

# Optimal combination of illusory and luminance-defined 3-D surfaces: A role for ambiguity

**Brittney Hartle**

Department of Psychology and Centre for Vision Research,  
York University,  
Toronto, Canada



**Laurie M. Wilcox**

Department of Psychology and Centre for Vision Research,  
York University,  
Toronto, Canada



**Richard F. Murray**

Department of Psychology and Centre for Vision Research,  
York University,  
Toronto, Canada



The shape of the illusory surface in stereoscopic Kanizsa figures is determined by the interpolation of depth from the luminance edges of adjacent inducing elements. Despite ambiguity in the position of illusory boundaries, observers reliably perceive a coherent three-dimensional (3-D) surface. However, this ambiguity may contribute additional uncertainty to the depth percept beyond what is expected from measurement noise alone. We evaluated the intrinsic ambiguity of illusory boundaries by using a cue-combination paradigm to measure the reliability of depth percepts elicited by stereoscopic illusory surfaces. We assessed the accuracy and precision of depth percepts using 3-D Kanizsa figures relative to luminance-defined surfaces. The location of the surface peak was defined by illusory boundaries, luminance-defined edges, or both. Accuracy and precision were assessed using a depth-discrimination paradigm. A maximum likelihood linear cue combination model was used to evaluate the relative contribution of illusory and luminance-defined signals to the perceived depth of the combined surface. Our analysis showed that the standard deviation of depth estimates was consistent with an optimal cue combination model, but the points of subjective equality indicated that observers consistently underweighted the contribution of illusory boundaries. This systematic underweighting may reflect a combination rule that attributes additional intrinsic ambiguity to the location of the illusory boundary. Although previous studies show that illusory and luminance-defined contours share many perceptual similarities, our model suggests that ambiguity plays a larger role in the perceptual

representation of illusory contours than of luminance-defined contours.

## Introduction

Illusory contours occur when boundaries are perceived in an image in the absence of a corresponding luminance gradient. Although illusory contours can be created under a wide range of conditions, the most well-known stimuli are those introduced by Kanizsa (1955). In Kanizsa figures, observers see a complete square even though the shape is specified only by the relative position of a set of inducing elements. In Figure 1, the central region of the two-dimensional (2-D) Kanizsa figure is perceived as being closer to the observer than the inducing elements and brighter than the background (Bradley & Dumais, 1984; Coren, 1972; Coren & Porac, 1983). Although lightness illusions do not always occur with Kanizsa configurations (Day, 1987; Dresch, Lorenceau, & Bonnet, 1990; N. Kogo, Strecha, Gool, & Wagemans, 2010), when they do occur, observers typically perceive illusory contours (He & Ooi, 1998; Prazdny, 1983).

In Kanizsa figures, the perceived depth order derived from occlusion plays a key role in creating illusory contours (Coren, 1972; Gillam & Nakayama, 2002; Kellman & Shipley, 1991; Rubin, 2001). However, in these 2-D images, depth information is qualitative. The apparent depth and shape of illusory surfaces are dramatically enhanced when Kanizsa figures are viewed stereoscopically (Carman & Welch, 1992; Ramachan-

Citation: Hartle, B., Wilcox, L. M., & Murray, R. F. (2018). Optimal combination of illusory and luminance-defined 3-D surfaces: A role for ambiguity. *Journal of Vision*, 18(4):14, 1–15, <https://doi.org/10.1167/18.4.14>.

<https://doi.org/10.1167/18.4.14>

Received October 12, 2017; published April 13, 2018

ISSN 1534-7362 Copyright 2018 The Authors



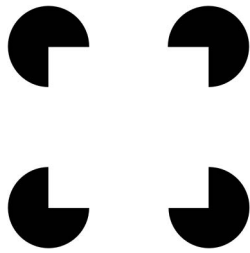


Figure 1. A Kanizsa square with four high contrast-inducing elements.

dran, 1986; Vreven & Welch, 2001). This is illustrated in Figure 2 in which the vertical edges of the inducers in the 3-D Kanizsa figure are rendered with binocular disparity, consistent with the presence of a curved white foreground surface.

Three-dimensional Kanizsa figures can be thought of as examples of disruptive coloration camouflage in which a large portion of an object's boundary has the same luminance as the background (Endler, 2006). To disambiguate a coherent surface, the visual system must combine information about the location of these camouflaged regions from the luminance-defined inducers and interpolated illusory edges. One difference between luminance-defined and illusory regions is that the location of the former is clearly defined by a change in the intensity of the physical stimulus, and the position of the latter is ambiguous and must be inferred.

There is considerable evidence that illusory and luminance-defined contours are integrated and even processed similarly by the visual system. It has been shown that illusory boundaries are perceptually similar to luminance edges in that they (a) share neural architecture that processes luminance-defined edges (Larsson et al., 1999; von der Heydt, Peterhans, & Baumgartner, 1984), (b) share similar perceptual illusions (Paradiso, Shimojo, & Nakayama, 1989; Smith & Over, 1979), and (c) exhibit rivalry with luminance edges (Fahle & Palm, 1991). More specifically, a number of studies have used 2-D Kanizsa figures to provide evidence of interaction between illusory and luminance-defined boundaries. For example, the inclusion of an illusory contour can improve the detectability of luminance-defined contours near contrast threshold using collinear facilitation (Dresp & Bonnet, 1995; Wehrhahn & Dresp, 1998). Others have shown that luminance contours superimposed on illusory contours can interfere with detection thresholds at suprathreshold contrasts (Dillenburger & Roe, 2010). Interestingly, contours are often completed irrespective of the stimulus attribute (i.e., luminance, temporal, or binocular disparity) used to create illusory contours (Poom, 2001). Clearly, illusory and luminance-defined regions interact to determine the per-

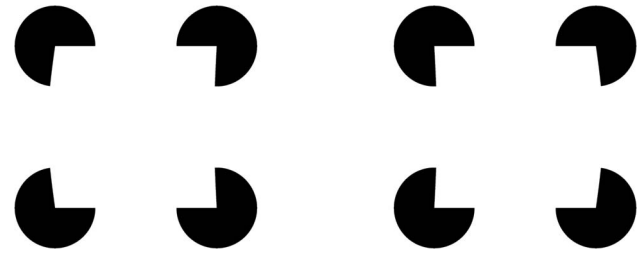


Figure 2. A stereo pair of a high-contrast Kanizsa figure. When cross-fused, the disparity at the inducing edges generates a percept of a 3-D crossed-disparity illusory surface in the absence of luminance-defined features in the central region.

ceived 3-D shape of illusory figures. The development of a quantitative model of the integration of illusory and luminance-defined features provides a compelling opportunity to evaluate the role of ambiguity in depth cue combination.

In other domains, researchers have modeled depth cue integration using Bayesian decision theory (Maloney & Landy, 1989). The combination of depth cues is commonly modeled as maximum likelihood estimation (MLE); it is assumed that the noise associated with each estimate is independent and Gaussian, and all Bayesian priors are uniform and noninformative (Landy, Maloney, Johnston, & Young, 1995). In this case, the combined-cue estimate can be calculated as a simple average of the single-cue estimates, weighted by each cue's reliability, which is the inverse of the cue's variance (Cochran, 1937). The MLE approach has provided a principled framework for modeling the fusion of multiple depth cues (Hillis, Watt, Landy, & Banks, 2004). In this approach, optimal cue integration maximizes reliability (Ernst & Banks, 2002; Landy et al., 1995). If we have an unbiased depth cue  $Q_i$  with variance  $\sigma_i^2$  and a second, independent and unbiased depth cue  $Q_l$  with variance  $\sigma_l^2$ , then the optimal MLE depth estimate  $\hat{Q}_c$  based on these two cues is

$$\hat{Q}_c = w_i Q_i + w_l Q_l, \quad (1)$$

where

$$w_i = \frac{1/\sigma_i^2}{(1/\sigma_i^2 + 1/\sigma_l^2)}$$

$$w_l = \frac{1/\sigma_l^2}{(1/\sigma_i^2 + 1/\sigma_l^2)}.$$

These weights are proportional to the inverse of the variances of the cue distributions, so greater weight is placed on the more reliable cue. By combining information from several depth cues, the visual system can estimate depth with greater precision than it can by relying on any single cue (Ernst & Banks, 2002; Knill & Saunders, 2002; Landy et al., 1995).

Here we use a similar MLE cue-combination approach to evaluate observers' ability to combine depth information from illusory and luminance-defined contours. Usually, MLE cue-combination studies assess the combination of distinct depth cues, such as texture and binocular disparity (e.g., Hillis et al., 2004). There is no evidence that depth is computed separately from illusory and luminance-defined contours, and in fact, as discussed above, there is considerable evidence that both types of contours are processed in the same manner. However, the optimal cue combination model described above does not need to be considered a mechanistic model. As a normative model, it describes the best possible combination of depth from illusory and luminance-defined contours, thus providing insight into the relationship between underlying sources of depth information that are otherwise difficult to dissociate within stereoscopic illusory stimuli. We created 3-D Kanizsa-like surfaces that were defined by both illusory and luminance contours (i.e., combined surfaces). As outlined below, we use a normative model to assess each observer's performance in the combined surface condition relative to the best performance we could expect given their performance in the single-cue conditions.

