

Ordinal configural cues combine with metric disparity in depth perception

Johannes Burge

Vision Science Program, University of California,
Berkeley, CA, USA



Mary A. Peterson

Department of Psychology, University of Arizona,
Tucson, AZ, USA



Stephen E. Palmer

Department of Psychology, University of California,
Berkeley, CA, USA



Prior research on the combination of depth cues generally assumes that different cues must be in the same units for meaningful combination to occur. We investigated whether the geometrically ordinal cues of familiarity and convexity influence depth perception when unambiguous metric information is provided by binocular disparity. We used bipartite, random dot stereograms with a central luminance edge shaped like a face in profile. Disparity specified that the edge and dots on one side were closer than the dots on the other side. Configural cues suggested that the familiar, face-shaped region was closer than the unfamiliar side. Configural cues caused an increase in perceived depth for a given disparity signal when they were consistent with disparity and a decrease in perceived depth when they were inconsistent. Thus, geometrically ordinal configural cues can quantitatively influence a metric depth cue. Implications for the combination of configural and depth cues are discussed.

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

Introduction

To provide useful information about the physical environment, the visual system must generate a reasonably accurate three-dimensional (3D) percept from optical information in two 2D retinal images. The actual 3D scene that gives rise to the images is geometrically underdetermined by this optical information, but the resulting ambiguity can be reduced by combining information from different cues relevant to the same environmental property. Depth cue combination is a topic on which there has been considerable recent research. An important assumption of this research has been that different cues must be in the same units for meaningful combination to take place (Landy, Maloney, Johnston, & Young, 1995). This study explores this assumption empirically by investigating whether ordinal information can influence depth perception when unambiguous metric information is present. The ordinal information comes from the configural cues of convexity and familiarity, important factors in determining figure-ground organization, and the metric information comes from binocular disparity, a potent factor in determining perceived depth.

Figure-ground organization occurs when two adjacent regions in the visual field are perceived as if one region (the “figure”) is nearer to the viewer and shaped by the common edge, whereas the other region (the “ground”) is farther from the viewer and not bounded by the common edge, appearing instead to extend behind the figure. Research on

figure-ground organization has focused primarily on identifying “configural” cues: stimulus properties that bias one region of a 2D display to be seen as nearer than the other and as shaped by the common edge (Palmer, 2002; Peterson & Skow-Grant, 2003). It is well known that the region that is more surrounded, smaller, more vertically oriented, higher in contrast, more symmetrical, bordered by more parallel contours, lower in the display, more convex, and more familiar is more likely to be seen as the nearer, figural region (Kanizsa & Gerbino, 1976; Peterson & Gibson, 1994a; Peterson, Harvey, & Weidenbacher, 1991; Rubin, 1915/1958; Vecera, Vogel, & Woodman, 2002). However, geometrical analyses of such factors indicate that metric information cannot be recovered from them. The shape of an occluding contour, for example, cannot specify the distance of either the occluding or the occluded surface; occlusion can only specify which one is closer. Geometrically, configural cues can therefore provide only ordinal information.

Perhaps because of the ordinal nature of configural information, figure-ground perception has been modeled by competitive interactions across an edge (e.g., Peterson, de Gelder, Rapcsak, Gerhardstein, & Bachoud-Lévi, 2000; Sejnowski & Hinton, 1987; Vecera & O'Reilly, 1998). The outcome of this activity is binary: one side (the figure) “wins” and appears shaped by the common edge, whereas the other side (the ground) “loses” and is not shaped by it. In some models (e.g., Sejnowski & Hinton, 1987; Vecera & O'Reilly, 1998), but not all (Peterson, 2003), perceived

depth ordering—the figure appearing closer than the ground—is also an outcome of the competition. This reflects the binary nature of standard phenomenological observations about figure-ground perception and is consistent with the geometrically ordinal nature of configural cues.

A very different picture of depth perception emerges from the literature on binocular disparity (Howard, 2002; Howard & Rogers, 2002). Horizontal disparity is a relative depth cue, but it can be interpreted metrically once distance and azimuth have been estimated, and empirical research has shown that metric information is indeed recovered (Backus, Banks, van Ee, & Crowell, 1999). In fact, it has been shown that geometrically available scaling parameters can metrically calibrate many different depth cues (e.g., disparity, motion parallax, and texture).

Recent work on depth cue combination has conceptualized the generation of a depth percept as a problem of statistical inference, specifying how the visual system should infer depth from noisy measurements and prior information. In this view, both the visual system's estimates of depth implied by various cues (likelihood functions) and by prior information (the prior probabilities) are modeled by probability distributions over metric space. Bayesian models allow optimal combination of such information to predict small, graded changes in depth perception that have been verified experimentally (e.g., Hillis, Ernst, Banks, & Landy, 2002; Hillis, Watt, Landy, & Banks, 2004; Knill, 1998). However, it is unclear how information from configural cues—indeed, from any geometrically ordinal cue—can be incorporated within this framework.

