Title: Natural-scene statistics predict the influence of the figure-ground cue of convexity on human depth perception

Abbreviated title: Natural-scene statistics influence human depth perception

Authors: Johannes Burge (corresponding)
360 Minor Hall, School of Optometry
Berkeley, CA 94720-2020
jburge@berkeley.edu

Charless C. Fowlkes
fowlkes@ics.uci.edu

Martin S. Banks
martybanks@berkeley.edu

Number Figures: 9

Supplementary: PDF document with mathematical proof

Material: Figure showing subsets of natural scene statistics

Number Pages: 24

Number of words: Abstract (188), Introduction (532), Discussion (1707) w/ refs.

Keywords: Depth perception, natural-scene statistics, visual perception, Bayesian inference, figure-ground cues, binocular disparity, cue combination, perceptual organization, visual cortex

Acknowledgments: Supported by AOF’s William C. Ezell Fellowship to JB and Research Grants NIH R01-EY12851 and NSF BCS-0617701 to MSB. The authors thank Ahna Girshick, Mike Landy, Jitendra Malik, Steve Palmer, and Mary Peterson for helpful discussion, and Zhiyong Yang, Dale Purves, Brian Potetz, and Tai Sing Lee for making their natural-scene data sets available.
Natural-scene statistics predict the influence of the figure-ground cue of convexity on human depth perception

Johannes Burge¹, Charless C. Fowlkes², Martin S. Banks¹,³
¹Vision Science Program, UC Berkeley, ²Department of Computer Science, UC Irvine, ³Department of Psychology & Wills Neuroscience Institute, UC Berkeley

ABSTRACT

The shape of the contour separating two regions strongly influences judgments of which region is “figure” and which is “ground”. Convexity and other figure-ground cues are generally assumed to indicate only which region is nearer, but nothing about how much the regions are separated in depth. To determine what depth information is conveyed by convexity, we examined natural scenes and found that depth steps across surfaces with convex silhouettes are likely to be larger than depth steps across surfaces with concave silhouettes. In a psychophysical experiment, we found that humans exploit this correlation. For a given binocular disparity, observers perceived more depth when the near surface’s silhouette was convex rather than concave. We estimated the depth distributions observers used in making those judgments: They were similar to the natural-scene distributions. Our findings show that convexity should be reclassified as a quantitative depth cue. They also suggest that the dichotomy between metric and non-metric depth cues is a false one and that many cues should be evaluated with respect to natural-scene statistics. Finally, the findings provide an ecological explanation for why figure-ground cues modulate disparity-sensitive cells in visual cortex.

Key words: Depth perception, natural-scene statistics, visual perception, Bayesian inference, figure-ground cues, binocular disparity, cue combination, perceptual organization, visual cortex
INTRODUCTION

Estimating three-dimensional (3D) layout from the two-dimensional retinal images is geographically under-determined, but biological visual systems solve the estimation problem nonetheless. This “inverse optics problem” has generated interest for centuries (Berkeley, 1709; von Helmholtz, 1867). The current view is that perceived depth is computed from features in the retinal images that provide information about depth in the scene. Theoretical and empirical research of these so-called depth cues has yielded an extensive taxonomy based on geometric analyses of the information they provide (Palmer, 1999; Bruce, Green, Georgeson, 2003). In this taxonomy, some cues, like binocular disparity, are called metric depth cues because they allow the absolute distance between two points in the scene to be recovered via simple trigonometry. Other cues, such as those indicating occlusion, are called ordinal depth cues because they do not directly indicate the distance between two objects; consequently, they are said to provide only “information about ordering in depth, but no measure of relative … distance” (Bruce et. al., 2003).

To estimate depth as accurately and precisely as possible, all relevant information should be used. Experimental evidence suggests that the visual system combines information from multiple metric depth cues in a statistically optimal fashion (Knill & Saunders, 2003; Hillis, Watt, Landy, Banks, 2004). However, it is unclear how information from ordinal and metric cues should be combined (Landy, Maloney, Johnston, Young, 1995). Ordinal information constrains only the sign of depth between pairs of surfaces, so the ordinal cue is either consistent with the metric cue and provides no additional numerical information, or the cues are inconsistent and it is not obvious how to resolve the conflict. And yet, the shape of an occluder’s silhouette affects the depth perceived from disparity; that is, the imaged shape of the occluding object changes the amount of depth that is perceived between the occluder and the background (Burge, Peterson, Palmer, 2005; Bertamini, Martinovic, Wuerger, 2008).

