

Integration of ordinal and metric cues in depth processing

Marco Bertamini

School of Psychology, University of Liverpool,
Liverpool, UK



Jasna Martinovic

School of Psychology, University of Liverpool,
Liverpool, UK



Sophie M. Wuerger

School of Psychology, University of Liverpool,
Liverpool, UK



J. Burge, M. A. Peterson, and S. E. Palmer (2005) reported that ordinal, configural cues of familiarity and convexity influence perceived depth even when unambiguous metric information in the form of binocular disparity is available. In their study, a shape that was both convex and familiar (i.e., a face) increased perceived depth in random dot stereograms if the shape was shown in the foreground and decreased perceived depth if it was shown in the background. It is generally assumed that luminance cues are necessary for pre-figural shape representation to influence figure-ground computations in this way (M. A. Peterson & B. S. Gibson, 1993); thus, Burge et al. (2005) had used a luminance edge. In this research, we asked whether configural cues need to be defined by luminance, contrast, or neither. For a sufficiently large disparity pedestal (about 2.5 arcmin), configural cues influenced perceived depth both for second-order contours and for contours defined only by disparity. The integration of ordinal and metric cues seems to be driven by the general saliency of the contours and not only by luminance information. This challenges the notion that the integration of such cues always needs to arise during figure-ground organization through early combinations of luminance-defined shape and binocular disparity.

Keywords: depth perception, figure-ground, disparity, configural cues, cue combination, occlusion, familiarity

Citation: Bertamini, M., Martinovic, J., & Wuerger, S. M. (2008). Integration of ordinal and metric cues in depth processing. *Journal of Vision*, 8(2):10, 1–12, <http://journalofvision.org/8/2/10/>, doi:10.1167/8.2.10.

Introduction

The process of image segmentation has two main aspects: Regions need to be identified in the image that relate to three-dimensional scene objects, and these regions need to be assigned a depth order. Although several types of cues exist in the image, this task is underspecified and the visual system must resort to rules or assumptions. Many of these rules have been described in the literature, for example, the classical Gestalt grouping principles can be seen as principles of segmentation. But how do different cues interact with each other? This is one of the central topics in current research because ambiguity can be greatly reduced by combining different cues. Converging information from multiple visual sub-systems reduces uncertainty, so this has an obvious adaptive appeal. However, it has been argued that different cues need to be in the same units if meaningful combination is to take place (Landy, Maloney, Johnston, & Young, 1995). This can be taken to imply that metric information and ordinal information provided by the processing of discrete configural properties belonging to an image cannot influence each other directly.

An important source of metric information about depth comes from binocular disparity because being spatially separated, each of the retinas gets a slightly different but overlapping image of the surrounding visual scene (Howard & Rogers, 2002). The amount of such horizontal disparity is directly correlated to the amount of depth. Consider a stereogram in which a region is divided in half along the vertical axis and the left side is specified as being in front of the right side through binocular disparity. Would the shape of the contour dividing the two regions have an effect on the magnitude of perceived depth or would it be completely irrelevant? The shape does not carry any metric information about depth, but it can carry cues to depth order, since several configural qualities of shape can *bias* the process of figure-ground stratification. Convexity is one type of ordinal cue that has long been considered important for figure-ground assignment: Convex regions tend to be perceived as figures and concave regions as ground (Kanizsa & Gerbino, 1976; Metzger, 1953). Bertamini and Lawson ([in press](#)) have recently found that convexity affects response time in random dot stereograms, in which depth order is not ambiguous. When the convex region was in front, responses were faster than when the concave region was in front. Besides convexity, another factor that does affect figure-ground

stratification is familiarity. It has been extensively studied by Peterson and collaborators (Gibson & Peterson, 1994; Peterson & Gibson, 1993, 1994a). They have shown that contours are more likely to be seen as boundaries of familiar shapes, and therefore in front of a background, when the shapes (such as a face or a sea horse) are depicted in the orientation in which they are typically seen (for a review, see Peterson & Skow-Grant, 2003).

This finding suggests that the familiarity of a shape is evaluated before figure-ground organization has been completed. In other words, object recognition operates on edges in the image and not only on the figural contours that are available after figure-ground segmentation (Peterson & Gibson, 1994a). As an explanation of this phenomenon, Peterson, de Gelder, Rapcsak, Gerhardstein, and Bachoud-Lévi (2000) proposed that shape properties such as familiarity and convexity create a competition, by mutual inhibition, between the opposite sides of a contour. The outcome of this competition is binary: One side gets figural status at the expense of the other side. In terms of what kind of edges are effective for object recognition, Peterson and Gibson (1993) have also argued that for shape to combine with disparity in depth perception, the edge between the two regions needs to be specified by luminance information, since it needs to arise early enough in processing to influence figure-ground organization.

A recent study by Burge, Peterson, and Palmer (2005) examined whether binocular disparity on the one hand and luminance-defined convexity and familiarity on the other jointly determine perceived depth in the presence of unambiguous metric information. Using the point of subjective equality (PSE) as a measure (the point at which the observer perceives the two stimuli as being the same), they found that observers perceived more depth in displays with a convex and familiarly shaped foreground (i.e., a face), and less depth in displays in which the convex and familiar shape was presented in the background region. From this, Burge et al. (2005) concluded that configural cues have an effect on metric depth perception. In other words, disparity is combined with familiarity and convexity, even though these cues are different in nature.

Burge et al. (2005) showed that metric and ordinal cues combine with each other. This is apparently at odds with the prediction that different cues must be in the same units for combination to occur. However, Burge et al. (2005) suggested a Bayesian explanation. They built their argument on the fact that occlusion relations between surfaces in natural visual scenes do not produce randomly distributed metric depth orders (Huang, Lee, & Mumford, 2000). Correlations between occlusion relations and metric depth provided by disparity could lead to a nonuniform likelihood distribution of metric depth values that result from occlusion in combination with the actual metric cues in the scene. Such statistical information coming from configural cues could be combined with any

other depth cue within the framework of Bayesian inference. This is not inconsistent with the modified weak fusion (MWF) model in Landy et al. (1995). In this model ordinal cues can influence the perception of metric depth if they are promoted to metric status or if they can disambiguate a stimulus in which depth order is ambiguous. In the case of stereograms, and in particular when disparity is far from threshold, both Burge et al. and the MWF model claim that to have an effect ordinal cues must be promoted to metric status. Therefore, the prediction is that a familiar shape in the foreground will increase the perceived metric depth and a familiar shape in the background will decrease perceived metric depth. This is consistent with the proposal that object recognition starts with edge information (Peterson & Gibson, 1994a) and inconsistent with the principle of unidirectional contour ownership (e.g., Nakayama & Shimojo, 1992; Nakayama, Shimojo, & Silverman, 1989), because if only figures have shape then there is no reason to expect a familiarly shaped background to differ from any other background. In other words, the effect of familiarity of the foreground shape does not in itself imply an effect of familiarity of the background region.