The empirical question we address in this work is whether geometrically ordinal depth information from the configural cues of familiarity and convexity combine with metric information from binocular disparity to influence depth perception. Surprisingly little research has examined this issue. Peterson and Gibson (1993) reported the most convincing evidence that configural cues affect perceived depth of stereoscopic displays, but they failed to settle the issue. They used stereograms in which adjacent black and white regions shared an edge whose shape suggested a familiar object (e.g., a face or seahorse in profile) on one side and whose binocular disparity suggested that the familiar region was either nearer to or farther from the observer than the unfamiliar region. When disparity suggested that the familiar region was nearer, observers usually reported perceiving two parallel planes separated in depth, with the familiar region in front (Figure 1a). When disparity suggested that the familiar region was farther, observers frequently reported that the familiar region appeared to be slanted in depth such that it was nearer at the central edge and farther at the outside edge (Figure 1b). Thus, two strikingly different depth interpretations resulted from the same disparity information, depending on how configural cues were aligned with it. This result therefore supports the conclusion that configural cues can influence perceived depth when a metric depth cue is present.

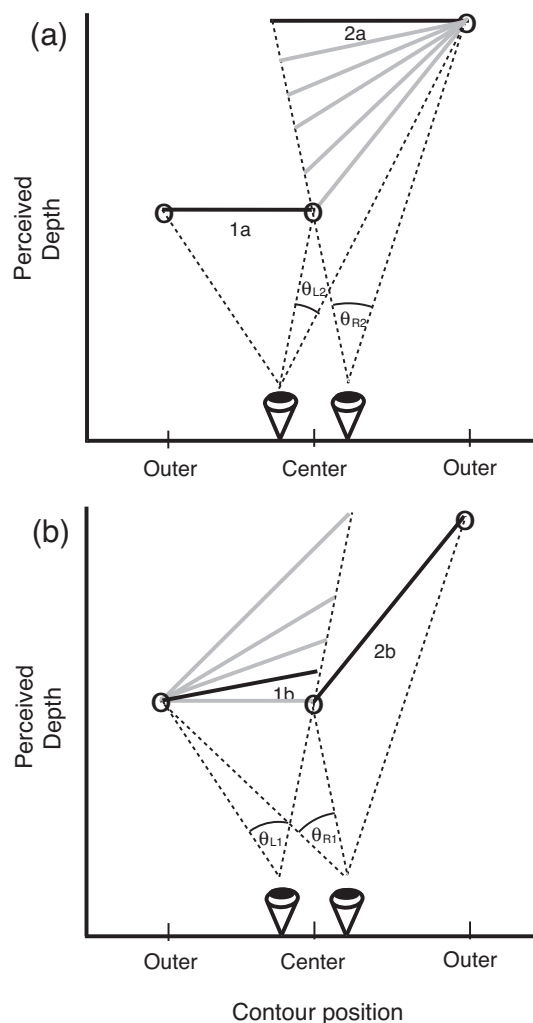


Figure 1. In Peterson and Gibson's (1993) displays, disparity provided unambiguous information about the depth of the contours (marked by circles), but because the regions lacked texture and because ownership of the central edge was not specified, multiple depth interpretations were consistent with the available disparity information. Gray lines and bold lines show some surfaces consistent with the disparity information. All gray surfaces subtend the same angles both in the left and in the right eye: θ_{L2} and θ_{R2} in (a) and θ_{L1} and θ_{R1} in (b). Bold lines show the most frequent depth percept when disparity suggested that the familiar region (region 1a) was in front (a) and when disparity suggested the familiar region (region 2b) was behind (b). Note that the central contour is owned by different surfaces in the two cases: by 1a in (a) and by 2b in (b).

Unfortunately, disparity information in Peterson and Gibson's displays was present only at the luminance edges and was ambiguous because many surfaces in depth were geometrically consistent with the displays (Peterson, 2003). The two regions were different widths in each eye, but because they lacked texture, local determination of a disparity signal was impossible except at the edges. Disparity unam-

biguously specified the position in depth of the central contour and of the two outside edges, but not the ownership of the central contour or the slant of the regions. Such displays are geometrically consistent with either a flat surface extending behind a near surface (see Figure 1a), or a farther surface slanting forward in depth to the central edge (see Figure 1b). Peterson and Gibson's results thus show that configural cues can influence the interpretation of *ambiguous* disparity information, but do not indicate what would happen if disparity information were unambiguous. The present experiments were designed to answer this question.

Experiment 1

To determine whether configural cues influence quantitative depth judgments based on binocular disparity, we constructed displays consisting of two regions covered in random dots, separated by a disparity defined depth step and a central vertical luminance edge whose shape was suggestive of a face in profile. Previous work by Peterson and Gibson (1993, 1994b) found this contour, because of convexity and familiarity, to be highly effective in biasing subjects to select one side as figural in bipartite non-stereoscopic displays. All other known configural cues were equated on both sides of the central edge. We paired the configural cues with disparity cues to create "consistent" stereograms (Figure 2a), in which configural cues and disparity specified the same side as in front (namely, the face side) and "inconsistent" stereograms (Figure 2b) in which they specified opposite sides as in front (the face side was in back). These labels therefore refer to the consistency or inconsistency of the sign of depth (left or right side in front) indicated by each cue.