## Experiment 1

The aim of our first experiment was to evaluate the perceived depth obtained from illusory and luminance-defined boundaries in stereoscopic curved surfaces to inform our MLE model. We assessed perceived depth estimates using variations of 3-D Kanizsa squares. Our stimulus conditions were designed to measure the depth defined by (a) illusory boundaries generated by stereoscopic inducing elements, (b) the luminance-defined disparity along the surface edge, and (c) a combined surface comprised of both.

### Observers

Seven observers (including authors BH and LW) participated in the study. Each observer's stereoacuity was assessed using the Randot™ test (Stereo Optical Co, Inc, Chicago, IL) to ensure that observers could detect depth from binocular disparities of at least 40 arcseconds. All observers had normal or corrected-to-normal vision. The same group of observers participated in both experiments reported here. One observer was not available to complete testing and was removed from the final analysis, resulting in a total of six observers. The experiments were approved by the York University Office of Research Ethics and followed the tenets of the Declaration of Helsinki.

### Stimuli: Rendering geometry

All Kanizsa figures were rendered as 3-D virtual objects in OpenGL using perspective projection with an asymmetric frustum configuration, using the Psychtoolbox package for MATLAB (MathWorks, Natick, MA; Brainard, 1997; Pelli, 1997). We configured OpenGL's projection matrix to match the viewing geometry in our modified Wheatstone mirror stereoscope with a viewing distance of 74 cm. To equate the test disparity for all observers, we set the lateral separation of the two projection frustums to equal each observer's interocular distance (IOD). Each 3-D Kanizsa figure was created by rendering four fronto-parallel black circles ( $0.81 \text{ cd/m}^2$ ), each with a diameter of  $0.8^\circ$  (1.1 cm) and a  $1.7^\circ$  (2.2 cm) separation between the centers of adjacent inducers. In stimulus conditions without illusory contours, the inducers faced outward (see Figure 4, Stimuli: Surface conditions). Here, the inducers were shifted diagonally by  $1.2^\circ$  (1.5 cm), so they abutted the corners of the occluding surface. The inducers were static with a  $90^\circ$  circular segment removed from the outermost edge and had zero disparity relative to the reference plane. The curvature of the surface edge was defined using a surface template that represented a half cycle of a sinusoid. The peak amplitude was calculated from the disparity at the peak and the observer's IOD using the conventional formula (see Howard & Rogers, 2012, pp. 152–154). All stimuli were rendered using the same curved surface template. Figure 3 illustrates the viewing geometry for Experiment 1.

### Stimuli: Surface conditions

Four stimulus conditions, (a) illusory, (b) low contrast, (c) combined, and (d) high contrast, were created for each observer (Figure 4). In the illusory condition, the luminance of the occluding surface was the same as the background (in all other conditions, this value differed). The low-contrast condition consisted of a stereoscopic surface with a low-contrast luminance value in the central region and rotated inducing elements. (See Procedure for details of luminance selection.) The combined condition consisted of a stereoscopic Kanizsa figure in which the central square and stimuli in the low-contrast condition had the same luminance. Lastly, the high-contrast control condition consisted of a black ( $0.81 \text{ cd/m}^2$ ) luminance-defined surface with rotated inducing elements. The surface was filled with black to create high-contrast edges, which are optimal for good stereoscopic acuity (McKee, 1983). All stimulus conditions had the same curvature along the surface edge, but in the illusory condition, the surface peak was camouflaged with

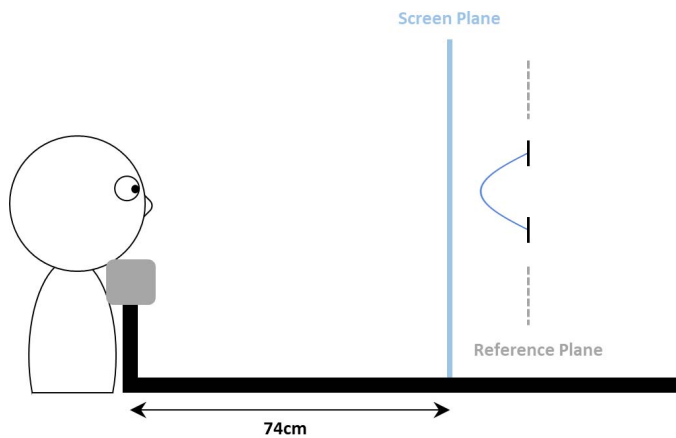


Figure 3. An illustration of the viewing geometry and stimulus as seen from the side (without the stereoscope mirrors and monitors). The reference plane had uncrossed disparity relative to the screen plane. The black inducers of the 3-D Kanizsa figure had zero disparity relative to the reference plane. The occluding surface extended in front of the reference plane, behind the screen plane, and in the shape of a half sinusoid.

respect to the background. Thus, the largest luminance-defined disparity in the illusory condition was the relative disparity at the tip of the inducing elements, and the largest disparity signal in the remaining surface conditions was at the surface peak.

All stimuli were presented at the center of the display on a gray background ( $50.3 \text{ cd/m}^2$ ) with an array of high-contrast ( $65.6 \text{ cd/m}^2$ ) outlined circles (radius of  $0.21^\circ$ ) above and below the central ( $5.2^\circ \times 10.5^\circ$ ) region. The arrangement of this pattern of circles was randomized on each trial, so they provided no consistent position cue but provided a strong fusion lock and reference plane. The stimulus and fusion field were presented at a standing uncrossed disparity of  $0.42^\circ$ . A circular disparity probe ( $22.6 \text{ cm/m}^2$ ) with a diameter of  $0.25^\circ$  was presented  $2.1^\circ$  to the left of the center of the screen. In preliminary testing, we assessed multiple lateral offsets ( $1.0^\circ$ ,  $1.5^\circ$ ,  $2.0^\circ$ , and  $2.5^\circ$ ) and determined that at displacements of  $2.0^\circ$  or greater there was no reliable influence of the probe on the interpolation of the surface.

### Apparatus

Stimuli were presented using the Psychtoolbox package (Brainard, 1997; Pelli, 1997) for MATLAB on a Mac OS X computer (Apple, Inc, Cupertino, CA). All stimuli were presented on a modified Wheatstone mirror stereoscope consisting of two LCD monitors (Dell U2412M; Dell Inc, Round Rock, TX) with a viewing distance of 74 cm and a fixed chin rest to

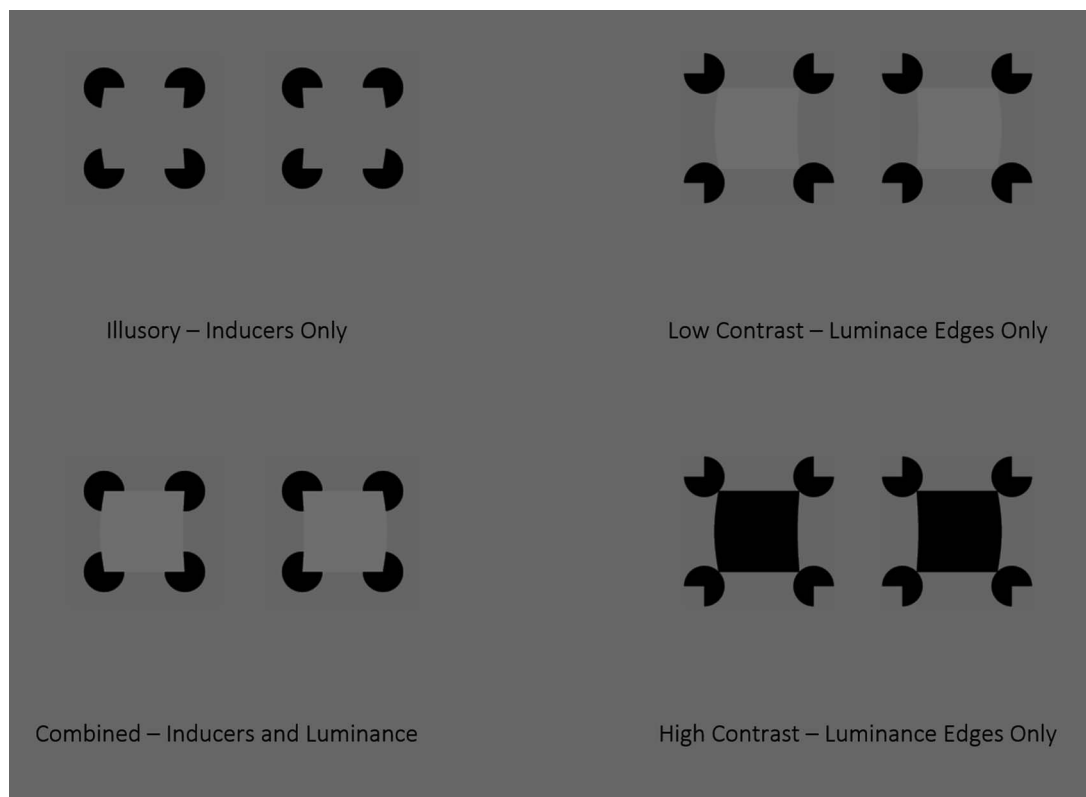


Figure 4. An illustration of the four stimulus conditions. All stereo pairs are arranged for cross-fusion. The feature available to support a disparity signal is indicated below each stimulus; see text for details.

maintain stable head position during testing. The monitor resolution was  $1,920 \times 1,200$  pixels with a refresh rate of 75 Hz. With these dimensions, each pixel subtended 1.26 arcminutes. Prior to testing, observers' IOD was measured using a Richter digital pupil distance meter, and all testing took place in a darkened room.