The main focus of our experiments was to investigate what kind of contour information is necessary when specifying the configural cue. As mentioned, earlier work on the effect of familiarity in stereograms suggested that luminance contours must be present in the display to influence figure-ground computations (Peterson & Gibson, 1993). This means that the availability of shape information should precede stereo fusion, which has implications for the time course of the integration process. Peterson and Gibson had come to that conclusion based on studies that used the number of figure-ground reversals as a measure of shape influence on figure-ground organization. In the current study we measured the effect of configural cues on perceived depth following the paradigm introduced by Burge et al. (2005), but, in addition to luminance contours, we tested second-order contours (defined by contrast) and dichoptic contours (defined solely by binocular disparity).

In summary, our study started with a replication of Burge et al. (2005) findings and subsequently attempted to extend it to other types of displays in which no luminance information was present. Configural effects on perceived metric depth should not extend to dichoptic contours if shape information needs to be present before stereo fusion to be combined with disparity information.

General method

Our stimuli and procedure were designed to be similar to those of Burge et al. (2005). Random dot stereograms contained two regions of approximately the same size

separated by a disparity-defined depth step along a central edge that had the shape of a face in profile (see [Figure 1](#)). This shape was chosen by Burge et al. due to its salience. A face promotes the selection of the surface it encloses as the figural side in nonstereoscopic bipartite displays (Peterson & Gibson, 1993, 1994a). Its effectiveness results from a joint effect of two figure-ground cues: convexity and familiarity. These shapes are also similar to the silhouetted face profiles recently studied by Davidenko (2007). He concluded that silhouettes are processed like regular face stimuli.

By adding disparity to the regions of a bipartite display, two types of displays were created. In the first one, the cues were consistent: Disparity information indicated the face-shaped region to be in front. In the second one, the cues were inconsistent: The face-shaped region was specified to be in the back. To measure the PSE, these displays were combined to create pairs comprising a standard stimulus and a comparison stimulus. The standard with fixed disparity was to be judged against the comparison of variable disparity. Four experimental conditions were thus produced: face in front in both standard and comparison (F-F); face in front in the standard paired with a non-face in front in the comparison (F-N); non-face in front in both standard and comparison (N-N); and non-face in front in the standard paired with a face in front in the comparison (N-F). To control for the possible effects of the location that was specified to be in front (left or right), these conditions were doubled, creating an equal number of left-sided and right-sided pairs.

We used a two-interval forced-choice (2IFC) task. The participants viewed one of the stimulus pairs (F-F, F-N, N-N, or N-F, left- or right-sided) and selected the interval that contained a greater depth separation between the two surfaces. The conditions F-F and N-N were included as control. In these conditions, it was expected that the perceived depth measured through the PSE should converge on the disparity of the standard stimulus. Burge et al. (2005) found that in the two experimental conditions, when the cues are inconsistent in one stereogram and consistent in the other, the face-shaped region was perceived as more separated from its background, and the non-face-shaped region was perceived as less separated from its background for the same amount of disparity. Therefore, we predicted that configural cues should affect the PSE in the manner observed by Burge et al.: When the face-shaped region is in front in the standard and the non-face is in front in the comparison (F-N), participants will need more disparity in the comparison to see the two stimuli as equal. The opposite should be the case when the non-face shaped side is in front in the standard and the face shaped is in front in the comparison (N-F): The participants will need less disparity to see the two stimuli as equal.

Random dot stereograms consisted of dots spread randomly over two central regions, surrounded by a square frame. The frame always had a 50% density and was always at a disparity-defined distance of 0.5 arcmin in front of the nearer surface. Dot density and disparity differed between experiments (see “Experiments”). In the standard stimulus, the farther region of the stereogram was fixed for any given

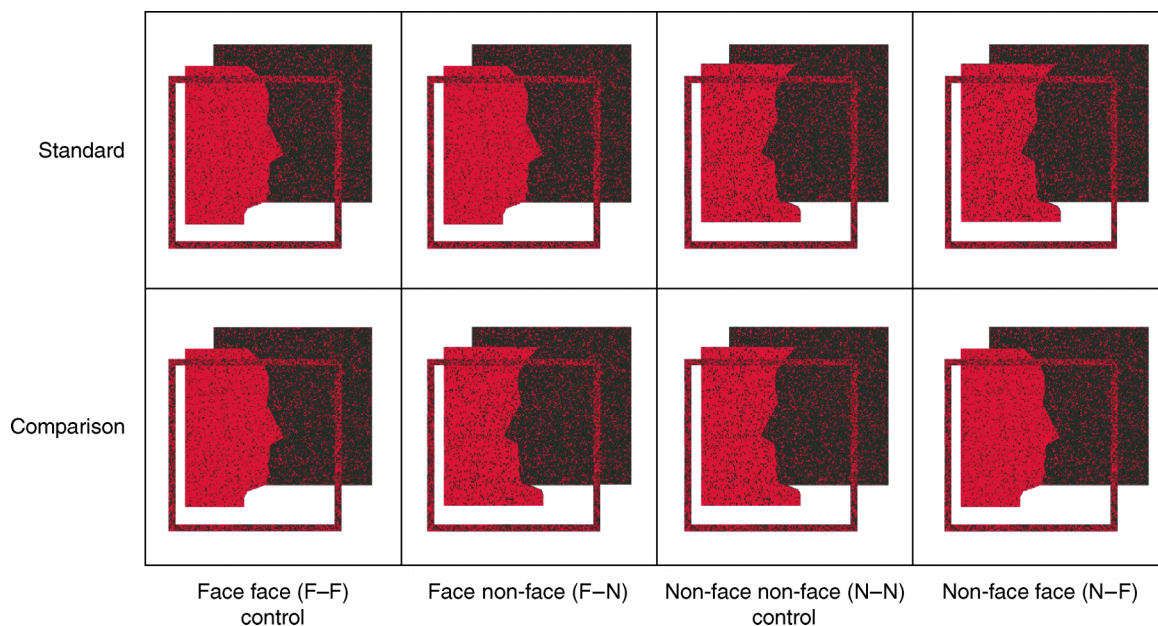


Figure 1. Stimuli for [Experiment 1](#). The three depth surfaces of the stereograms have been separated for clarity. In one case the front surface was a Face, and in another it was a Non-face. The four conditions were defined by the pair of a standard stimulus and a comparison stimulus (F-F, F-N, N-F, and N-N). Note that order of presentation was random, so F-N does not mean that the Face was necessarily in the first interval. During the experiment the standard stimulus remained constant while the disparity of the comparison was adjusted according to a one-up, one-down staircase procedure to sample points at or near the 50% point of the psychometric function.