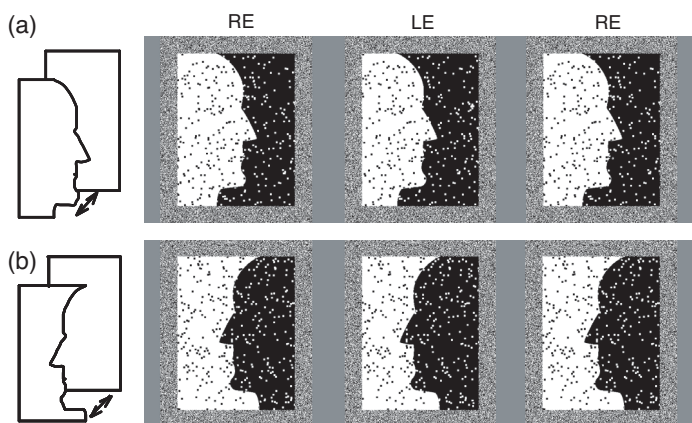


Figure 2. Cross-fuse the left two images or divergently fuse the right two images to see stimulus examples in depth. (a). Consistent stereogram: disparity information and configural cues both indicate the white region to be in front. (b). Inconsistent stereogram: disparity specifies the white region to be in front and configural cues suggest the black region to be in front. The schematic to the left indicates the type of depth percept that should result from the corresponding stereo pairs.

Participants were shown different pairs of consistent and inconsistent displays in a two-interval forced-choice (2IFC) procedure and were asked to select the interval in which they saw greater depth separation between the two regions. We used a one-up/one-down staircase procedure to measure the point of subjective equality (PSE) for a variable comparison stereogram relative to a standard stereogram whose disparity was always 7.5 arcmin. When the standard and the comparison displays were identical (i.e., consistent standard vs. consistent comparison trials and inconsistent standard vs. inconsistent comparison trials), the PSE should converge on the disparity of the standard. When the standard and comparison displays were different (i.e., consistent standard vs. inconsistent comparison trials and inconsistent standard vs. consistent comparison trials), configural cues either should change the PSE, if they have a quantitative effect on metric depth judgment, or should not change the PSE, if they do not have such an effect.

Suppose that the face-shaped side is perceived to be slightly closer than a non-face side would be, given the same disparity. Then subjects should see less depth in an inconsistent display than in a consistent display with the same disparity signal. Hence, with a consistent standard display and an inconsistent comparison display subjects should require *more* disparity for the depth separation in the comparison display to appear identical to that in the standard (PSE > 7.5 arcmin). In contrast, with an inconsistent standard and a consistent comparison, *less* disparity should be required for depth separation in the comparison to appear identical to that in the standard (PSE < 7.5 arcmin). If this result is observed, we will have shown a quantitative effect of configural cues on metric depth perception, suggesting that the face side appears slightly closer due to its configural properties.

Methods

Participants

Thirteen University of California Berkeley students participated. Ten were naïve to the experimental hypothesis. Participants were compensated at the rate of \$10 per hour. All observers had normal stereovision as determined by the Titmus stereo test.

Stimuli and apparatus

Stereoscopic displays showed two adjacent, opaque, equal-area, high-contrast regions covered in random dots and separated by an edge suggesting a face in profile on one side (see Figure 2). Binocular disparity specified that the edge and the dots in the two regions lay at two different viewing distances. In consistent displays, disparity specified the face side as closer than the other region (Figure 2a). In inconsistent displays, disparity specified the face side as farther away (Figure 2b). A frame surrounded the display at

a disparity-specified distance nearer than either of the two regions of interest.

Equal numbers of consistent and inconsistent displays were constructed in which the nearer region was on the right or on the left. The near region, as specified by disparity, was always bright red with black dots. The far region was always black with bright red dots. The frame was a dense random dot field with equal numbers of bright red and black dots so that each region differed from it by equal contrast. The entire stereogram was surrounded by a dim red field, with a luminance equal to the average luminance of the stereogram.

The displays were created offline. Original images were obtained from the OMEFA stimulus set (<http://www.u.arizona.edu/~mapeters/thelab.html>). Two copies were made of the original, one for each eye. Dots of equal size (~1.5 arcmin) were sprinkled randomly with the same density (~250 dots/deg²) on both regions. The disparity signal was created by shifting corresponding regions in each eye's image horizontally by equal amounts. The central edge and the dots from the near stimulus region were defined at the projection plane. The dots from the far stimulus region were defined by disparity as being behind the projection plane. Finally, a frame made of smaller random dots (~15 arcsec) and 2.5 arcmin nearer than the projection plane surrounded the figure-ground regions.

Two stimuli were presented sequentially on each trial, a standard and a comparison. Standard displays always had a depth pedestal of 7.5 arcmin of disparity, which corresponded to ~40 cm at the viewing distance of 3.25 m. The disparity of the comparison varied from 0.5 arcmin to 15 arcmin in 0.5-arcmin steps. (The exact disparity varied slightly depending on each subject's interpupillary distance.) The different depth pedestals were achieved by changing the disparity of the far plane so that observers could not base their judgments on the depth separation between the frame and the near plane.

Displays were presented on a CRT, 28.4-cm high and 38.7-cm wide. Screen resolution was 1600 x 1024 pixels. Each image (frame included) measured 125-mm high (2.2 deg) and 100-mm wide (1.8 deg). The frame was 10-mm wide (0.18 deg visual angle).

CrystalEyes liquid-crystal shutter glasses allowed separate presentation of the left-eye and right-eye images. Images were drawn on alternate frames so that each eye's image was drawn only when the corresponding shutter was open. Each eye's image was re-drawn at 50 Hz. To minimize cross-talk, only the red phosphor was used. The room was otherwise dark.

