contour are Bayesian priors and transformational invariants. Under the transformational invariant explanation, potential interpolation options would be examined for simplicity under image transformations. One such transformation might be rotation around the line of sight, such that when fixated centrally, a disk has self-similarity (simplicity) under rotation. Thus, the roundish recess generated by a proto-interpolation process might be guided toward the form of a disk because of its figural simplicity under rotation.

The other possible mechanism is the Bayesian prior, which says that, regardless of figural properties, the interpolation is governed by the probability of having encountered recesses of various shapes in the past. The highest probability shapes then tend to govern the shape that the present recess is perceived to have. It should be noted that the Bayesian hypothesis is entangled with the transformational invariance hypothesis in two respects. One is that the recesses encountered in the past may themselves be biased toward a rotationally invariant sample because many of them may have been made by a rotationally invariant tool such as a drill (or a molding derived from such a tool). In this respect, our modern visual systems brought up in a "carpentered" artificial environment might be very different from those brought up in a primitive cave environment, in which accurately disciform recesses would be extremely rare. The other is that, conversely, the average probability function for recess shapes would tend toward rotational invariance even though the individual examples encountered might not themselves have been circular, simply by regression toward the mean. On this argument, the Bayesian prior might tend toward simplicity principles in the average even if they were not manifested in individual examples. In general, then, it is likely to be difficult to disentangle the mechanism enforcing the simplicity interpretation unless studies are done to specifically distort the near-term Bayesian prior and determine its degree of applicability in particular cases.

Keywords: disparity, depth perception, 3D surface and shape perception, filling-in, propagation, Bayesian prior

Acknowledgments

This paper was supported by AFOSR grant #FA9550-09-1-0678. The authors thank Spero Nicholas for implementing the local depth propagation hypothesis (Figures 3, 5). XL thanks the New Jersey Medical School Office of Research and Sponsored Programs for support.

Commercial relationships: none.
Corresponding author: Christopher W. Tyler.
Email: cwt@ski.org.
Address: Smith-Kettlewell Eye Research Institute, San Francisco, CA, USA.

References

Bach, M. (2002). Neon color spreading. *100 Visual Phenomena & Optical Illusions Website,* Internet site: http://michaelbach.de/ot/col_neon/index.html (Accessed July 18, 2013).

Bressan, P., Mingolla, E., Spillmann, L., & Watanabe, T. (1997). Neon color spreading: A review. *Perception, 26,* 1353–1366.

Gove, A., Grossberg, S., & Mingolla, E. (1995). Brightness perception, illusory contours, and corticogeniculate feedback. *Visual Neuroscience, 12*(6), 1027–1052.

Grossberg, S. (1999). How does the visual cortex work? Learning, attention and grouping by the laminar circuits of the visual cortex. *Spatial Vision, 12,* 163–186.

Grossberg, S., Kuhlmann, L., & Mingolla, E. (2007). A neural model of 3D shape-from-texture: Multiple-scale filtering, boundary grouping, and surface filling-in. *Vision Res, 47,* 634–672.

Huang, P. C., Chen, C. C., & Tyler, C. W. (2012). Collinear facilitation over space and depth. *Journal of Vision, 12*(2):20, http://www.journalofvision.org/content/12/2/20, doi:10.1167/12.2.20. [PubMed] [Article]

Huang, T., & Russell, S. (1998). Object identification: A Bayesian analysis with application to surveillance. *Artificial Intelligence, 103,* 77–93.

Downloaded from jov.arvojournals.org on 10/20/2019

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

Likova, L. T., & Tyler, C. W. (2003). Peak localization of sparsely sampled luminance patterns is based on interpolated 3D object representations. *Vision Research, 43,* 2649–2657.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: W. H. Freeman and Company.

Moghaddam, B. (2001). Principal manifolds and probabilistic subspaces for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24,* 780–788.

Rue, H., & Hurn, M. A. (1999). Bayesian object identification. *Biometrika, 86,* 649–660.

Stormont, D. P. (2007). An online Bayesian classifier for object identification. *IEEE International Workshop on Safety, Security and Rescue Robotics, 2007,* 1–5.

Tsirlin, I., Wilcox, L.M. & Allison, R.S. (2011). Disparity biasing in depth from monocular occlusions. *Vision Research, 51,* 1699–1711.

Tyler, C. W. (2005). Spatial form as inherently three-dimensional. In M. R. M. Jenkin & L. R. Harris (Eds.), *Seeing spatial form* (pp. 67–88). Oxford, UK: Oxford University Press.

Tyler, C. W., & Kontsevich, L. L. (1995). Mechanisms of stereoscopic processing: Stereoattention and surface perception in depth reconstruction. *Perception, 24,* 127–153.

van Tuijl, H. F. J. M. (1975). A new visual illusion: Neonlike color spreading and complementary color induction between subjective contours. *Acta Psychologica, 39,* 441–445.

Varin, D. (1971). Fenomeni di contrasto e diffusione cromatica nell organizzazione spaziale del campo percettivo. *Rivista di Psicologia, 65,* 101–128.

Zimmerman, G. L., Legge, G. E., & Cavanagh, P. (1995). Pictorial depth cues: A new slant. *Journal of the Optical Society of America A, 12,* 17–26.

Downloaded from jov.arvojournals.org on 10/20/2019