experiment (e.g., it was 1.1 arcmin in [Experiment 1](#)); this value is referred to as the pedestal. In the comparison stimulus, the farther region could vary in depth, and went from being coplanar with the nearer surface to extending 12.7 arcmin in depth, with a step size of 0.5 arcmin. Each stimulus image subtended an angle of 4.8 by 4.8 deg, with the frame being 0.18 deg wide. The stimuli were presented centrally on a black screen, preceded by a fixation cross, which the participants were instructed to look at. The participants were seated in a dark room at a distance of 2 meters from the screen. The experiment was run on a Macintosh computer connected to a Sony F500TD monitor, with a resolution of 1,280 by 1,024 pixels run at 120 Hz. A C program was used to generate the stimuli, control their presentation, and collect data. Some of the VideoToolbox functions were used (Pelli, 1997). Two stereo images were presented using a NuVision infrared emitter and stereoscopic shutter glasses. Due to the interleaving of the left-eye and right-eye images, the effective vertical resolution and the refresh rate were halved (640 pixels at 60 Hz).

Each trial consisted of two intervals: a standard and a comparison stimulus, each presented for 1 s with a 0.5-s interstimulus interval (see [Figure 2](#)). The interval with the standard stimulus was randomly selected. The participants pressed buttons on a game pad to indicate the interval in which they perceived a bigger depth separation between the two regions. Intertrial intervals varied with the participants' response times. A total of eight conditions (F-F, F-N, N-N, and N-F, left- and right-sided) was given to each participant, via randomly interleaved one-up, one-down staircases that

varied the depth separation between the two surfaces in the comparison stimulus. The one-up, one-down reversal rule samples around the 50% point of the psychometric function (Levitt, 1970) and was used by Burge et al. (2005) because of its suitability for PSE measurements. Each staircase terminated after 12 reversals, which usually took around 20 to 40 trials, depending on the speed of convergence. Total number of trials per condition was 4 times this value because each participant completed four blocks of trials, each containing eight staircases (one for each condition). They were preceded by 18 practice trials, containing easy stimulus pairs from all conditions.

We fit a cumulative normal function to the raw psychometric data for each observer and each condition and derive the maximum likelihood estimate (MLE) of the mean and the standard deviation of the cumulative normal distribution. The mean of the psychometric function (the 50% point) is the PSE, and shifts in PSE between conditions are informative about changes in the perceived depth. The standard deviation is inversely related to the slope of the psychometric function and reflects the reliability of the observer's judgment (Treutwein & Strasburger, 1999). Combined effects of standard and comparison type were analyzed with a 2×2 repeated measures ANOVA. In addition, t tests against the pedestal value and paired t tests between control and experimental conditions were performed. Differences between the experiments containing the same type of contour but different pedestal ([Experiments 2a and 2b](#) and [Experiments 3a and 3b](#)) were examined using independent t tests for differences in PSEs and standard deviations.

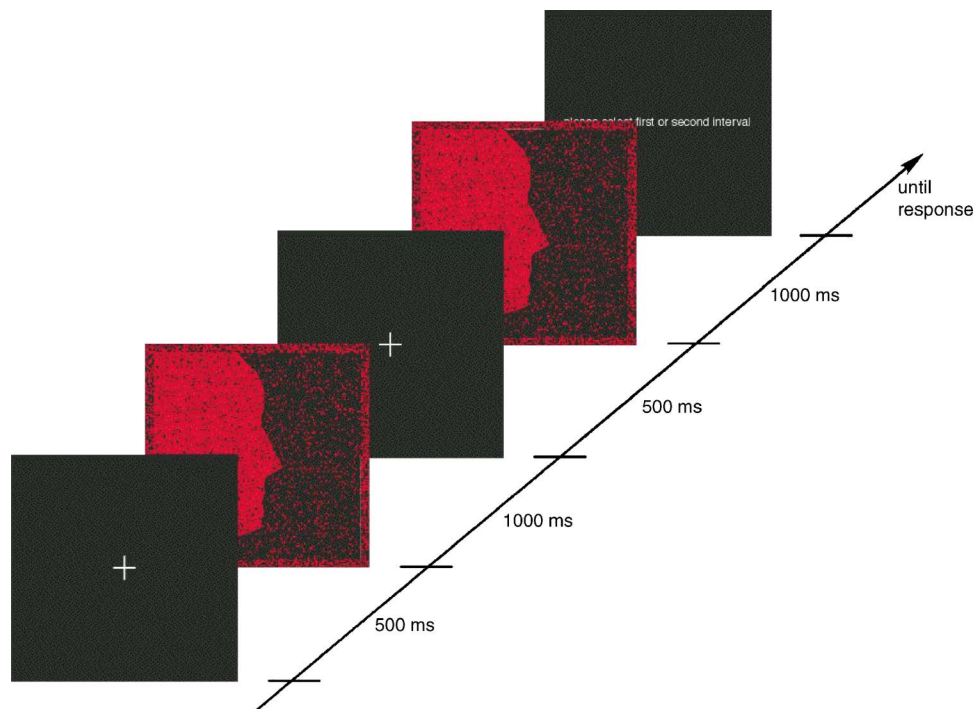


Figure 2. Two stimuli were presented for 1 s each, with a 0.5-s interstimulus interval. Observers had to make a forced choice on which foreground had greater depth separation from the background.

To evaluate whether differences between control and experimental conditions were significant for individual observers, we used a bootstrapping method (Wichmann & Hill, 2001) to estimate the variability of the individual PSE estimates. For each condition and each observer we resampled ($n = 1000$) the observed data and then estimated the PSE using the same MLE method as for the original data. We then conducted one-sample t tests on the PSE differences (experimental vs. control) and obtained p values for each observer.

Experiment 1: Replication

This experiment was a replication of the Burge et al. (2005) study. The only significant difference in our stimuli was that the disparity pedestal in the standard stimulus was fixed at 1.1 arcmin, while in Burge et al. (2005) it was 7.5 arcmin. A trade off exists between small and large values for the pedestal. A large value may make the shape more clearly visible, but it also increases the role of monocular (unmatched) regions relative to the role of binocular disparity per se. In other words, a smaller pedestal minimizes the difference between right and left images. Perhaps more importantly, for a fixed step size, a larger pedestal makes the task harder because the proportional (perceptual) change associated with a step would be smaller. Our study used a small depth separation in the standard to make the task relatively easier.