Procedure

Each trial consisted of two intervals: one containing the standard display and the other containing the comparison display. The interval containing the standard was randomly chosen. Participants were asked to indicate, via key

presses, whether the display in the first or second interval had a greater apparent separation in depth between the right and left sides (2IFC paradigm). No feedback was provided. Each display was presented for 1 s with an inter-stimulus interval of 0.5 s. Intertrial intervals were approximately 0.5 s, although they varied with the subject's response time. A chin rest was used to keep viewing distance constant.

Each participant was exposed to the same eight conditions. Within a condition, the same side (left or right) was specified by disparity to be in front in both intervals. Two "control" conditions (consistent standards versus consistent comparisons and inconsistent standards versus inconsistent comparisons) and two "experimental" conditions (consistent standards versus inconsistent comparisons and inconsistent standards versus consistent comparisons) were independently varied with front side.

The disparity-specified depth separation of the comparison stimulus was varied with a one-up/one-down staircase procedure. This particular reversal rule samples points at or near the 50% point of the psychometric function (Levitt, 1971). Each staircase terminated when it had reversed 12 times. Four staircases were collected for each condition from each participant. For each participant PSE estimates were obtained by performing a maximum-likelihood fit to all the raw psychometric data from a given condition (Wichmann & Hill, 2001a). We report the average of these values across subjects as the condition PSE.¹ We also present the individual subject PSEs.

Participants completed four blocks of trials, each of which contained eight randomly interleaved one-up/one-down staircases, one for each condition. Each block contained 180-220 trials, depending on the speed of convergence. Participants received two staircases of practice, randomly chosen from the eight conditions, which they ran until completion.

Results and discussion

The data were analyzed in a three-factor analysis of variance (ANOVA) with three within-subject factors: condition type (control vs. experimental), standard type (consistent vs. inconsistent), and side in front by disparity (left vs. right). Because there was no main effect of side ($p = .576$) and no interactions of side with other factors [$p = .296$ (standard x side) and $p = .623$ (condition x side)], we pooled the raw data across side and re-fit the psychometric functions. Two subjects were excluded from the analysis because their data were unacceptably noisy. (In both experimental conditions, their 95% confidence intervals exceeded 5 arcmin; whereas no other subjects exceeded 2.5 arcmin in any condition. Excluding the deviant subjects in the analysis did not change the significance of the effects reported in the next paragraph.)

As shown in Figure 3, the results show a quantitative effect of configural cues on depth perception. The ANOVA showed a significant main effect of standard, $F_{(1,10)} = 6.896$; $p < .025$, and a significant interaction between standard and condition, $F_{(1,10)} = 16.719$; $p < .002$. PSEs obtained in the control conditions were close to the pedestal (7.5 ± 0.1 arcmin). In the experimental conditions, the PSE obtained for an inconsistent comparison (mean = 9.21; CI = 7.77-10.65) was greater than that obtained in the control condition (mean = 7.57; CI = 6.83-8.31) and the PSE obtained for a consistent comparison (mean = 5.84; CI = 5.01-6.68) was less than that obtained in the control condition (mean = 7.48; CI = 6.94 - 8.02). Furthermore, experimental PSEs were significantly different from the standard pedestal of 7.5 arcmin, which lies outside the 95% confidence intervals for both means. Thus, with these stimuli, the configural cue was, on average, equivalent to approximately 1.6 arcmin of disparity.

Data from individual subjects are presented in Figure 4. While there is significant intersubject variation, in both experimental conditions 10 of 11 subjects showed the same trend. Error bars are 95% confidence intervals as deter-

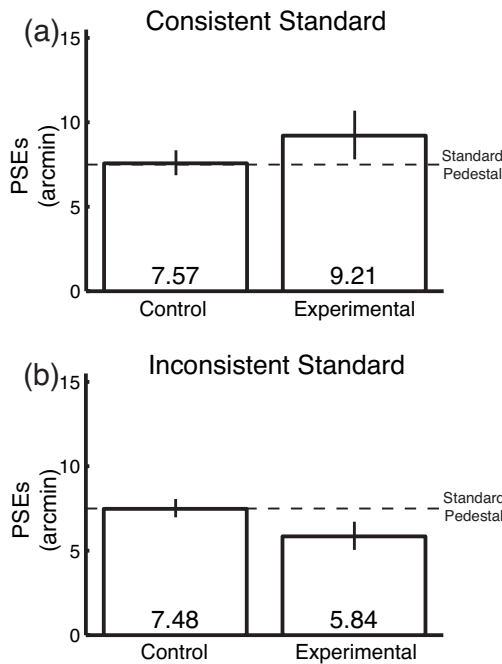


Figure 3. Results from Experiment 1: PSEs from the control and experimental conditions. Error bars are 95% confidence intervals as determined by a two-factor ANOVA. (a). With a consistent standard, more disparity was required (1.6 arcmin) with an inconsistent comparison (experimental condition) than with a consistent comparison (control condition) for the same apparent depth separation. (b). With an inconsistent standard, less disparity (1.6 arcmin) was required with a consistent comparison (experimental condition) than with an inconsistent comparison (control condition) to have the same apparent depth separation. Configural cues were therefore worth approximately 1.6 arcmin of disparity in these displays.