## Procedure

Previous lightness-matching experiments with Kanizsa figures tend to report contrasts that fall below the critical contrast range for disparity detection (Legge & Gu, 1989) with contrast thresholds as low as 1% to 3% (Li & Guo, 1995). Thus, prior to comparing perceived depth in the illusory and luminance-defined conditions, it was necessary to roughly equate perceived contrast in each condition while ensuring that the disparity of the low-contrast luminance surface was suprathreshold. To do so, we capitalized on the fact that reducing the contrast of a stimulus at a fixed disparity can make it appear more distant (Fry, Bridgman, & Ellerbrock, 1949; Rohaly & Wilson, 1999; Schor & Howarth, 1986). We used a two-interval, forced choice (2IFC) depth-discrimination paradigm in which observers compared the perceived depth of an illusory and luminance-defined surface with equivalent relative disparity along the inducing edge. The intensity of the luminance-defined surface was varied to find a luminance value at which the perceived depth of the illusory and luminance-defined peaks was roughly equivalent for each observer.

In all experiments, observers were asked to judge the perceived depth at the peak of the curved surface. To do so, we used a task in which observers judged the relative depth of the peak of the surface and a probe of known disparity. This reference is necessary given that the edges of the 3-D Kanizsa figure are illusory and, therefore, make disparity estimation more difficult. The point of subjective equality (PSE) for each stimulus was measured using a depth-discrimination paradigm. In disparity-probe procedures, the depth estimate of the test object is obtained by comparing the relative depth between the probe and test object. However, it is important to note that the task can also be completed implicitly by comparing the relative disparity of the probe and test object (Howard & Rogers, 2012). Therefore, in this paradigm, the PSEs represent the disparity at which the probe and surface peak are perceived as equivalent. To evaluate accuracy, the observed PSEs were compared to the relative disparity of the template used to generate the stimuli for each condition.

On each trial, observers initially fixated a Nonius cross at the center of the screen and aligned the vertical contours of the cross to fixate on the zero-disparity

reference plane. Once the cross was aligned, observers pressed a game-pad button to display the stimulus for 320 ms. This viewing time ensured that there was sufficient time for the illusory surface to form (approximately 140 to 200 ms; see I. Kogo, Liinasuo, & Rovamo, 1993; Reynolds, 1981; Ringach & Shapley, 1996) while restricting the amount of time observers had to complete a vergence eye movement. Although observers are capable of initiating vergence within 160 to 200 ms (Tulunay-Keesey & Jones, 1976; Westheimer & Mitchell, 1969; Yang, Bucci, & Kapoula, 2002), the time to complete a vergence eye movement can be upward of 800 ms (Rashbass & Westheimer, 1961). On all trials, observers were asked to indicate whether the disparity probe was located in front of or behind the peak of the surface using a game pad. Prior to each test session, observers performed a brief practice session of five trials per disparity so that we could choose an appropriate step size. For each stimulus, the disparity probe was presented at nine disparity levels. The four stimulus conditions were tested in separate blocks with blocks shown in random order. Stimuli were presented in random order 30 times each for a total of 270 trials per condition. The probe ranged in disparity from  $0.06^\circ$  to  $0.17^\circ$ . Observers made depth judgments for each surface condition over a range of disparities, the largest of which was well below the diplopia threshold. The inducer disparity (i.e., the disparity at the tip of the inducing element) of the reference stimulus was fixed at  $0.09^\circ$ . For the luminance-defined surface conditions, inducer disparity refers to the disparity at the same position along the vertical contour as the tip of the inducer.

We used a maximum likelihood method to fit a normal cumulative distribution function to the empirical psychometric function, and the PSE was computed as the 50% response point for each test condition for all observers ( $n = 6$ ). The analysis was performed using R statistical software and bootstrapped 95% confidence intervals (CIs) were calculated using Monte Carlo simulation methods run 1,000 times for each data set (Wichmann & Hill, 2001a; Wichmann & Hill, 2001b).

## Results

Figure 5 shows the PSE estimates for each observer in each of the four stimulus conditions. To assess the differences in PSE, a repeated-measures ANOVA examined the effect of stimulus condition on the mean PSE. The analysis revealed a significant difference in mean PSE across the four surface conditions,  $F(3, 15) = 123.05$ ,  $p < 0.001$ ,  $\eta^2 = 0.95$ . The differences in the PSE between stimulus conditions were examined using pairwise  $t$  tests with Benjamini and Hochberg's (1995) correction for false discovery rate. For all observers, the estimated disparity of the low- and high-contrast peaks

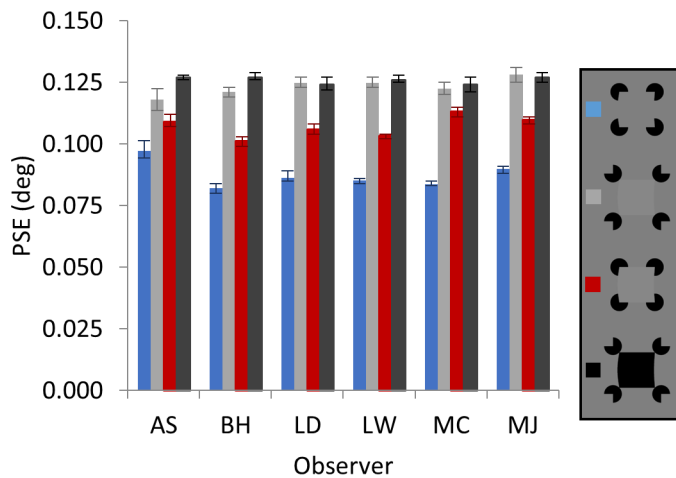


Figure 5. PSEs for all observers ( $n = 6$ ) in each of the four stimulus conditions: illusory (blue), low-contrast (gray), combined (red), and high-contrast (black). Error bars represent 95% CI.

was similar, and estimates of the illusory and combined surface peaks were shifted downward. Pairwise  $t$  tests on the means confirmed that the difference between the illusory ( $PSE_I = 0.087^\circ$ ,  $SE = \pm 0.002$ ), low-contrast ( $PSE_{LC} = 0.123^\circ$ ,  $SE = \pm 0.001$ ), and combined ( $PSE_C = 0.107^\circ$ ,  $SE = \pm 0.002$ ) surface conditions were all significant ( $p < 0.01$ ). There was no significant difference in perceived depth between the low- and high-contrast ( $PSE_{HC} = 0.126^\circ$ ,  $SE = \pm 0.001$ ) luminance-defined surfaces ( $p = 0.17$ ). Importantly, the PSEs obtained using the high- and low-contrast surfaces closely matched the disparity at the peak of the surface template for all observers. This confirms that observers could accurately localize the peak of these surfaces and that the disparity probe did not introduce a bias.<sup>1</sup> On average, PSEs were lower when the surface was illusory than in the low-contrast condition ( $p < 0.001$ ). This indicates that the trajectory of the interpolated surface was shallower for these illusory surfaces than in the surface template used to create the stimulus. In the combined condition, the peak was consistently localized as lying between the illusory and low-contrast surface peaks ( $p < 0.001$  and  $p = 0.001$ , respectively). Thus, when the surface was defined by luminance edges that occluded the inducing elements, the presence of the inducers consistently *reduced* the perceived depth at the peak of the surface. This occurred even though, when presented on its own, the luminance-defined signal was consistently matched to a larger disparity.