In this experiment, as in Burge et al. (2005), the central edge between the two surfaces was defined by binocular disparity but also by a luminance difference. The front surface was red with 10% black dots and the background surface was black with 10% red dots.

Method

Participants

Fourteen participants were recruited from the University of Liverpool campus; three of them were experienced observers and the remaining 11 were naïve. Out of those, two had to be removed from the sample because the psychometric function could not be fitted properly due to noisy data. All participants had normal or corrected-to-normal vision. Stereoacuity was recorded using the TNO stereotest and ranged between 30 and 60 arcsec.

Stimuli and procedure.

In this experiment, the pedestal was set at 1.1 arcmin and the dot density was 10%. Further details on stimuli and procedure are in “General method.” The procedure and the timing are summarized in Figure 2.

Results and discussion

In a preliminary analysis, the side factor (whether the foreground was on the left or on the right) did not show any significant effects, which replicates the findings of Burge et al. (2005). Therefore the side factor was collapsed and the psychometric functions recalculated.

A 2×2 repeated measures ANOVA with the factor condition type (control vs. experimental) and standard type (face in front or face in back) was performed. In Figure 3 the average PSE (in arcmin) for all four conditions is shown, demonstrating an effect of configural cues on perceived depth. There was a significant main effect of standard, $F(1, 11) = 9.14$, $p < 0.01$, and a highly significant interaction between the two factors of standard and condition, $F(1, 11) = 21.68$, $p < 0.001$. For experimental conditions, the PSE was higher than the pedestal in the F-N condition, $t(11) = -3.29$, $p < 0.01$, and lower than the pedestal in the N-F condition, $t(11) = 4.16$, $p < 0.01$. The PSE was higher than in control condition when F was the standard, F-N vs. F-F: $t(11) = 3.09$, $p < 0.01$, and lower when N was the standard, N-F vs. N-N: $t(11) = -3.24$, $p < 0.01$. On average, the configural cue was worth 0.17 ± 0.03 arcmin of disparity. For the control conditions, the PSE converged upon the pedestal, F-F: $t(11) = -0.26$, n.s.; N-N: $t(11) = 0.45$, n.s.

The lower graph of Figure 3 is a boxplot showing the difference between PSE and pedestal, and it illustrates the variability between individual participants. To evaluate whether these differences were significant for the individual observers, we estimated the reliability of the individual PSE estimates using bootstrapping (see “General method”) and performed t tests for each observer. We found that the differences between experimental and control conditions were present in 10 of 12 participants for F-N vs. F-F and in 11 of 12 participants for N-F vs. N-N; these differences were significant for each of the 10 (11) observers ($p < 0.01$). This confirms conclusions based on the t tests across all observers. The size of the effect varied between observers, but such variability was also reported by Burge et al. (2005) and is in accordance with reports of large inter-subject variability in cue-combination studies (Hillis, Ernst, Banks, & Landy, 2002).

In conclusion, with stimuli similar to those used by Burge et al. (2005), but with a smaller pedestal, we confirmed an effect of configuration (familiarity and convexity) on the PSE for perceived depth.

Experiment 2: Contours defined by contrast

In Burge et al. (2005) and in our Experiment 1 there was an effect of configuration using displays in which a

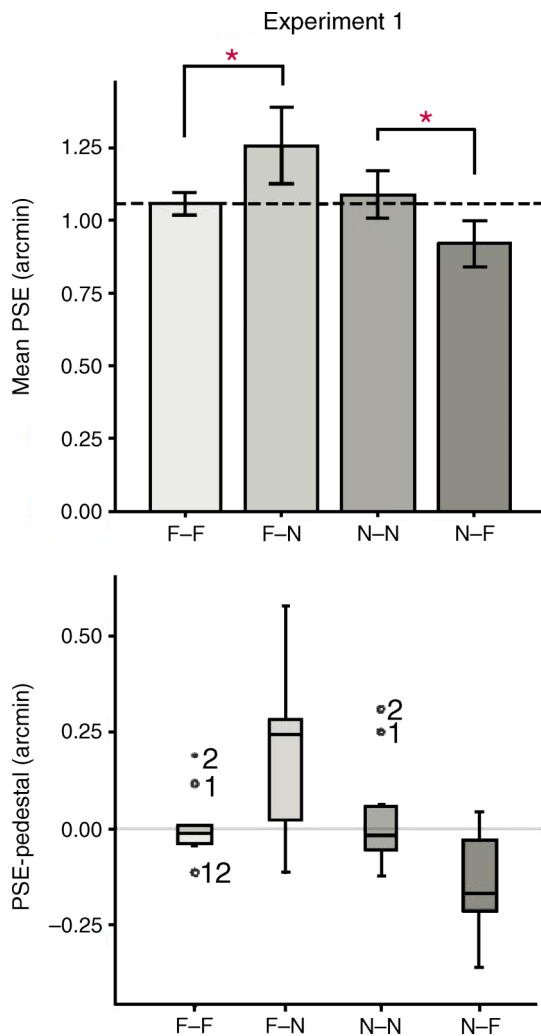


Figure 3. Results from Experiment 1. (Upper) Bar plots of PSEs for the four conditions, arranged according to the shape of the surface that is in the front in the standard and in the comparison: Face Face (F-F), Face Non-face (F-N), Non-face Non-face (N-N), and Non-face Face (N-F). Error bars depict 95% confidence intervals. (Lower) Boxplots of the PSEs in the sample. Midlines indicate medians, ends of boxes indicate 25th and 75th percentiles, ends of lines indicate 10th and 90th percentiles, and circles indicate outliers.

luminance contrast creates salient monocular contours. In Experiment 2 we introduced two changes: We eliminated the luminance information and we increased dot density.

Foreground and background were created with equal numbers of dark and light red dots (see Figure 4). The average luminance on both sides was equal and the contour was defined only by a contrast difference. Dot density has also changed. Instead of having 10% of dots sprinkled upon each of the surfaces as in Experiment 1, here dot density was 50% on each surface. This was necessary to create a second-order contour between them.¹

We conducted two versions of this experiment. In the first (Experiment 2a), the depth separation in the standard

stimulus was 1.1 arcmin, as in Experiment 1, and in the second (Experiment 2b), the depth separation in the standard stimulus was 2.6 arcmin. A larger pedestal was introduced to ensure that the contour was clearly visible, because a poorly defined dichoptic shape would make detecting a configural effect harder.

Method

Participants

For Experiment 2a, 12 participants were recruited from the University of Liverpool campus: 3 experienced and 9 naïve observers. All participants had normal or corrected-to-normal vision. Stereoacuity was recorded using the TNO stereotest and ranged between 15 and 240 arcsec.