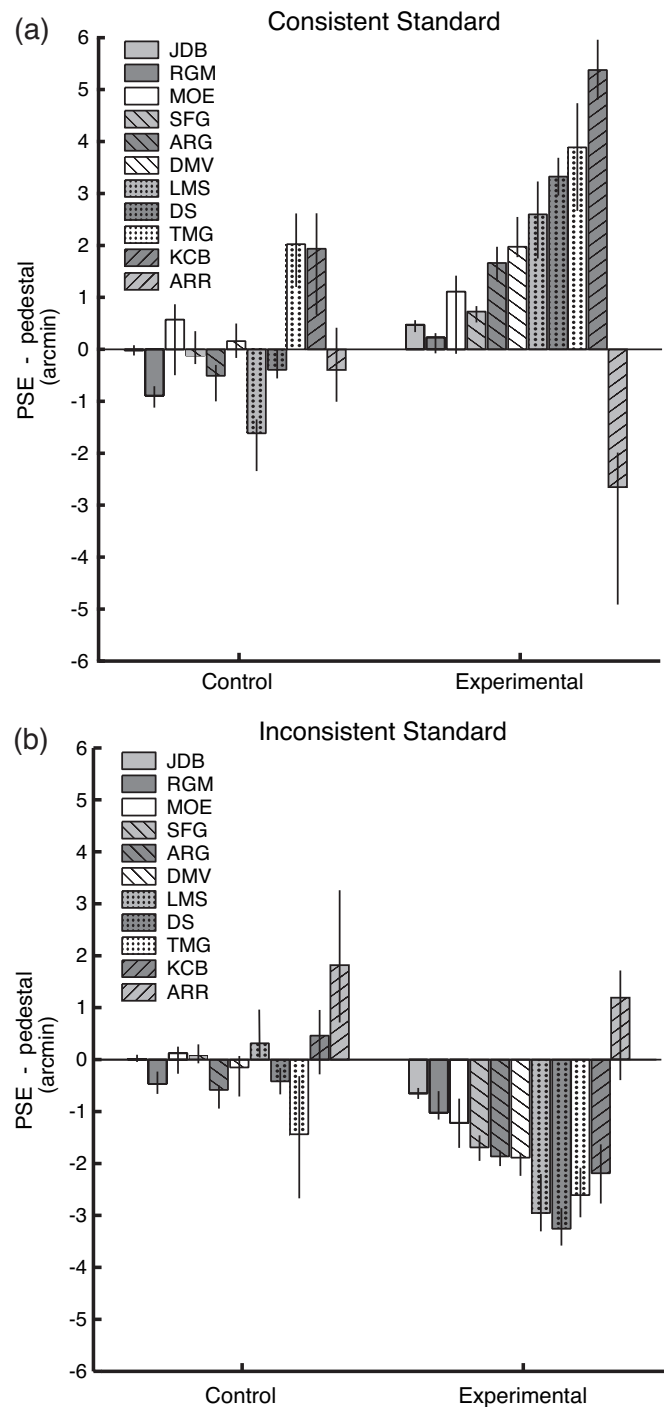


Figure 4. Individual subject PSEs from Experiment 1 plotted as deviations from pedestal. For each condition data from individual subjects are listed from left to right. Error bars are 95% confidence intervals on the PSE estimates from the maximum likelihood fit to each subject's data. (a). PSEs from control and experimental conditions with a consistent standard. (b). PSEs from control and experimental conditions with an inconsistent standard. Ten of the 11 subjects (ARR) included in the analysis trended in the same direction in both experimental conditions. As expected, in the control conditions, PSEs fluctuated randomly on either side of the pedestal value.

mined by Wichmann and Hill's bootstrapping routine (2001b). Although configural cues changed the depth percept by a different amount in each subject, the effect was qualitatively the same in nearly all subjects. Marked inter-subject variability has been well documented in the cue-combination literature (Hibbard, Bradshaw, Langley, & Rogers, 2002; Hillis et al., 2002).

Experiment 2

The results of [Experiment 1](#) are consistent with the hypothesis that configural cues combine with disparity information to affect the depth percept of a metric depth interval geometrically specified by disparity. Configural cues were on average equivalent to approximately 1.6 arcmin of disparity in the unambiguous, stereoscopic displays used. For the same disparity signal, subjects saw less depth in inconsistent displays than in the consistent displays. However, it is not clear whether participants saw more depth in the consistent displays, less depth in the inconsistent displays, or whether both effects were present. Knowing which of these occurred is the first step in understanding the process underlying the combination of configural cues and disparity information. To better determine the structure of the effect, we asked subjects to compare the amount of depth seen in consistent or inconsistent displays relative to a display whose edge shape was neutral with respect to configural cues. We used a sine-wave contour (see [Figure 5](#)) because it equates configural cues in both regions (the two sides being identical except for an 180° rotation).

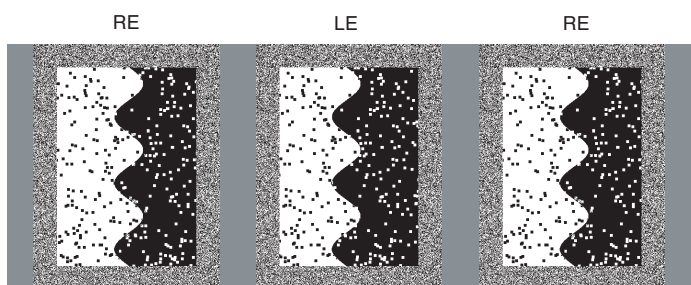


Figure 5. Stereogram with neutral configural cues. Cross-fuse the two left images or divergently fuse the two right images to see an example of the sine-wave stimuli used in [Experiment 2](#). A control experiment was performed with these stimuli to show that the orientation of this contour did not affect the depth percept.