## Discussion

### *Illusory surface interpolation*

Experiment 1 showed that the perceived disparity at the peak of the curvilinear illusory surface was

consistently shallower than specified by the sinusoidal template used to generate the stimuli. A comparison of the PSEs in the illusory condition to the relative disparity at the tip of the inducing elements revealed that on average the perceived peak of the surface ( $0.09^\circ$ ,  $SE = \pm 0.002$ ) was approximately the same as the disparity at the tip of the inducing edge. In this respect, the interpolation seen here is consistent with studies of surface structure (Anderson, 2003) and disparity interpolation (Mitchison & McKee, 1987) that show, when a disparity signal is interpolated across an ambiguous region, the interpolated disparity tends to be equivalent to the disparity of the nearest unambiguous element. This pattern of results is consistent with observers simply matching the disparity of the inducer tips rather than relying on the perceived depth of the interpolated surface. However, if this were the case, then the tips of the inducers should appear to bend toward the observer in depth. Instead, inspection of the stimuli reveals that the disparity signal at the inducer tip appears to be assigned to the illusory surface while the inducer itself appears almost fronto-parallel (Figure 4). To determine if this impression was shared by observers, in a follow-up study we evaluated the perceived depth of the inducer tips and illusory surface boundary. If observers explicitly relied on the depth of the inducer tips in the preceding experiment, then we would predict that the perceived offset in depth (depth magnitude) of these two regions would be the same; furthermore, they should show the same dependence on disparity. Here we asked observers to estimate the perceived depth in the region of the tip of one of the inducing elements. On half the trials, they were told to indicate the depth of the high-contrast inducer tip, and on the remaining trials, they were asked to base their judgments on the perceived depth of the adjacent (illusory) surface at that location. The trial type was randomized and indicated by displaying the word “tip” or “surface” prior to the stimulus presentation. Observers had unlimited time to make their responses and indicated depth magnitude using a custom-built (and previously validated) haptic sensor strip (see Deas & Wilcox, 2014; Hartle & Wilcox, 2016). We tested four observers using three inducer tip disparities ( $0^\circ$ ,  $0.04^\circ$ , and  $0.09^\circ$ ).

We found that observers reported significantly more depth when asked to report the depth of the illusory surface (Figure 6); furthermore, these estimates increased significantly as a function of disparity. However, estimates explicitly based on the inducer tip showed no such dependence. Although the inducer depth estimates appear to increase from zero disparity, the fact that observers indicated when they perceived zero depth removed the uncertainty in depth estimates at zero disparity that is caused by variability in finger placement at the low end of the scale. The inclusion of this variability causes a bias toward overestimation of

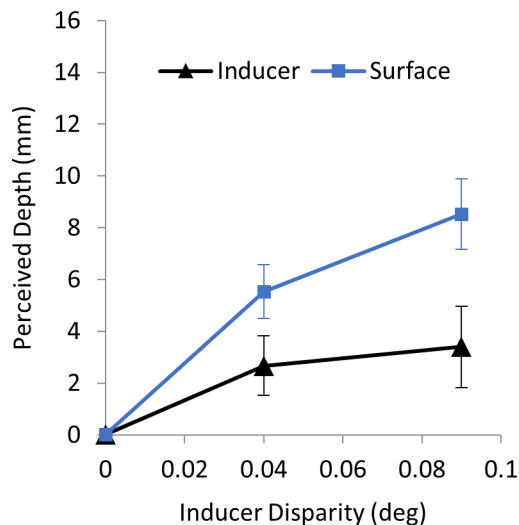


Figure 6. Mean depth estimates ( $n = 4$ ) for the inducer (black triangle) and the surface (blue squares) at the tip of the inducing element. Error bars represent 1 SEM.

depth at zero disparity regardless of the stimulus. However, in this analysis, the pairwise  $t$  test confirmed the increase from zero to the first test disparity for inducer depth estimates is not significant ( $p = 0.05$ ) even with the variability at zero disparity excluded. Overall, these results confirm that observers can and do report the depth at the peak of the illusory surface separately from the depth at the inducer edge. These results are also consistent with the well-known boundary ownership phenomenon reported in figure-ground literature (Anderson, 2003; Anderson, Singh, & Fleming, 2002), whereby the depth from disparity can disambiguate boundary ownership along the contour by assigning the depth to one side of the contour or the other but not both (Anderson & Julesz, 1995).

### Combined surface

In the combined condition, the introduction of a luminance-defined disparity signal at the boundary of the Kanizsa figure caused the perceived location of the peak to become shallower compared to luminance-defined surface peak. Given that the relative disparity along the surface edge was equivalent in the combined and low-contrast conditions, this result seems counterintuitive. Experiment 1 provided perceived depth estimates for illusory contours  $Q_i$ , luminance-defined contours  $Q_l$ , and the combined cue condition  $Q_c$ . The goal of Experiment 2 was to estimate the variance associated with these perceived depth estimates. The variance associated with the combined MLE estimate is

$$\hat{\sigma}_c^2 = \frac{\sigma_i^2 \sigma_l^2}{\sigma_i^2 + \sigma_l^2}. \quad (2)$$

Given that the presence of inducing elements strongly affected the location of the surface peak (despite a suprathreshold luminance-defined disparity signal and higher contrast at the corners of the surface), the local shape information at the inducers could have strongly influenced the perceived location of the surface peak because the shape information provided by the inducers was more reliable than the disparity-defined peak of the low-contrast luminance edge. If true, we would expect to find that depth judgments are more precise for the illusory surface than for the low-contrast, luminance-defined surface. We evaluate this prediction in Experiment 2.

## Experiment 2

To estimate the variance of perceived depth estimates for the surface conditions in Experiment 1, we used a 2IFC depth-discrimination task. This paradigm avoids the influence of the probe used in Experiment 1 on the variance of depth estimates by comparing two stereoscopic surfaces directly. Thus, the reliability estimate for each surface condition from Experiment 2 represents only the variance of perceived depth estimates of each surface condition.

### Observers

The observers in Experiment 1 also participated in this study. One observer was removed from the final analysis because he or she performed at chance in the low-contrast condition, resulting in a total of six observers.

### Stimuli

The four stimulus conditions tested here used the same stimuli described in Experiment 1. All stimuli were presented along with the fusion pattern and Nonius cross described in Experiment 1. The inducer disparity values (i.e., disparity at the tip of the inducers) used in each condition were sampled symmetrically around the reference disparity of  $0.09^\circ$ . The range of disparities was the same for all observers. The figures were presented with one of nine crossed inducer disparities (step size of  $0.02^\circ$ ).

### Procedure

The just noticeable difference (JND) for each stimulus condition was measured using a 2IFC paradigm and the method of constant stimuli. The

reference stimulus was randomly presented in the first or second interval, and the test stimulus was presented in the other interval. The reference stimulus had an inducer disparity of  $0.09^\circ$  and was always the same surface condition as used for the test stimulus. Each stimulus was viewed for 320 ms, separated by a 750 ms interstimulus interval during which the observers viewed a Nonius cross. On each trial, observers were asked to indicate which of the two intervals contained the surface with more depth between the plane with the inducing elements and the peak of the surface. Prior to each test session, observers completed four trials per test disparity to familiarize themselves with the stimuli and task. The four stimulus conditions were assessed in separate blocks, and in each block the nine test disparities were randomly presented 30 times each for a total of 270 trials per condition. Condition order was randomized across observers, and observers had a short break between each block of trials. To estimate the variance for each condition, the empirical psychometric function was fit using a normal cumulative distribution function using MLE, and bootstrapped 95% CI were calculated using Monte Carlo methods (Wichmann & Hill, 2001a; Wichmann & Hill, 2001b).

## Results and discussion

The JND was computed for each condition for all observers. JNDs represent 1  $SD$  of the fitted cumulative Gaussian function divided by  $\sqrt{2}$  to account for the 2IFC procedure (Green & Swets, 1974). Figure 7 shows the JND estimates for each observer in each of the four stimulus conditions. Observers made relatively precise estimates of the depth at the surface peak for all stimuli.

The repeated-measures ANOVA (which was the same as described in Experiment 1) revealed a significant difference in the mean JNDs between stimulus conditions,  $F(3, 15) = 18.22$ ,  $p < 0.001$ ,  $\eta^2 = 0.53$ . Thus, as predicted, precision varied across the four surface conditions. Pairwise  $t$  tests showed that despite the absence of luminance-defined boundaries or features in the illusory surface, mean estimates were equally precise in the illusory ( $JND_I = 0.014^\circ$ ,  $SE = \pm 0.002^\circ$ ) and high-contrast ( $JND_{HC} = 0.013^\circ$ ,  $SE = \pm 0.001^\circ$ ) surface ( $p = 0.63$ ) conditions. Importantly, the average precision was poorer in the low-contrast surface condition ( $JND_{LC} = 0.025^\circ$ ,  $SE = \pm 0.003^\circ$ ) than in any other condition ( $p = 0.01$ ); thus, the perceived depth of the low-contrast luminance-defined surface was *less reliable* than the depth of an illusory surface with the same dimensions. The reliability of the mean estimates in the illusory and combined ( $JND_C = 0.014^\circ$ ,  $SE = \pm 0.002^\circ$ ) conditions was similar ( $p = 0.73$ ), suggesting that the presence of the inducers improved

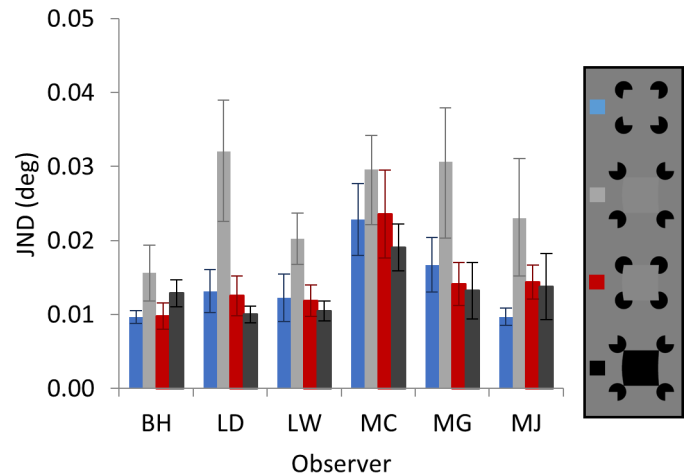


Figure 7. JNDs for each observer in each condition: illusory (blue), low-contrast (gray), combined (red), and high-contrast (black). Error bars represent 95% CI.

the precision of depth estimates, bringing them to a level similar to that of a salient, high-contrast, luminance-defined surface.