For Experiment 2b, 12 participants were recruited from the University of Liverpool campus: 4 experienced and 8 naïve observers. All participants had normal or corrected-to-normal vision. Stereoacuity was recorded using the TNO stereotest and ranged between 15 and 60 arcsec.

Stimuli and procedure

The pedestal was set at 1.1 arcmin (Experiment 2a) or 2.6 arcmin (Experiment 2b) and the dot density was 50%. Average luminance was matched for the foreground and the background to 13.8 cd/m^2 (see Figure 4). Further details on stimuli and procedure can be found in “General method” and are presented in Figure 2.

Results and discussion

The side factor did not show any significant effects and was therefore collapsed and the psychometric functions calculated. Experiments 2a and 2b were analyzed separately. A 2×2 repeated measures ANOVA with the factor condition type (control vs. experimental) and standard type (face in front or face in back) was performed. Results are shown in Figure 5.

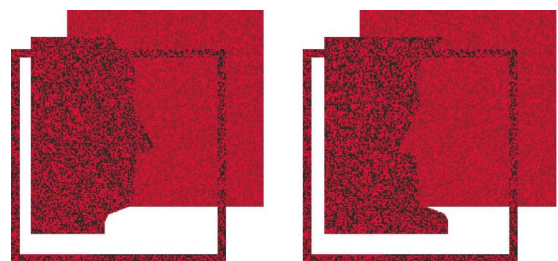


Figure 4. Stimuli for Experiment 2. In the foreground, half of the dots were red and half were black. In the background, half of the dots were dark red and half were light red.

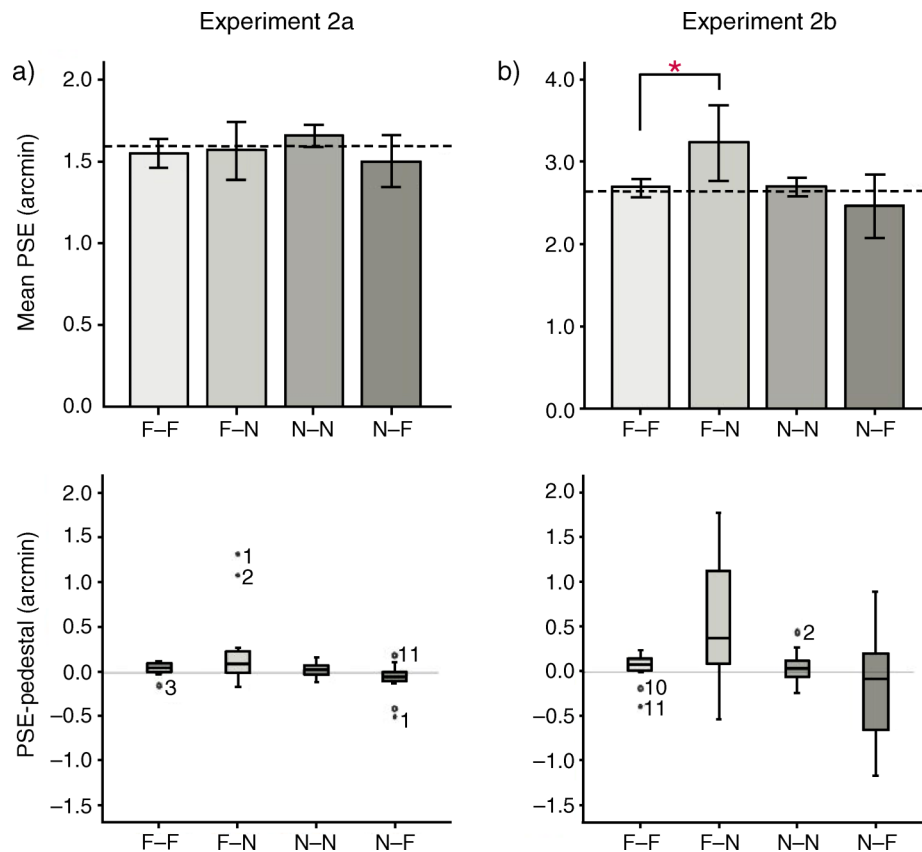


Figure 5. Results from Experiments 2a and 2b. (Upper) Bar plots of PSEs for the four conditions, arranged according to the shape of the surface that is in the front in the standard and in the comparison: Face Face (F-F), Face Non-face (F-N), Non-face Non-face (N-N), and Non-face Face (N-F). Error bars depict 95% confidence intervals. (Lower) Boxplots of the PSEs in the sample. Midlines indicate medians, ends of boxes indicate 25th and 75th percentiles, ends of lines indicate 10th and 90th percentiles, and circles indicate outliers.

Experiment 2a

The results indicate that there was a trend for configural cues to affect perceived depth, but no effect of condition or interaction: standard type, $F(1, 11) = 3.26$, $p = 0.098$; condition type, $F(1, 11) = 0.10$, n.s.; interaction, $F(1, 11) = 2.87$, n.s. For the control conditions, the PSE converged upon the pedestal both when face was in front, F-F: $t(11) = 1.28$, n.s., and when non-face was in front in the standard, N-N: $t(11) = 0.56$, n.s. The experimental conditions did not significantly differ from the pedestal, F-N, $t(11) = 1.74$, n.s.; N-F, $t(11) = -1.77$, n.s. Experimental and control conditions did not differ from each other when face was in front, F-N vs. F-F: $t(11) = -1.51$, n.s.; for only 7 of 12 observers this difference was significant (one-sample t test: $p < 0.01$). There was a trend toward a difference between experimental and control conditions when non-face was in front in the standard, N-F vs. N-N: $t(11) = 1.87$, $p = 0.09$; for 8 of 12 observers this difference reached significance (one-sample t test; $p < 0.01$). Figure 5a shows some variability between participants, with 2 of them showing a large effect of configuration on metric cues. Meanwhile, the PSEs for the majority of participants were more narrowly distributed around the pedestal.