[Experiment 2](#) should reveal whether configural cues can both increase and decrease perceived depth separation depending on how they are correlated with disparity information. Before conducting [Experiment 2](#), we conducted a preliminary experiment to determine whether the sine-wave stimuli were indeed neutral. We found there was no change in perceived depth depending on whether the sine-

wave display's near surface was locally convex or concave at the near surface's uppermost portion (i.e., the orientation of the sine-wave contour). The resulting PSEs converged on values very close to the disparity of the standard (7.5 ± 0.1 arcmin).

Methods

Participants

The participants were nine experienced psychophysical observers, six of whom had participated in [Experiment 1](#). Six were naïve. Three of the subjects from [Experiment 1](#) were still naïve when they ran in [Experiment 2](#). All were compensated at the rate of \$10 per hour.

Stimuli and apparatus

The central edge of the stimuli used in [Experiment 2](#) was either a sine wave or a face contour. The sine-wave edge was locally convex at the near surface's uppermost portion. The face stimuli were the consistent and inconsistent stereograms from [Experiment 1](#).

Procedure

The design of [Experiment 2](#) was similar to that of [Experiment 1](#) except that a face stimulus (consistent or inconsistent) was shown in one of the two intervals while a sine-wave stimulus was shown in the other interval. Each participant was exposed to the same eight conditions. Within a condition, the same side (left or right) was in front in both intervals. Two "sine-wave standard" conditions (paired with either consistent face or inconsistent face comparisons) and two "sine-wave comparison" conditions (paired with either consistent face or inconsistent face standards) were independently varied with front side just as the conditions were in [Experiment 1](#).

Results and discussion

We conducted two separate two-factor ANOVAs, one on the sine-wave comparison conditions and one on the sine-wave standard conditions, and we examined the effects of side and consistency in each. Although there was a main effect of side for the sine standard condition ($p < .029$), there was no main effect in the sine comparison condition ($p < .079$), and side did not interact significantly with either of the other factors [$p = .353$ (sine standard) and $p = .646$ (sine comparison)]. In the interest of keeping our data analysis consistent across experiments and because doing so did not affect the significance of the effect in which we are most interested, we pooled the raw data across side and refit the psychometric functions. We then ran two ANOVAs on the remaining four conditions.

Subjects saw more depth in displays with consistent configural cues and less depth in displays with inconsistent configural cues than in displays with neutral configural cues and the same disparity signal (see Figure 6). A main effect of consistency was obtained both when the sine-wave display was the standard stimulus, $F_{(1,8)} = 14.51$; $p < .002$, and when it was the comparison, $F_{(1,8)} = 41.84$; $p < .0001$. Importantly, the standard pedestal lay outside the 95% confidence intervals of the marginal means for consistency in both ANOVAs. When the sine-wave stimulus was the standard, PSEs for consistent conditions (mean = 6.78; CI = 6.22-7.35) were lower than the standard pedestal of 7.5 arcmin, and PSEs for inconsistent conditions (mean = 9.14; CI = 8.52-9.77) were greater than the standard pedestal. Equivalent results were found when the sine-wave stimulus was the comparison stimulus: PSEs significantly higher (mean = 8.38; CI = 7.81-8.94) than the pedestal were obtained in the consistent conditions and PSEs significantly lower (mean = 6.07; CI = 5.45-6.69) than the pedestal were obtained in the inconsistent conditions. Here, configural cues were worth between 0.7 and 1.6 arcmin of disparity, depending on the condition. We have therefore shown configural cues can both increase and decrease the amount of perceived depth relative to that seen in a neutral display with the same disparity signal.

Individual subject data can be viewed in Figure 7. While there is again significant subject-to-subject variation,

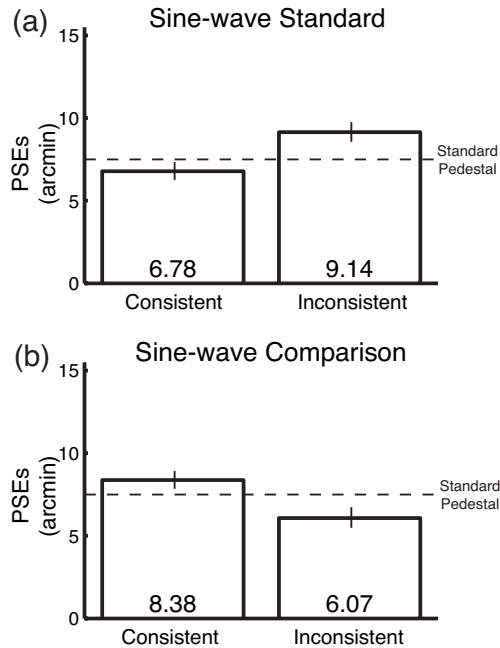


Figure 6. Results from Experiment 2. (a). PSEs from the consistent and inconsistent conditions with a sine-wave standard. (b). PSEs from the consistent and inconsistent conditions with sine-wave comparison. Error bars are 95% confidence intervals as determined by two ANOVAs. Configural cues can both increase and decrease the amount of depth perceived depending on how they are aligned with disparity information.