This counter-intuitive result can be understood by considering the depth information potentially provided by the involved depth cues; perhaps it will be productive to consider the statistical relationship between each cue and its associated distribution of depths in the natural environment (Brunswik & Kamiya, 1953; Hoiem, Efros, Hebert, 2005; Saxena, Chung, Ng, 2005). Depth can be directly estimated from disparity although disparity provides less information than suggested from the viewing geometry because uncertainty arises due to noise in disparity measurements (Cormack, Landers, Ramakrishnan, 1997) and in other signals required to interpret measured disparities (Backus, Banks, van Ee, Crowell, 1999). Now consider cues to occlusion. There is no a priori geometric reason that the shape of an image region should provide metric depth information, but we propose that the convexity of an image region (i.e., the convexity of a silhouette) is statistically correlated with depth in natural viewing, and therefore that convexity imparts information about metric depth. If such a relationship exists, any system (human, animal, or machine) could exploit it when estimating depth. We measured natural-scene statistics and found a systematic relationship between convexity and depth. In a parallel psychophysical study, we found that humans use this relationship when estimating 3D layout. These results demonstrate a useful and previously unrecognized role for natural-scene statistics in depth perception.

METHODS FOR NATURAL-SCENE ANALYSIS

To evaluate the hypothesis that image-region shape provides useful metric depth information, we investigated the relationship between figure-ground cues and depth in natural scenes. Depth is defined as the difference in viewing distance for two points along a line of sight. We focused the analysis on the figure-ground cue of convexity because convexity can be readily measured in images of natural scenes (Fowlkes, Martin, Malik, 2007) and because convexity is known to affect figure-ground assignment in human perception (Metzger, 1953; Kanizsa & Gerbino, 1976).
We measured the joint statistics of convexity and depth in a collection of indoor and outdoor scenes. To compute these statistics, we analyzed ~450,000 contour points sampled from a collection of luminance and range images (spatially co-registered RGB and time-of-flight laser-range data) from two published studies (Potetz & Lee, 2003; Yang & Purves, 2003a). Fig. 1a is an example of a luminance image and Fig. 1b is the corresponding range image. Many of the luminance images in the two published reports were noisy and blurred, so we selected a well-lit and sharp subset of 35 images (23 from Yang & Purves, 2003a and 12 from Potetz & Lee, 2003) that had localizable, high-contrast contours in a variety of indoor and outdoor scenes. To identify a representative set of region-bounding contours, five people, who were naïve to the experimental hypotheses, hand-selected contours in the luminance images using a previously published procedure (Martin, Fowlkes, Tal, Malik, 2001). The contour selectors were instructed to mark all image-region boundaries in the luminance image that were “important”. Using a computer mouse and software tool, they segmented each image into disjoint regions. The tool allowed them to zoom in and refine boundaries until they were satisfied with the result. The contour selectors generally did not mark shadows or the contours of image regions subtending less than 2.9° (~20 pixels). To assess the consistency of segmentation, two people marked each image. The agreement in their segmentations was measured using the Local Consistency Error measure (Martin et. al., 2001). The median LCE was 0.13, which is slightly larger than previously reported values of 0.08 (Martin et. al., 2001).
Figure 1. Luminance and range images. a) An example luminance image. b) Corresponding range image. Blue indicates nearer distances, yellow intermediate distances, and red farther distances. c) Close-up of the luminance image with a representative hand segmentation overlaid. d) Close-up of the associated range image with the same segmentation overlaid. e) The same image as d, but with convexity flags added. Flags point towards the image region that was classified as more convex. Longer flags correspond to larger convexity values.

For each point along a selected contour, we computed the convexity of the local image regions on either side of the contour using a previously published measure of local convexity (Fowlkes et al., 2007). A circular analysis window of fixed radius (2.16°, 15 pixels) was centered at each point along a selected contour. We sampled pairs of points inside both regions and recorded the fraction of pairs for which a line segment connecting them lay completely within the region. Convexity, c, is the log ratio of the two fractions. Fig. 1e shows convexity along each selected contour in the expanded region from Fig. 1a. Finally, we examined the frequency of different depths as a function of region convexity.