## General discussion

In Experiment 1, a disparity-probe paradigm was used to measure perceived depth for stereoscopic illusory, luminance, and combined surfaces. To evaluate the precision of these estimates, in Experiment 2 observers compared the shape of two stereoscopic surfaces using a depth-discrimination task. Taken together, these experiments suggest that the illusory surface supports more reliable depth estimation than the luminance-defined surface and that the presence of inducing elements has a strong influence on the depth of the combined surface peak. In the following section, these data were used to examine the combination of illusory and luminance-defined contours using the MLE cue-combination model given by Equation 1.

As outlined earlier, the illusory and luminance-defined depth estimates in our MLE model are the PSEs from the depth-discrimination task in Experiment 1, and the standard deviation ( $\sigma$ ) of the estimates are the JNDs measured in the 2IFC task in Experiment 2. The linear model in Equation 1 was used to predict PSEs and JNDs in the combined condition ( $PSE_c$  and  $\sigma_c$ , respectively) for each observer who had full data sets from both Experiments 1 and 2 ( $n = 5$ ). To assess whether the observed  $PSE_c$  and  $\sigma_c$  estimates were consistent with MLE model predictions, we compared the 95% CI of the observed and predicted  $PSE_c$  and  $\sigma_c$  for each observer (Figure 8). The results revealed that the  $\sigma_c$  of observed depth estimates were consistently



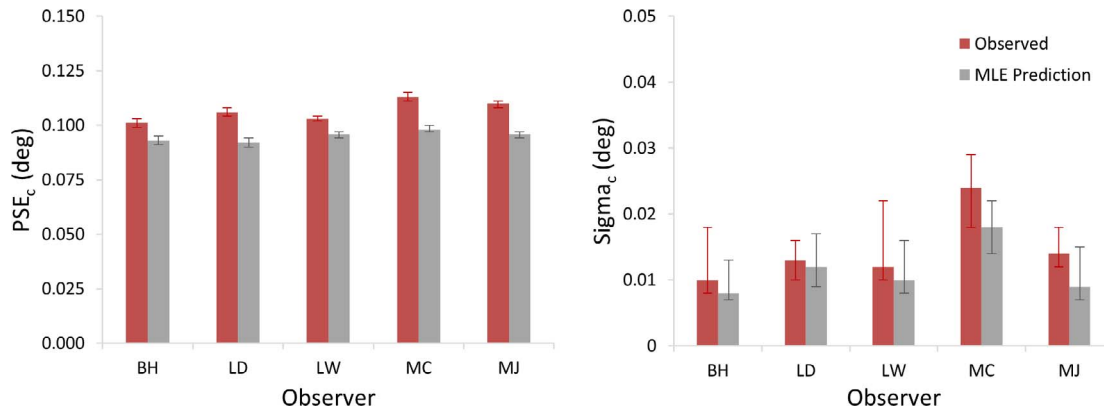


Figure 8. Observed and predicted PSEs and sigmas for the combined condition for each observer. Error bars represent 95% CI.

within the 95% CI of the MLE predictions for all five observers. However, comparison of the 95% CI for the PSE<sub>c</sub> showed that the observed PSEs were much higher than the predicted PSEs for all observers.

Although the observed  $\sigma_c$  are consistent with the MLE model predictions, the observed PSEs show a systematic bias; depth from the illusory surface is consistently underweighted, resulting in a combined estimate that is larger than MLE predictions. Observers underweight the depth estimate from illusory boundaries despite the considerable influence of the illusory boundary on the location of the surface peak in Experiment 1. Given that the results of Experiment 2 showed that the depth from illusory boundaries was more reliable than the depth from low-contrast luminance-defined edges, the MLE model predicts that when estimating the depth of the combined surface observers should assign even more weight to the illusory boundaries.

## A model of ambiguous depth cues

Why do observers underweight depth estimates from illusory contours? One possibility is that illusory contours are especially ambiguous depth cues. Two sources of ambiguity are (a) noise in the measurements of the visual system (e.g., internal errors in estimating disparity; see Cormack, Landers, & Ramakrishnan, 1997) and (b) many-to-one relationships between properties of the external environment and retinal images. Although noise contaminates depth estimates from both illusory and luminance-defined contours, the many-to-one relationship may have a much greater impact on illusory contours. As outlined in the Introduction, in these Kanizsa figures illusory contours are the visual system's attempt to perceive partly camouflaged surfaces. Images of camouflaged surfaces are ambiguous because hidden sections of the surface can only be inferred from relatively distant image features. One consequence of this ambiguity is addi-

tional uncertainty in the location of the camouflaged surface. In a Kanizsa square, for example, even if the visual system accurately estimates the 3-D orientations of the corners of the central square where they partly occlude the inducers, there is a range of plausible surface shapes that could connect the corners across the empty image regions between the inducers.

In this section, we suggest that the classic MLE cue-combination model fails to account for our findings on depth perception from illusory contours because it does not take this kind of ambiguity into account. As outlined below, it is possible to incorporate ambiguity into classic cue-combination models, and the resulting model gives a better account of our findings with illusory contours.

Suppose we have a one-dimensional family  $I(v)$  of images that depict partly camouflaged surfaces. We assume that the images are ambiguous: An image  $I(v)$  may depict many partly camouflaged surfaces that have the same visible components but different camouflaged (i.e., invisible) components.

The observer views a randomly chosen image  $I(V)$  and judges the depth of a point of interest in the camouflaged region. Here  $V$  is a random variable whose value is the parameter  $v$  that picks out the randomly chosen image. The depth of the point of interest is also a random variable  $D$ . To model the ambiguous information that each image  $I(v)$  provides about  $D$ , we assume that  $D$  is conditionally distributed as

$$P(D = d | V = v) = \phi(d, v, \sigma_D). \quad (3)$$

Here  $\phi(x, \mu, \sigma)$  is the normal probability density function, and  $\sigma_D$  is a parameter that quantifies the depth ambiguity of the images. Thus, given an image  $I(v)$ , the depth of the point of interest follows a normal distribution with mean  $v$  and standard deviation  $\sigma_D$ . Equation 3 implies that the image parameter  $v$  is not arbitrary: We have parameterized the family of images  $I(v)$  such that the mean depth of the point of interest over all partly camouflaged surfaces depicted by the image  $I(v)$  is equal to  $v$ . That is, the images are parameterized by the

mean depths that they depict at the camouflaged point of interest.

As in classic cue-combination models, we assume that the observer views the image  $I(V)$  and computes an unbiased but noisy depth cue  $X$  for the point of interest, given by

$$X = V + N_X. \quad (4)$$

Here  $N_X$  is a normally distributed random variable with mean zero and standard deviation  $\sigma_X$ . We assume that  $V$  and  $N_X$  are independent. The observer uses the depth cue  $X$  to estimate the depth  $D$  of the point of interest. In the Appendix, we show that, under the model outlined here, the probability distribution of the depth cue  $X$  given true depth  $D = d$  is closely approximated by

$$P(X = x|D = d) = \phi\left[x, d, (\sigma_X^2 + \sigma_D^2)^{1/2}\right]. \quad (5)$$

The maximum likelihood depth estimate  $\hat{d}$  is the value of  $d$  that maximizes this expression, so  $\hat{d} = x$ , and the uncertainty associated with this estimate is  $(\sigma_X^2 + \sigma_D^2)^{1/2}$ . Thus, in a single-cue depth-judgment task, the observer simply uses the value of the depth cue  $X$  to estimate the depth  $D$  of the point of interest. The ambiguity does not affect the observer's maximum likelihood depth estimate or their JND in the single-cue condition. This means that in a single-cue depth-discrimination task such as our Experiment 2, we can use the slope of the psychometric function to estimate the depth cue noise parameter  $\sigma_X$ , and this slope is not affected by the ambiguity parameter  $\sigma_D$ .