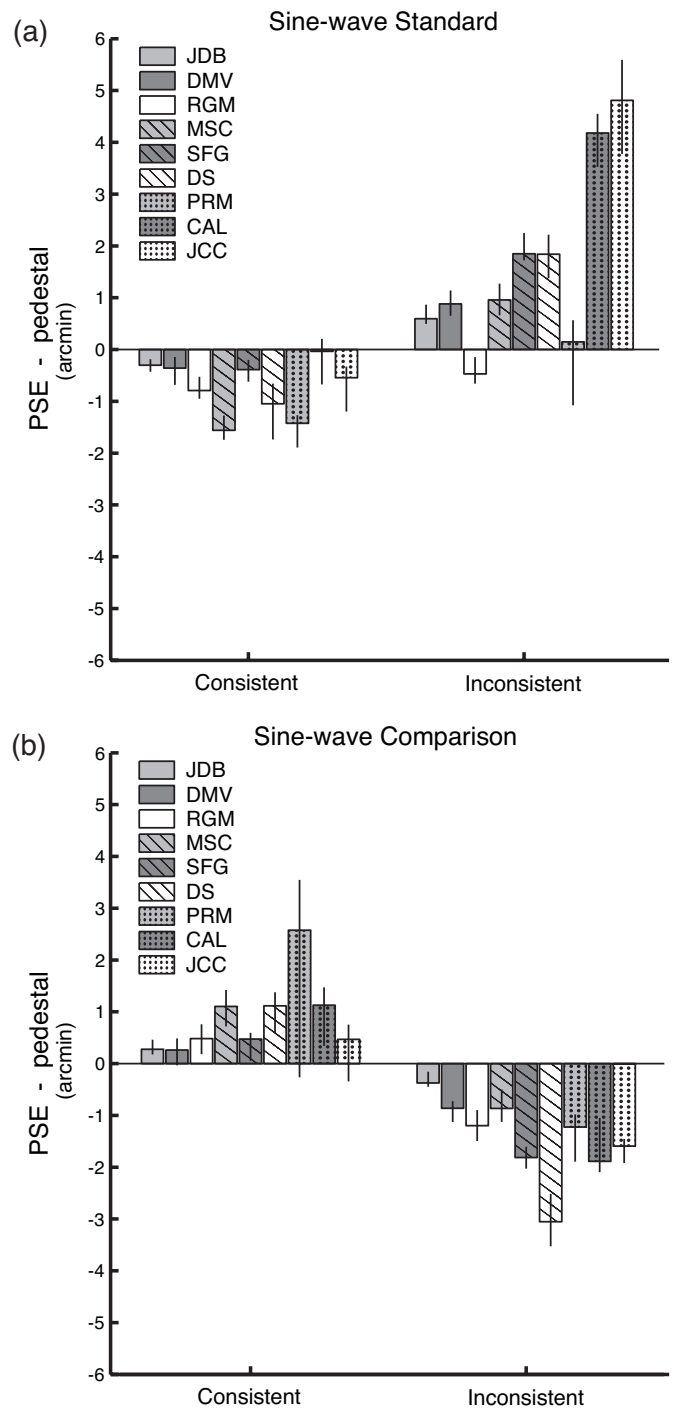


Figure 7. Individual subject PSEs from Experiment 2 plotted as deviations from pedestal. For each condition data from individual subjects are listed from left to right. (a). PSEs from the consistent and inconsistent conditions with a sine-wave standard. (b). PSEs from the consistent and inconsistent conditions with sine-wave comparison. Error bars are 95% confidence intervals on the PSE estimates from the maximum likelihood fit to each subject's data. While there is significant intersubject variability, all nine subjects trended in the same direction on all four conditions, except subject RGM in the inconsistent condition with a sine-wave standard.

in the inconsistent condition with the sine-wave standard, eight of nine subjects trended in the same direction. In the other three conditions, all nine subjects trended in the same direction. The intersubject variation we observed was not unexpected, given the extant literature.

General discussion

Experiment 1 showed that subjects perceive more depth in consistent displays than in inconsistent displays for the same amount of disparity. Experiment 2 showed that subjects saw more depth in consistent displays and less depth in inconsistent displays than they did in neutral displays with the same disparity signal. These results show conclusively, for the first time, that configural cues have a metric effect on depth perception in the presence of unambiguous disparity information.

The metric nature of these effects not only demonstrates that these two geometrically very different cues combine, but poses a more general problem: How can the visual system combine geometrically ordinal cues with unambiguous metric cues to yield metric effects? If Landy et al. (1995) are correct that different cues must be in the same units for combination to occur, standard assumptions about the ordinal nature of the depth information available from configural cues must be altered in some way. Indeed, investigation into how configural cues are interpreted metrically may provide a window into a general process employed by the visual system to bring geometrically ordinal cues into register with geometrically metric cues.

One possibility is that the visual system relies on the accumulation of statistical information about the natural environment. Recent research analyzing range images has shown that not all possible depth values are equally likely in natural scenes (Huang, Lee, & Mumford, 2000). An occlusion relation between surfaces thus implies a nonuniform likelihood distribution of depth values that could, in principle, be built in by evolution or be learned by monitoring the correlation between occlusions and metric depth values specified by disparity and other classic depth cues. Once the statistical likelihood of a metric depth value given a geometrically ordinal depth cue (such as convexity or shape familiarity) has been internalized, it can be combined with metric information from other depth cues within the framework of Bayesian inference.