Experiment 2b

There was a trend for configural cues to affect perceived depth, $F(1, 11) = 4.03$, $p = 0.07$, and a significant effect of condition, $F(1, 11) = 8.43$, $p < 0.05$. There was also a trend for an interaction between these two factors, $F(1, 11) = 4.18$, $p = 0.07$. It is evident from Figure 5 that the PSEs converged upon the pedestal in control conditions both when face was in front in the standard, F-F: $t(11) = 0.62$, n.s., and when non-face was in front, N-N: $t(11) = 0.81$, n.s. The experimental condition when face was in front significantly differed from the pedestal, F-N, $t(11) = 2.75$, $p < 0.05$, but this was not the case when non-face was in front, N-F, $t(11) = -1.13$, n.s. Experimental and control conditions differed significantly from each other when the face was in front in the standard, F-N vs. F-F: $t(11) = -2.70$, $p < 0.05$; at the individual level, this effect was significant for 8 of 12 observers ($p < 0.01$). There was no overall difference when the non-face was in front in the standard, N-F vs. N-N: $t(11) = 1.21$, n.s. The effect was significant for 8 of 12 observers ($p < 0.01$), but the average effect size was too small to reach significance when averaged across observers. On average, the configural cue was worth 0.38 ± 0.14 arcmin of disparity. It can be seen from Figure 5b that the PSEs for experimental

conditions tended to be broadly distributed, while the PSEs for the control conditions converged on the pedestal and were more narrowly distributed.

Between-subjects analysis of Experiments 2a and 2b

Prior to the analysis, three participants were removed because they did both of the experiments. Independent t tests were performed on PSEs and overall SD s; corrected degrees of freedom were used when equal variances assumption was violated. To normalize the effect of the PSE between conditions, the pedestal was first subtracted from the PSE for each participant. PSEs did not differ significantly from each other for any of the experimental conditions, F-F: $t(16) = -0.30$, n.s.; F-N: $t(9.16) = 1.60$, n.s.; N-N: $t(16) = 0.03$, n.s.; and N-F: $t(8.62) = -0.21$, n.s. On the contrary, SD s significantly differed between experiments for each condition, F-F: $t(8.33) = 2.64$, $p < 0.05$; F-N: $t(16) = 2.12$, $p = 0.05$; N-N: $t(8.56) = 2.38$, $p < 0.05$; and N-F: $t(8.58) = 2.73$, $p < 0.05$.

The finding of [Experiment 2a](#) might be taken as evidence that luminance cues are necessary for ordinal cues to exert an effect on metric cues. However, [Experiment 2a](#) showed a trend for an effect of configural cues in the ANOVA, due to a tendency to perceive less depth when the non-face was in front. Thus, an alternative explanation would be that the effect of configural on metric cues might still exist, but because the pedestal was set at a value of 1.1 arcmin, our step size of 0.5 arcmin led to the display being only two steps away from being co-planar with the front surface. This may have been too crude to capture the effect of less depth, which should be perceived when face was in front, while still managing to show the effect of more depth in the experimental condition in which the non-face was in front in the standard. In [Experiment 2b](#), all parameters were the same as in [Experiment 2a](#) except for the pedestal, which was reset to 2.6 arcmin. As predicted, clearer evidence of a configural effect on perceived depth was found in [Experiment 2b](#). This shows that (i) other types of contours can interact with metric information on depth order and that (ii) the type and size of the effect depend on the amount of disparity in the pedestal. A comparison between the two experiments indicates that while there was an overall increase in uncertainty with an increase in the pedestal (reflected by larger SD s), the perceptual effect itself did not change (no significant differences in PSEs).

Experiment 3: Disparity-defined contours

[Experiment 3](#) was designed to examine if the quantitative effect of configural cues on metric depth obtained in [Experiments 1](#) and [2](#) would persist when all information in

the image is only binocularly specified and does not contain any monocular luminance or contrast cues. After pilot tests, the depth separation in the standard stimulus was set to 1.6 arcmin for [Experiment 3a](#) and 2.6 arcmin for [Experiment 3b](#). This was done to ensure that the face-shaped contour was easily visible. The central edge between the surfaces was defined only through binocular disparity information because both surfaces had an equal number of red and black dots. Therefore, the two surfaces of the central stimulus display had exactly the same monocular properties.

Method

Participants

For [Experiment 3a](#), 13 participants were recruited from the University of Liverpool campus: 1 experienced and 12 naïve observers. One participant had to be removed from the sample because the staircases failed to converge properly. All participants had normal or corrected-to-normal vision. Stereoacuity was recorded using the TNO stereotest and ranged between 30 and 60 arcsec.

For [Experiment 3b](#), 12 participants were recruited from the University of Liverpool campus: 4 experienced and 8 naïve observers. All participants had normal or corrected-to-normal vision. Stereoacuity was recorded using the TNO stereotest and ranged between 30 and 60 arcsec.

Stimuli and procedure

The pedestal was set at 1.6 arcmin ([Experiment 3a](#)) or 2.6 ([Experiment 3b](#)) arcmin and the dot density was 50%. Unlike all previous experiments, there was no monocular difference at all between foreground and background surfaces. Further details on stimuli and procedure are in “[General method](#).”

Results and discussion

The side factor did not show any significant effects, so it was collapsed and the psychometric functions recalculated. [Experiments 3a](#) and [3b](#) were analyzed separately. A 2×2 repeated measures ANOVA with the factor condition type (control vs. experimental) and standard type (face in front or face in back) was performed.

Experiment 3a

The results indicate that there were no effects of configural cues on perceived depth in this experiment: standard type, $F(1, 11) = 0.58$, n.s.; condition type, $F(1, 11) = 0.27$, n.s.; and interaction, $F(1, 11) = 2.27$, n.s. For the control conditions, the PSE converged upon the pedestal in both conditions: F-F: $t(11) = -1.22$, n.s. and