In a multiple-cue depth-judgment task, an observer who follows the MLE model combines the depth estimates from individual cues, weighted according to their reliability. In the Appendix, we show that under the model outlined here, the reliability of the depth cue  $X$  is  $r_X = (\sigma_X^2 + \sigma_D^2)^{-1}$ , which is smaller than the reliability  $r_X = \sigma_X^{-2}$  we would expect from a classic cue-combination model that considers only depth cue noise and does not take depth ambiguity into account. As a result, a maximum likelihood observer who understands depth-cue ambiguity will assign a lower weight to an ambiguous depth cue than we would expect from the slope of their psychometric function in a single-cue depth-discrimination task. This is precisely what we found in our experiments on depth judgments from illusory and luminance-defined contours (Figure 8). Thus, despite the presence of ambiguity in the illusory contour cue in both the single-cue and multiple-cue depth-judgment tasks, ambiguity only affects the weights assigned in the multiple-cue condition while the JND in the single-cue condition remains unaffected.

We estimated the value of the ambiguity parameter  $\sigma_D$  for each observer using his or her observed PSE for the combined surface (Figure 9). The ambiguity parameter  $\sigma_D$  for observers BH, LD, LW, MC, and MJ

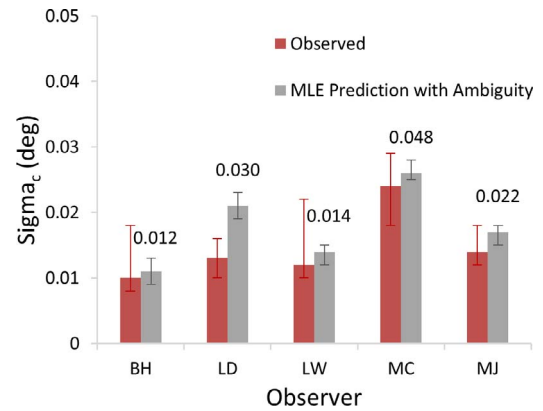


Figure 9. Observed and predicted standard deviations for the combined condition for each observer. Data labels show the estimated intrinsic ambiguity for each observer. Error bars represent 95% CI.

were 0.012, 0.030, 0.014, 0.048, and 0.022, respectively. The estimated intrinsic ambiguity due to the ambiguity of the illusory contour is on average 1.8 times the observed standard deviation of illusory depth estimates. This estimate seems reasonable given it is within an order of magnitude of the observed standard deviation of the perceived depth of the combined surface for all observers. Thus, very little depth ambiguity is necessary to account for our findings. In addition, comparison of the observed standard deviations in Figure 8 to the predicted standard deviations in Figure 9 shows that the inclusion of intrinsic ambiguity in the model yields predicted standard deviations that are within the 95% CI of the observed standard deviations for four of the five observers. Given that the inclusion of intrinsic ambiguity produces a combined estimate that is consistent with our observed estimates, our results are consistent with the explanation that observers attribute additional uncertainty to the position of illusory contours and consequently underweight their contribution to perceived depth.

The addition of this intrinsic ambiguity increases the uncertainty in the location of the illusory surface peak beyond what is expected from internal noise alone. We propose that the intrinsic ambiguity is responsible for the systematic underweighting of the depth from the illusory surface. Although previous studies have shown that illusory and luminance-defined contours are processed similarly (Larsson et al., 1999; von der Heydt et al., 1984), our results demonstrate that there is a crucial difference: Illusory contours have additional intrinsic ambiguity beyond what is expected based on luminance-defined contours. Despite the shape invariance and consistency of 3-D illusory surfaces generated from Kanizsa figures (Carman & Welch, 1992), the visual system accounts for the ambiguity in the position of illusory contours by reducing their contribution to the perceived depth of stereoscopic surfaces.

## Correlated error

Given that the illusory and luminance cues share the high-contrast portion of the surface edge, it is possible that internal cue estimates were correlated. In the model of correlated error proposed by Oruç, Maloney, and Landy (2003), if two cues are correlated with correlation  $\rho$ , then the optimal choice of weight for a single cue is

$$w_1 = \frac{r_1 - \rho\sqrt{r_1 r_2}}{r_1 + r_2 - 2\rho\sqrt{r_1 r_2}}, \quad (6)$$

where the value of  $r_i = \frac{1}{\sigma_i^2}$  is the reliability of each single cue estimate. This model accounts for a suboptimal choice of weights by correcting the reliability of each single cue condition by  $-\rho\sqrt{r_1 r_2}$ . As the correlation  $\rho$  increases, additional weight is placed on the more reliable cue. The derivative of Equation 6 with respect to the correlation  $\rho$  is

$$\frac{dw_1}{d\rho} = \frac{(r_1 - r_2)\sqrt{r_1 r_2}}{(r_1 + r_2 - 2\rho\sqrt{r_1 r_2})^2}. \quad (7)$$

When cue 1 is more reliable ( $r_1 > r_2$ ), the derivative is positive, and when cue 2 is more reliable ( $r_1 < r_2$ ), the derivative is negative. Thus, increasing the correlation  $\rho$  between the two cues always increases the weight assigned to the more reliable cue. If there were significant correlation between illusory and luminance-defined contours, then greater weight would be placed on the depth from illusory boundaries because it was significantly more reliable than depth from low-contrast luminance-defined edges in Experiment 2. Thus, the presence of a correlation between illusory and luminance-defined contours cannot account for the underweighting of illusory boundaries seen in our data. Importantly, this finding shows that the ambiguity model makes predictions in the opposite direction of a correlation model. Our ambiguity model can provide researchers with a valuable tool for understanding cue-combination behavior that departs from the predictions of the standard MLE model.

## Summary

The aim of this series of experiments was to assess the combination of perceived depth defined by illusory boundaries and luminance-defined edges in stereoscopic curved surfaces using a MLE cue-combination paradigm. Experiment 1 demonstrated that the addition of the inducing elements critically impacts estimates of depth from binocular disparity, and Experiment 2 confirmed our hypothesis that the presence of inducers forms a more reliable surface percept by increasing the precision of depth estimates of the surface peak. The

evaluation of the combination of illusory and luminance-defined stereoscopic surfaces suggests that intrinsic ambiguity in the position of illusory boundaries influences observers' depth estimates when combining this information with a luminance-defined signal. The reliability of the position of the illusory surface represents measurement noise as well as the intrinsic ambiguity in the position of the camouflaged boundary.

Although our analysis provides a normative MLE model of the results, it does not rule out the possibility that surface perception under these conditions involves more complex processes. For example, our current model does not consider the contribution of 2-D depth cues (e.g., occlusion and luminance relationships), which differed across conditions. The combination of such quantitative and qualitative depth cues is complex and poorly understood (Landy et al., 1995). Previous approaches to this issue have focused on how cues are combined when they are either consistent or in conflict; however, it has been shown that in cases of cue conflict there are large individual differences in the way that cues are combined (Cavanagh, 1987). For instance, Knill (2007) showed the weights assigned to stereoscopic and figural cues to surface slant depend on the level of cue conflict. He concluded their combination is represented by a mixture model that allows prior models to differ between observers depending on their interpretation of sensory cues. Recently, N. Kogo, Drozdowska, Zaenen, Alp, and Wagemans (2014) assessed the perceived depth of planar Kanizsa figures by varying the structure and polarity of the inducers and compared the perceived depth to similar Kanizsa-like figures with inducers that do not create illusory boundaries. They proposed a nonlinear dynamic weighting model to describe the combination of occlusion cues and depth from binocular disparity in Kanizsa configurations, which suggests consistent cues may work together to enhance depth perception of illusory surfaces and reduce the ambiguity of individual cues. Other studies have suggested that global geometry (such as the consistency of 2-D and 3-D curvature) may contribute to the integration of individual cues beyond a simple weighted linear summation (Stevens, Lees, & Brookes, 1991). It is possible that such interactions occurred in our surface configurations but were not directly evaluated by our cue-combination methodology.

## Conclusions

Our normative model provides a starting point for further investigation of interactions between depth from disparity, occlusion features, and luminance relationships in stereoscopic Kanizsa figures. Our results suggest that observers combine perceived depth

from illusory and luminance-defined surfaces according to a linear combination rule that takes into account intrinsic ambiguity in the true location of the camouflaged boundary. Although illusory and luminance-defined contours share many perceptual similarities, our model suggests that the visual system processes the precision of depth information from these two types of contours differently. Despite the high precision of perceived depth estimates for illusory boundaries, the visual system appears to take into account the intrinsic ambiguity in the position of illusory boundaries when combining their depth with depth estimated from luminance-defined elements.