RESULTS OF NATURAL-SCENE ANALYSIS

Fig. 2a shows the distribution of convexities for the contours we analyzed. Fig. 2b shows the statistical relationship between convexity and depth, where depth is the difference in the viewing distances of the two surfaces on opposite sides of a contour. The most likely depth is approximately zero because many contours correspond to rapid changes in surface orientation or changes in reflectance due to surface markings. Nonetheless, depth is clearly correlated with convexity. For all positive depths, a region is always more likely to be near than far if its bounding contour makes it convex. Put another way, if the occluding surface is convex, the distance between the occluding and the occluded surface is likely to be larger than if the contour is concave.
METHODS FOR PSYCHOPHYSICAL EXPERIMENTS

We conducted three psychophysical experiments. Two experienced observers participated in Experiment 1. Seven and ten observers participated in Experiments 2 and 3, respectively. All had normal visual acuity and stereopsis. With the exception of one observer in Experiment 3, all were naïve to the experimental hypotheses.

The stimuli were presented on a CRT (28.4×38.7 cm; 1600×1024 pixels) and subtended 4.4×3.6°. The two eyes were stimulated separately with liquid-crystal shutter glasses (CrystalEyes®, StereoGraphics, Inc.) synchronized to the display. The frame rate was 100 Hz, so each eye’s image was refreshed at 50 Hz. To minimize crosstalk between the two eyes’ images, only the red phosphor was used. The room was otherwise dark. Viewing distance was 325 cm in all but the second condition of Experiment 3 in which it was 85 cm. The long viewing distance was used for most of the experiments because it is associated with a low reliability of depth from disparity (and therefore makes it easier to see an effect of convexity) and because it minimized the
reliability of focus cues (and therefore avoided contamination by such cues) (Watt, Akeley, Ernst, Banks, 2005). The observer’s head position was stabilized with a chin and forehead rest. The stimuli were two adjacent equal-area regions textured with randomly positioned dots (~250 dots/deg$^2$), separated by a luminance contour that was either curved (radius of curvature = 10.5 cm) or straight (Fig. 3). One region was black with red dots and the other was red with black dots. Disparity specified that one region was in front of the other. The contour had the same disparity as the nearer region. There were three kinds of stimuli—consistent, inconsistent, and neutral. In “consistent” stimuli, the near region, as specified by disparity, was made convex by the contour (Fig. 3a). For “neutral” stimuli, the contour was a vertical line (Fig. 3b). In “inconsistent” stimuli, the disparity-specified near region was concave (Fig. 3c). A textured frame surrounded the two regions; it aided binocular fusion and had a crossed disparity of 2.5 arcmin relative to the nearer region. Other known figure-ground cues (e.g., size, surroundedness (Rubin, 1921), contrast (Driver & Baylis, 1996), lower region (Vecera, Vogel, Woodman, 2002) were equated on both sides of the contour. Stimulus features that could not be equated, such as brightness, were the same in the two stimuli presented on each trial. With this design, we can attribute changes in perceived depth to changes in region convexity.

**Figure 3.** Examples of the experimental stimuli. The upper row contains stereograms that can be cross-fused to reveal two regions at different distances. The lower row depicts the disparity-specified depth for each stimulus type. a) “Consistent” stimulus: disparity and convexity both indicate that the white region is nearer than the black region. b) “Neutral” stimulus: disparity specifies that the white region is nearer, while convexity does not indicate which region is nearer. c) “Inconsistent” stimulus: disparity specifies that the white region is nearer and convexity suggests that it is farther. A reader who examines the stereograms closely might perceive a difference in depth separation between the different stimuli. But such an effect might not be apparent because, as our data show, the perceptual effect is small.