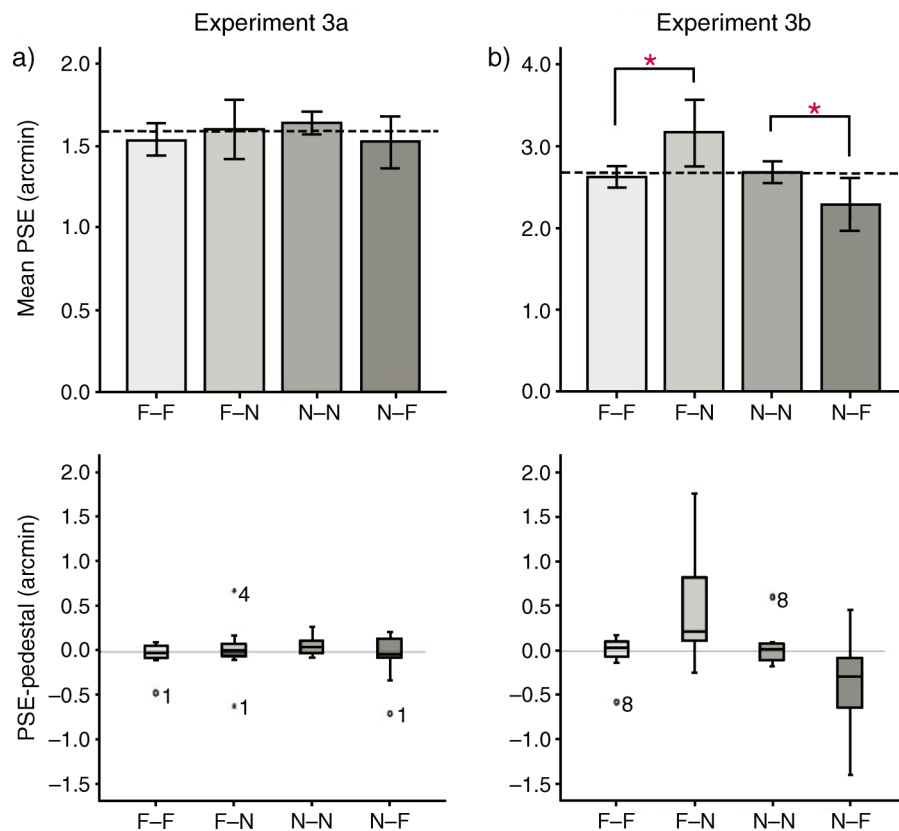


Figure 6. Results from Experiments 3a and 3b. (Upper) Bar plots of PSEs for the four conditions, arranged according to the shape of the surface that is in the front in the standard and in the comparison: Face Face (F-F), Face Non-face (F-N), Non-face Non-face (N-N), and Non-face Face (N-F). Error bars depict 95% confidence intervals. (Lower) Boxplots of the PSEs in the sample. Midlines indicate medians, ends of boxes indicate 25th and 75th percentiles, ends of lines indicate 10th and 90th percentiles, and circles indicate outliers.

N-N: $t(11) = 0.09$, n.s. The experimental conditions did not significantly differ from the pedestal either: F-N, $t(11) = 1.59$, n.s. and N-F, $t(11) = -0.92$, n.s. Experimental and control conditions did not differ from each other: F-N vs. F-F: $t(11) = -0.88$, n.s. and N-F vs. N-N: $t(11) = 1.29$, n.s. This was confirmed by the t tests for the individual subjects: only 5 of 12 observers showed a significant difference between F-F and F-N ($p < 0.01$), and 8 of 12 showed a significant difference between N-N and N-F ($p < 0.01$). Figure 6 shows that almost all of the PSEs fell around the pedestal, except for one participant who consistently underestimated the amount of depth in three of four conditions. It is also important to note that during the debrief, a few participants reported that they had some difficulties in clearly seeing the face-shaped contour once the depth had decreased to levels around the pedestal in both the standard and the comparison.

Experiment 3b

The results showed a quantitative effect of configural cues on perceived depth, as shown in Figure 6. There was a significant main effect of standard, $F(1, 11) = 8.59$, $p < 0.05$, and a significant interaction between the two factors of standard and condition, $F(1, 11) = 7.29$, $p < 0.05$.

For the control conditions, the PSE converged upon the pedestal when face was in front, F-F: $t(11) = -1.04$, n.s., but there was a trend for it to differ from the pedestal when non-face was in front, N-N: $t(11) = 0.48$, $p = 0.06$. The experimental condition was significantly different from the pedestal when face was in front in the standard: More depth was needed to perceive the two stimuli as being the same, F-N: $t(11) = 3.24$, $p < 0.01$. There was a trend for less depth needed if the standard contained the non-face shape, N-F: $t(11) = -1.86$, $p = 0.09$. Importantly, control and experimental conditions differed significantly from each other, F-N vs. F-F: $t(11) = -2.82$, $p < 0.05$ and N-F vs. N-N: $t(11) = -2.39$, $p < 0.05$. On an individual level, 10 of 12 observers showed a significant difference between F-N and F-F ($p < 0.01$), and 9 of 12 showed a significant difference between N-F and N-N ($p < 0.01$). On average, the configural cue was worth 0.42 ± 0.13 arcmin of disparity.

Between-subjects analysis of Experiments 3a and 3b

Prior to the analysis, one participant was removed because he did both of the experiments. Independent t tests were performed on PSEs and on overall SDs; corrected degrees of freedom were used when equal variances assumption was violated. The pedestal was subtracted from the PSE

before the analysis to normalize the data. PSEs did not differ significantly from each other for the majority of experimental conditions, (F-F: $t(20) = 0.07$, n.s.; N-N: $t(20) = -0.30$, n.s.; and N-F: $t(14.54) = -1.59$, n.s. In the F-N condition only there was a significant change, $t(13.98) = 2.16$, $p < 0.05$, with less depth perceived in the experiment with the larger pedestal. *SDs* did not significantly differ between experiments for most conditions, F-F: $t(20) = 0.72$, n.s.; N-N: $t(20) = 0.51$, n.s.; and N-F: $t(20) = 1.05$, n.s.; again, the only change occurred in the F-N condition, where there was a trend for an increase in response uncertainty for the experiment with the larger pedestal, $t(11.14) = 1.90$, $p = 0.08$.

Conclusions

Experiment 1 was a replication of the findings of Burge et al. (2005), and confirmed that configural cues of familiarity and convexity affect the perceived metric depth in the presence of unambiguous disparity information.

In Burge et al. and our own **Experiment 1**, configural cues were defined by luminance information. We investigated whether luminance information is necessary for configural cues to affect perceived depth, as Peterson and Gibson's (1993, 1994a) model would imply. Contrast-defined surfaces were tested in **Experiment 2**, while in **Experiment 3** the surfaces only differed in binocularly defined depth. We found that even in the absence of luminance-defined contours, configural cues influenced perceived depth.

A few research questions remain open. First, different cues can be tested. In both Burge et al. (2005) and in our experiments, familiarity and convexity are confounded. They could be isolated, and other depth cues could be tested, because a large set of factors are known to affect figure-ground segmentation, for example, closure or symmetry (e.g., Kanizsa & Gerbino, 1976; Kovács & Julesz, 1993). Data from Burge, Fowlkes, and Banks (2007) and our own preliminary results suggest a role of convexity independently of familiarity.

Second, more could be done to ensure that the configural effect reported by Burge et al. and by our work is not affected by response selection. Depth order in Burge et al. and in our own stimuli was unambiguous. However, if the disparity difference between the two intervals is small, the judgment will be uncertain, and observers may be biased by the shape at the time of the response selection. How likely is this possibility? It means that observers would have been biased to choose a sinusoidal shape when there was a face in the background in the second experiment of Burge et al. (2005). Note, however, that the non-face foreground stimulus is more concave than either the face stimulus or the sinusoidal stimulus. A bias against concavity would therefore be consistent with the available evidence.