Could our results be explained by a theory in which cues are *reduced* to the weakest scale before combination takes place (Birnbaum, 1983)? Hel-Or and Edelman (1994) showed that a set of interlocking ordinal depth relationships, recovered from multiple sources at multiple different depths, can converge on a metric representation of space. This theory cannot account for our results, however, because our stimuli (excluding the frame) contained only two surfaces at two depths.

Understanding how configural cues combine with disparity is clearly an important topic for further research. The present results suggest that binary competitive frameworks will not provide an adequate account of figure-ground organization, in so far as it relates to depth perception. Instead, a more quantitative framework is necessary, ideally one that is compatible with the current depth perception literature, perhaps including Bayesian analysis and signal detection theory.

Acknowledgments

We wish to thank Martin Banks, Carmel Levitan, Rob Meyerson, Dhanraj Vishwanath, Laura Walker, and Simon Watt for their helpful suggestions and Martin Banks for his generous contribution of equipment and laboratory space. This research was supported by National Institutes of Health Training Grant in Vision Science T32 EY07043 (JB), National Science Foundation Grant BCS 0425650 (MAP), and National Institutes of Health National Eye Institute Grant R01-EY12851 (equipment).

Commercial relationships: none.

Corresponding author: Johannes Burge.

Email: jburge@berkeley.edu.

Address: 509 Minor Hall, Vision Science Program, University of California, Berkeley, CA 94720-2020.

Footnotes

¹Data were originally analyzed by averaging each staircase's reversal points for a convergence point, averaging the convergence points for the subject PSE, and finally averaging across subjects for the condition PSE. During the review process the concern was expressed that this method might reduce variability among subject PSEs and might artificially increase the chance of obtaining statistical significance. Accordingly, we re-did our analysis. The significance levels of our effects were essentially unchanged.

References

- Backus, B. T., Banks, M. S., van Ee, R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, 39, 1143-1170. [PubMed]
- Birnbaum, M. H. (1983). Scale convergence as a principle for the study of perception. In H. Geissler (Ed.), *Modern issues in perception* (pp. 319-335). Amsterdam: North-Holland.
- Hel-Or, Y., & Edelman, S. (1994). A new approach to qualitative stereo (pp. 316-320). In S. Ullman & S. Peleg (Eds.), *Proceedings of the 12th International Conference on Pattern Recognition*. Washington, DC: IEEE Press.

- Hibbard, P. B., Bradshaw, M. F., Langley, K., & Rogers, B. J. (2002). The stereoscopic anisotropy – individual differences and underlying mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 469-476. [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]
- Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optional cue combination. *Journal of Vision*, 4(12), 1-3, <http://journalofvision.org/4/12/1/>, doi:10.1167/4.12.1. [PubMed][Article]
- Howard, I. P. (2002). *Seeing in depth: Volume 1, Basic mechanisms*. Ontario, Canada: I Porteous Publishing.
- Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth: Volume 2, Depth perception*. Ontario, Canada: I Porteous Publishing.
- 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*. New York: Springer Publishing Co.
- Knill, D. C. (1998). Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Research*, 38, 2635-2656. [PubMed]
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. J. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35, 389-412. [PubMed]
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Optical Society of America*, 49(2), 467-477. [PubMed]
- Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: Bradford Books, MIT Press.
- Palmer, S. E. (2002). Perceptual organization in vision. In *Sensation and perception* (pp. 177-234). In H. Pashler (Ed.), *Stevens handbook of experimental psychology* (Vol. 1, 3rd ed.). New York: Wiley.
- Peterson, M. A. (2003). On figures, grounds, and varieties of amodal surface completion. In R. Kimchi, M. Behrmann, & C. Olson (Eds.), *Perceptual organization in vision: Behavioral and neural perspectives* (pp. 87-116). Mahwah, NJ: LEA.
- Peterson, M. A., de Gelder, B., Rapcsak, S. Z., Gerhardstein, P. C., & Bachoud-Lévi, A.-C. (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 inputs 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., Harvey, E. H., & Weidenbacher, H. L. (1991). Shape recognition contributions to figure-ground organization: Which route counts? *Journal of Experimental Psychology: Human Perception and Performance*, 17, 1075-1089. [PubMed]
- Peterson, M. A., & Skow-Grant, E. (2003). Memory and learning in figure-ground perception. *Cognitive Vision: Psychology of Learning and Motivation*, 42, 1-34.
- Rubin, E. (1958). Figure and ground. In D. Beardslee & M. Wertheimer (Ed. and Trans.), *Readings in perception* (pp. 35-101). Princeton, NJ: Van Nostrand. (Original work published in 1915)
- Sejnowski, T. J., & Hinton, G. E. (1987). Separating figure from ground with a Boltzmann machine. In M. Arbib & A. Hanson (Eds.), *Vision, brain, and cooperative computation*. Cambridge, MA: MIT Press.
- Vecera, S. P., & O'Reilly, R. C. (1998). Figure-ground organization and object recognition processes: An interactive account. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 441-462. [PubMed]
- Vecera, S. P., Vogel, E. K., & Woodman, G. F. (2002). Lower region: A new cue for figure-ground assignment. *Journal of Experimental Psychology: General*, 131, 194-205. [PubMed]
- Wichmann, F. A., & Hill, J. J. (2001a). The psychometric function: I. Fitting, sampling and goodness-of-fit. *Perception & Psychophysics* 63(8), 1293-1313. [PubMed]
- Wichmann, F. A., & Hill, J. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics* 63(8), 1314-1329. [PubMed]