*Keywords:* illusory contours, depth perception, stereopsis, disparity interpolation, cue combination

## Acknowledgments

This research was supported by NSERC funding to L. M. Wilcox and R. F. Murray and an Ontario Graduate Scholarship to B. Hartle. The authors would also like to thank James H. Elder for his suggestions regarding the ambiguity of depth cues from illusory contours.

Commercial relationships: none.

Corresponding author: Brittney Hartle.

Email: brit1317@yorku.ca.

Address: Centre for Vision Research, Lassonde Building, York University, Toronto, Canada.

## Footnote

<sup>1</sup> To evaluate whether the contrast polarity (black inducers on a gray background) was important to the pattern of results, in a followup study a subset of observers ( $n = 3$ ) compared the perceived peak of a high-contrast white surface to the black surface used in Experiment 1. There was no significant difference in the PSE of the peaks.

## References

- Anderson, B. L. (2003). The role of occlusion in the perception of depth, lightness, and opacity. *Psychological Review*, *110*(4), 785–801.
- Anderson, B. L., & Julesz, B. (1995). A theoretical analysis of illusory contour formation in stereopsis. *Psychological Review*, *102*, 705–743.
- Anderson, B. L., Singh, M., & Fleming, R. W. (2002). The interpolation of object and surface structure. *Cognitive Psychology*, *44*, 148–190.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, *57*, 289–300.
- Bradley, D. R., & Dumais, S. T. (1984). The effects of illumination level and retinal size on the depth stratification of subjective contour figures. *Perception*, *13*, 155–164.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.
- Bromiley, P. (2003). Products and convolutions of Gaussian probability density functions. *Tina-Vision Memo*, *3*(4), 1.
- Carman, G. J., & Welch, L. (1992, December 10). Three-dimensional illusory contours and surfaces. *Nature*, *360*, 585–587.
- Cavanagh, P. (1987). Reconstructing the third dimension: Interactions between color, texture, motion, binocular disparity, and shape. *Computer Vision, Graphics, and Image Processing*, *37*, 171–195.
- Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society*, *4*, 102–118.
- Coren, S. (1972). Subjective contours and apparent depth. *Psychological Review*, *79*(4), 359–367.
- Coren, S., & Porac, C. (1983). Subjective contours and apparent depth: A direct test. *Perception & Psychophysics*, *33*, 197–200.
- Cormack, L. K., Landers, D. D., & Ramakrishnan, S. (1997). Element density and the efficiency of binocular matching. *Journal of the Optical Society of America*, *14*(4), 723–730.
- Day, R. H. (1987). Cues for edge and the origin of illusory contours: An alternative approach. In S. Petry & G. E. Meyer (Eds.), *The perception of illusory contours* (pp. 53–61). New York: Springer-Verlag.
- Deas, L. M., & Wilcox, L. M. (2014). Gestalt grouping via closure degrades suprathreshold depth percepts. *Journal of Vision*, *14*(9):14, 1–13, <https://doi.org/10.1167/14.9.14>. [PubMed] [Article]
- Dillenburger, B., & Roe, A. W. (2010). Influence of parallel and orthogonal real lines on illusory contour perception. *Journal of Neurophysiology*, *103*, 55–64.
- Dresp, B., & Bonnet, C. (1995). Subthreshold summation with illusory contours. *Vision Research*, *35*(8), 1071–1078.
- Dresp, B., Lorenceau, J., & Bonnet, C. (1990). Apparent brightness enhancement in the Kanizsa

- square with and without illusory contour formation. *Perception*, 19(4), 483–489.
- Endler, J. (2006). Disruptive and cryptic coloration. *Proceedings of the Royal Society of London B: Biological Sciences*, 273, 2425–2426.
- Ernst, M. O., & Banks, M. S. (2002, January 24). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(24), 429–433.
- Fahle, M., & Palm, G. (1991). Perceptual rivalry between illusory and real contours. *Biological Cybernetics*, 66, 1–8.
- Fry, G. A., Bridgman, C. S., & Ellerbrock, V. J. (1949). The effect of atmospheric scattering on binocular depth perception. *American Journal of Optometry*, 29, 9–15.
- Gillam, B., & Nakayama, K. (2002). Subjective contours at line terminations depend on scene layout analysis, not image processing. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 43–53.
- Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. Huntington, NY: Robert E. Krieger.
- Hartle, B., & Wilcox, L. M. (2016). Depth magnitude from stereopsis: Assessment techniques and the role of experience. *Vision Research*, 125, 64–75.
- He, Z. J., & Ooi, T. L. (1998). Illusory-contour formation affected by luminance contrast polarity. *Perception*, 27(3), 313–335.
- Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4(12):1, 967–992, <https://doi.org/10.1167/4.12.1>. [PubMed] [Article]
- Howard, I. P., & Rogers, B. J. (2012). *Perceiving in depth: Volume 2, Stereoscopic vision*. Oxford, UK: Oxford University Press.
- Kanizsa, G. (1955). Margini quasi-percettivi in campi con stimolazione omogenea [Quasi-perceptual margins in homogeneously stimulated fields]. *Rivista di Psicologia*, 49, 7–30. In Petry S. & Meyer, G. E. (1987). *The Perception of Illusory Contours*. New York, NY: Springer.
- Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, 23, 141–221.
- Knill, D. C. (2007). Learning Bayesian priors for depth perception. *Journal of Vision*, 7(8):13, 1–20, <https://doi.org/10.1167/7.8.13>. [PubMed] [Article]
- Knill, D. C., & Saunders, J. (2002). Humans optimally weight stereo and texture cues to estimate surface slant. *Journal of Vision*, 2(7): 400, <https://doi.org/10.1167/2.7.400>. [Abstract]
- Kogo, I., Liinasuo, M., & Rovamo, J. (1993). Spatial and temporal properties of illusory figures. *Vision Research*, 33, 897–901.
- Kogo, N., Drozdowska, A., Zaenen, P., Alp, N., & Wagemans, J. (2014). Depth perception of illusory surfaces. *Vision Research*, 96, 53–64.
- Kogo, N., Strecha, C., Gool, L. V., & Wagemans, J. (2010). Surface construction by a 2-D differentiation-integration process: A neurocomputational model for perceived border ownership, depth, and lightness in Kanizsa figures. *Psychological Review*, 117(2), 406–439.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35, 389–412.
- Larsson, J., Amunts, K., Gulyas, B., Malikovic, A., Zilles, K., & Roland, P. (1999). Neuronal correlates of real and illusory contour perception: Functional anatomy with PET. *European Journal of Neuroscience*, 11, 4024–4036.
- Legge, G. E., & Gu, Y. (1989). Stereopsis and contrast. *Vision Research*, 29, 989–1004.
- Li, C., & Guo, K. (1995). Measurements of geometric illusions, illusory contours and stereo-depth at luminance and colour contrast. *Vision Research*, 35(12), 1713–1720.
- Maloney, L. T., & Landy, M. S. (1989). A statistical framework for robust fusion of depth information. *SPIE Visual Communications and Image Processing IV*, 1199, 1154–1163.
- McKee, S. P. (1983). The spatial requirements for fine stereoacuity. *Vision Research*, 23(2), 191–198.
- Mitchison, G. J., & McKee, S. P. (1985, May 30). Interpolation in stereoscopic matching. *Nature*, 315, 402–404.
- Mitchison, G. J., & McKee, S. P. (1987). Interpolation and the detection of fine structure in stereoscopic matching. *Vision Research*, 27(2), 295–302.
- Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, 43, 2451–2468.
- Paradiso, M. A., Shimojo, S., & Nakayama, K. (1989). Subjective contours, tilt aftereffects, and visual cortical organization. *Vision Research*, 29, 1205–1213.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.