Third, convexity, as discussed above, may be a factor on its own, but in addition, convexity creates a difference

in the likelihood that observers will be looking at the foreground or the background surface. If we assume that observers fixate the center of the display as instructed, at stimulus onset they will be looking at the foreground surface in the convex condition (e.g., face in front) and at the background surface in the concave condition (e.g., non-face in front). This is a confound hard to avoid when concave and convex stimuli are compared, but the role of different fixation planes could be tested in a separate experiment. In the context of ambiguous figure-ground displays there is good evidence that fixation is an important factor (Peterson & Gibson, 1994b).

Another aspect of our results was that the configural effect for second-order or purely disparity-defined contours was only observed for large pedestals of approximately 2.5 arcmin. There are two possible explanations for this. On the one hand, cue integration may depend on the reliability of the cue, as advocated by standard models of cue combination (Ernst & Banks, 2002). Therefore, as the reliability of one cue decreases the influence of other cues increases. This is consistent with a larger effect of a configural cue on perceived depth at larger disparity pedestals. On the other hand, the task may need to be difficult to reveal an effect of shape information because if selection between the two intervals is straightforward there is no bias in favor of either shape. In other words, participants may be biased to select the familiar shape as the foreground more often when their confidence in their judgments decreases. Note also that we found large inter-observer variability, as did Burge et al. (2005). Even when the majority of observers behaved similarly, the size of the effect varied from one person to another.

In conclusion, we found that the metric effect of figure-ground biases can be driven by both texture and disparity-defined edges. This is an important constraint in how such a phenomenon can be modeled. It challenges the notion that the integration of such cues always needs to arise during figure-ground organization through early combinations of luminance-defined shape and binocular disparity. Further research will have to discriminate between a high-level process that biases the choice between two stimuli with similar perceived metric depth and a promotion of ordinal information to metric status. Moreover, the effectiveness of a number of configural cues (convexity and familiarity in our case) could be compared. It would also be useful to measure whether depth values get assigned to different configurations in the absence of disparity information. The problem in that situation is one of depth-order ambiguity; nevertheless, the final goal should be to map all the conditions under which ordinal cues are promoted to metric status.

Acknowledgments

Jasna Martinovic was supported by a Study Visit Grant from the Experimental Psychology Society and by a grant from the British Academy. The authors would like to

thank Christopher Nolan for assistance with data collection and Matthias Gondan for helpful suggestions on the data analysis.

Commercial relationships: none.

Corresponding author: Marco Bertamini.

Email: M.Bertamini@liverpool.ac.uk.

Address: School of Psychology, University of Liverpool, Eleanor Rathbone Building, Bedford Street, South Liverpool, L69 7ZA, UK.

Footnote

¹The dots carry the disparity and are therefore essential to create surfaces in the frontoparallel plane, as discussed in Burge et al. (2005). For low density (e.g., 10%), however, there is a risk that the dots are seen as sprinkled on top, but not necessarily attached to a surface with solid color red for the foreground and black for the background. We are not claiming that this actually happened; nevertheless, such a possibility is eliminated when 50% of the dots have one color and 50% have the other color.

References

- Bertamini, M., & Lawson, R. (in press). Rapid figure-ground responses to stereograms reveal an advantage for a convex foreground. *Perception*.
- Burge, J., Fowlkes, C. C., & Banks, M. S. (2007). Configural cues, disparity, and depth perception: Internalization of natural scene statistics. *Perception*, 36, 180.
- Burge, J., Peterson, M. A., & Palmer, S. E. (2005). Ordinal configural cues combine with metric disparity in depth perception. *Journal of Vision*, 5(6):5, 534–542, <http://journalofvision.org/5/6/5/>, doi:10.1167/5.6.5. [PubMed] [Article]
- Davidenko, N. (2007). Silhouetted face profiles: A new methodology for face perception research. *Journal of Vision*, 7(4):6, 1–17, <http://journalofvision.org/7/4/6/>, doi:10.1167/7.4.6. [PubMed] [Article]
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433. [PubMed]
- Gibson, B. S., & Peterson, M. A. (1994). Does orientation-independent object recognition precede orientation-dependent recognition? Evidence from a Cuing paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 299–316. [PubMed]
- Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between senses. *Science*, 298, 1627–1630. [PubMed]
- Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth* (vol. 2, *Depth perception*). Toronto: Porteous.
- Huang, J., Lee, A. B., & Mumford, D. (2000). Statistics of range images. *Proceedings of the IEEE Conference on Computational Vision and Pattern Recognition*, 1, 324–331.
- Kanizsa, G., & Gerbino, W. (1976). Convexity and symmetry in figure-ground organization. In M. Henle (Ed.), *Vision and artifact* (pp. 25–32). New York: Springer Publishing Co.
- Kovács, I., & Julesz, B. (1993). A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proceedings of The National Academy of Science of the United States of America*, 90, 7495–7497. [PubMed] [Article]
- 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. [PubMed]
- Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49, 467–477. [PubMed]
- Metzger, W. (1953). *Gesetze des Sehens* (2nd ed). Frankfurt/M.: W. Kramer.
- Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, 257, 1357–1363. [PubMed]
- Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception*, 18, 55–68. [PubMed]
- Pelli, D. (1997). The Video Toolbox software for visual psychophysics transforming numbers into movies. *Spatial Vision*, 10, 437–442. [PubMed]
- Peterson, M. A., de Gelder, B., Rapcsak, S. Z., Gerhardstein, P. C., & Bachoud-Lévi, A. (2000). Object memory effects on figure assignment: Conscious object recognition is not necessary or sufficient. *Vision Research*, 40, 1549–1567. [PubMed]
- Peterson, M. A., & Gibson, B. S. (1993). Shape recognition contributions to figure-ground organization in three-dimensional displays. *Cognitive Psychology*, 25, 383–429.
- Peterson, M. A., & Gibson, B. S. (1994a). Must figure-ground organization precede object recognition? An assumption in peril. *Psychological Science*, 5, 253–259.

- Peterson, M. A., & Gibson, B. S. (1994b). Object recognition contributions to figure-ground organization: Operations on outlines and subjective contours. *Perception & Psychophysics*, *56*, 551–564. [[PubMed](#)]
- Peterson, M. A., & Skow-Grant, E. (2003). Memory and learning in figure-ground perception. In B. Ross & D. Irwin (Eds.), *Cognitive vision: Psychology of learning and motivation* (pp. 1–34). San Diego, CA: Academic Press.
- Treutwein, B., & Strasburger, H. (1999). Fitting the psychometric function. *Perception & Psychophysics*, *61*, 87–106. [[PubMed](#)]
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function. II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, *63*, 1314–1329. [[PubMed](#)] [[Article](#)]