- Poom, L. (2001). Visual inter-attribute contour completion. *Perception*, *30*, 855–865.
- Prazdny, K. (1983). Illusory contours are not caused by simultaneous brightness contrast. *Perception and Psychophysics*, *34*(4), 403–404.
- Ramachandran, V. S. (1986). Capture of stereopsis and apparent motion by illusory contours. *Perception & Psychophysics*, *39*, 361–373.
- Rashbass, C., & Westheimer, G. (1961). Disjunctive eye movements. *Journal of Physiology*, *159*, 339–360.
- Reynolds, R. I. (1981). Perception of an illusory contour as a function of processing time. *Perception*, *10*, 107–115.
- Ringach, D., & Shapley, R. (1996). Spatial and temporal properties of illusory contours and amodal boundary completion. *Vision Research*, *36*(19), 3037–3050.
- Rohaly, A. M., & Wilson, H. R. (1999). The effects of contrast on perceived depth and depth discrimination. *Vision Research*, *39*, 9–18.
- Rubin, N. (2001). The role of junctions in surface completion and contour matching. *Perception*, *30*, 339–366.
- Schor, C. M., & Howarth, P. A. (1986). Suprathreshold stereo-depth matches as a function of contrast and spatial frequency. *Perception*, *15*(3), 249–258.
- Smith, A. T., & Over, R. (1979). Motion aftereffect with subjective contours. *Perception & Psychophysics*, *25*, 95–98.
- Stevens, K., Lees, M., & Brookes, A. (1991). Combining binocular and monocular curvature features. *Perception*, *20*, 425–440.
- Tulunay-Keesey, U., & Jones, R. M. (1976). The effect of micromovements of the eye and exposure duration on contrast sensitivity. *Vision Research*, *16*, 481–488.
- von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984, June 15). Illusory contours and cortical neuron responses. *Science*, *224*, 1260–1262.
- Vreven, D., & Welch, L. (2001). The absence of depth constancy in contour stereograms. *Perception*, *30*, 693–705.
- Wehrhahn, C., & Dresch, B. (1998). Detection facilitation by collinear stimuli in humans: Dependence on strength and sign of contrast. *Vision Research*, *38*(3), 423–428.
- Westheimer, G., & Mitchell, D. E. (1969). The sensory stimulus for disjunctive eye movements. *Vision Research*, *9*, 749–755.
- Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception and Psychophysics*, *63*(8), 1293–1313.
- Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception and Psychophysics*, *63*(8), 1314–1329.
- Yang, Q., Bucci, M. P., & Kapoula, Z. (2002). The latency of saccades, vergence, and combined eye movements in children and in adults. *Investigative Ophthalmology & Visual Science*, *43*(9), 2939–2949.

## Appendix

Here we expand on the model of ambiguous depth cues outlined in the main text. The maximum likelihood depth estimate  $\hat{d}$ , given depth cue  $X = x$ , is the value of  $d$  that maximizes  $P(X = x | D = d)$ . To find this conditional probability, we start by partitioning over the possible values of the stimulus image parameter  $V$ :

$$P(X = x | D = d) = \int_{-\infty}^{\infty} P(X = x | V = v \ \& \ D = d) P(V = v | D = d) dv \quad (\text{A1})$$

We assume that given the image parameter  $V$ , the true depth  $D$  of the camouflaged point gives no additional information about the observer's depth cue  $X$  (i.e.,  $X$  is conditionally independent of  $D$ , given  $V$ ):

$$\int_{-\infty}^{\infty} P(X = x | V = v) P(V = v | D = d) dv \quad (\text{A2})$$

Bayes' theorem gives

$$= \int_{-\infty}^{\infty} P(X = x | V = v) P(D = d | V = v) P(V = v) / P(D = d) dv. \quad (\text{A3})$$

Equations 3 and 4 give expressions for these conditional probabilities.

$$\int_{-\infty}^{\infty} \phi(x, v, \sigma_X) \phi(d, v, \sigma_D) P(V = v) / P(D = d) dv, \quad (\text{A4})$$

$$\int_{-\infty}^{\infty} \phi(v, x, \sigma_X) \phi(v, d, \sigma_D) P(V = v) / P(D = d) dv. \quad (\text{A5})$$

We use the fact that the point-wise product of normal probability density functions is a scaled normal probability density function (Bromiley, 2003).

$$= \int_{-\infty}^{\infty} \phi \left[ x - d, 0, (\sigma_X^2 + \sigma_D^2)^{1/2} \right] \phi(v, \mu', \sigma') P(V = v) / P(D = d) dv, \quad (\text{A6})$$

where  $\mu' = (\sigma_X^{-2}x + \sigma_D^{-2}d)/(\sigma_X^{-2} + \sigma_D^{-2})$  and  $\sigma' = (\sigma_X^{-2} + \sigma_D^{-2})^{-1/2}$

$$= \phi \left[ x, d, (\sigma_X^2 + \sigma_D^2)^{1/2} \right] \times \int_{-\infty}^{\infty} \phi(v, \mu', \sigma') P(V = v) / P(D = d) dv. \quad (\text{A7})$$

Equation A7 contains a weighted integral of  $P(V = v)/P(D = d)$  with weights given by a normal probability density function whose value is negligible outside the range  $\mu' \pm 3\sigma'$ . Within this small range, if the priors on depicted depth  $V$  and actual depth  $D$  are weak, then the ratio  $P(V = v)/P(D = d)$  is approximately constant, and we use this approximation here. Letting this constant be  $k$ ,

$$\approx k\phi \left[ x, d, (\sigma_X^2 + \sigma_D^2)^{1/2} \right] \int_{-\infty}^{\infty} \phi(v, \mu', \sigma') dv. \quad (\text{A8})$$

Now the term inside the integral is a probability density function, which integrates to one.

$$= k\phi \left[ x, d, (\sigma_X^2 + \sigma_D^2)^{1/2} \right]. \quad (\text{A9})$$

Thus, the maximum likelihood depth estimate is approximately  $\hat{d} = x$ , and the uncertainty associated with this estimate is  $(\sigma_X^2 + \sigma_D^2)^{1/2}$ . As explained in the main text, in a single-cue depth-discrimination task, a maximum likelihood observer simply uses the value of the unbiased depth cue  $X$  to judge the depth of the point of interest, and we can recover  $\sigma_X$  from the slope of the observer’s psychometric function.

In a depth-discrimination task in which two depth cues,  $X$  and  $Y$ , are available, the maximum likelihood depth estimate depends on both cues. Here we assume that  $X$  is unbiased but ambiguous in the manner described by the model outlined here and that  $Y$  is unbiased and unambiguous as in classic cue-combination models. We also assume that  $X$  and  $Y$  are conditionally independent given the true depth  $D$ . The maximum likelihood depth estimate is the value of  $d$  that maximizes

$$P(X = x, Y = y | D = d) = P(X = x | D = d) P(Y = y | D = d), \quad (\text{A10})$$

$$= k\phi \left[ d, x, (\sigma_X^2 + \sigma_D^2)^{1/2} \right] \phi(d, y, \sigma_Y^2). \quad (\text{A11})$$

Again, we use the fact that the point-wise product of normal probability density functions is a scaled normal probability density function.

$$= k\phi \left[ x - y, 0, (\sigma_X^2 + \sigma_D^2 + \sigma_Y^2)^{1/2} \right] \phi(d, \mu'', \sigma''), \quad (\text{A12})$$

where  $\mu'' = \frac{r_X}{r_X+r_Y}x + \frac{r_Y}{r_X+r_Y}y$

and  $\sigma'' = (r_X + r_Y)^{-1/2}$

with  $r_X = (\sigma_X^2 + \sigma_D^2)^{-1}$  and  $r_Y = \sigma_Y^{-2}$ .

Equation A12 is maximized when  $d = \mu''$ , so the maximum likelihood depth estimate is a weighted sum of depth cues  $X$  and  $Y$  with weights determined by the reliabilities  $r_X$  and  $r_Y$ , much as in classic cue-combination models. However, here the ambiguous depth cue  $X$  has reliability  $r_X = (\sigma_X^2 + \sigma_D^2)^{-1}$ , and so the greater the ambiguity term  $\sigma_D^2$ , the less reliable the cue, and the lower the weight it receives in the optimal weighted sum in Equation A12.

In classic cue-combination models, the “reliability” of a depth cue  $Y$  can mean both the inverse of the cue’s variance,  $r_X = \sigma_X^{-2}$ , and the value  $r_X$  used to construct the cue’s weight  $r_X/(r_X + r_Y)$  in an optimal weighted sum of depth cues because these values are equal. In our revised model, these values are not necessarily equal. In this case, it may be more descriptive to refer to the inverse variance  $\sigma_X^{-2}$  as the cue’s “stability,” because it describes the cue’s variability from trial to trial, and to keep the term “reliability” for the uncertainty measure  $(\sigma_X^2 + \sigma_D^2)^{-1}$  that describes how precisely the depth cue estimates true depth. Thus, for example, a noiseless but ambiguous depth cue could be completely stable ( $\sigma_X^{-2} = \infty$ ) but only partly reliable ( $(\sigma_X^2 + \sigma_D^2)^{-1} < \infty$ ).

We hope it is clear that this is a highly general model of ambiguity in cue combination that should be useful for understanding a wide range of tasks in addition to depth discrimination from illusory contours.

We conclude with a minor clarification about our model. As pointed out in the main text, Equation 3 implies that the image parameter  $v$  is not arbitrary in that the images  $I(v)$  are parameterized by the mean depth  $v$  that they depict at the point of interest. This is a genuine restriction on the set of stimulus images because in general there could be distinct images  $I_1$  and  $I_2$  that depict distinct sets of partly camouflaged surfaces that have the same mean depth  $v$  at the point of interest whereas in a one-dimensional family of images  $I(v)$  there can only be one image for each value of  $v$ . However, this restriction is valid in many depth judgment tasks, including the experiments we report here, in which the stimulus images are a one-dimensional family  $I(v)$  of stereoscopic Kanizsa squares, each of which can be interpreted as depicting a range of partly camouflaged surfaces with a unique mean depth  $v$  at the point midway along a vertical illusory contour.