**Experiment 1**

Two stimuli were presented sequentially on each trial: a standard and comparison. Observers indicated the stimulus in which they perceived more depth between the red and black regions. The disparity of the standard was fixed at one of eight values: 2.5–20 arcmin in 2.5-arcmin steps. These disparities corresponded to simulated depths of ~12–157 cm. The disparity of the comparison varied about the standard disparity in 0.5-arcmin steps; the minimum and maximum values were 0.5 and 30 arcmin. Only the disparity of the far region changed, so observers could not base judgments on the depth separation between the frame and the near region. We used a 2-interval, forced-choice procedure. The standard and comparison were presented for 1 sec each in random order with an
inter-stimulus period of 0.5 sec. The disparity-specified near region was on the left for half the trials (i.e., it was on the left for both intervals on a given trial) and the right for the other half. Observers indicated the interval that contained the greater apparent separation in depth between the near and far regions. No feedback was provided. Five stimulus pairings were presented at each of the eight standard disparities: neutral-consistent (NvC), neutral-inconsistent (NvI), neutral-neutral (NvN), consistent-consistent (CvC), and inconsistent-inconsistent (IvI). The set of judgments for a given stimulus pairing and standard disparity yielded a psychometric function: the percentage of trials in which the comparison stimuli was said to have greater depth as a function of its disparity-specified depth. The disparity-specified depth separation in the comparison stimulus was varied with 1-down/2-up and 2-down/1-up staircase procedures. These staircases tend to sample points near the 29% and 71% points of the psychometric function (Levitt, 1971). Each staircase terminated after 12 reversals. Four staircases were run for each condition for both observers. The observers completed 40 blocks of trials. Each block presented all five conditions randomly interleaved at one standard disparity. Blocks consisted of 320-450 trials. We observed no systematic effect for left-vs-right positioning of the near side, so left and right data were combined. This resulted in ~320 observations per psychometric function. Psychometric data from all 40 conditions (5 pairings x 8 standard disparities) were used in the subsequent analyses. For each observer and condition, we obtained estimates of the point of subjective equality (PSE)—the disparity value of the comparison stimulus that on average had the same apparent separation in depth as the standard stimulus—and estimates of the just-noticeable difference (JND) defined as the 84% on the psychometric function—by fitting a cumulative Gaussian to the psychometric data (Wichmann & Hill, 2001).

Experiment 2

Seven observers participated in Experiment 2. The stimuli and procedure were the same as in Experiment 1 with the following exceptions. Instead of convex, concave, and straight contours, there were only convex and concave contours. Instead of five stimulus conditions, there were four: consistent-consistent, inconsistent-inconsistent, consistent-inconsistent, and inconsistent-consistent. The standard stimulus had only one value: 15 arcmin. A 1-down/1-up staircase procedure was used to estimate the PSE: the disparity value of the comparison stimulus that had on average the same apparent depth as the standard stimulus. There were ~200 trials per condition for each observer.

Experiment 3

This experiment was identical to Experiment 2 with two exceptions. 1) We used the method of adjustment instead of the 2-interval, forced-choice procedure. We did so to check for the possibility that response bias contaminated the results of Experiments 1 and 2 (this issue is explained in the Results). 2) The stimuli were presented at two viewing distances (325 and 85 cm) instead of one (325 cm). We used the two distances to determine how changes in the reliability of depth from disparity (greater at 85 than at 325 cm) affected the influence of convexity.

RESULTS OF PSYCHOPHYSICAL EXPERIMENTS

In three experiments, we investigated whether humans exploit the information available from image-region convexity when estimating metric depth. Fig. 4 illustrates the expected influence of contour shape on depth judgments. The upper row depicts the probability distributions associated with consistent stimuli (Fig. 3a) in which convexity and depth both indicate that one side is nearer. The posterior distribution, given by the product of the distributions in the left panel, is shifted slightly toward less depth than specified by disparity. The lower row depicts the probability distributions associated with inconsistent stimuli (Fig. 3c). The product of the two distributions in the left panel is now shifted toward less depth than occurs with consistent stimuli. Therefore, the observer should perceive less depth with inconsistent stimuli (concave-silhouetted occluders) than
with consistent stimuli (convex-silhouetted occluders) when the disparity-specified depth is the same.

![Figure 4](image.png)

**Figure 4.** Predicted depth percepts for different combinations of convexity and disparity. The stimulus is composed of two regions separated by a contour. The left panels show probability distributions associated with convexity and disparity expressed over depth. The abscissa is the depth between the regions on the two sides of the contour. Positive numbers indicate that the putative figural region is closer than the opposing ground region. The blue and red curves are schematics representing the probability distributions derived from the natural-scene statistics (Results, Fig. 2) for convex and concave reference regions, respectively. The black curves represent the distribution over depths derived from an intrinsically noisy disparity signal that specifies that the assigned region is nearer than the opposing region. \( f^{-1}(\cdot) \) maps disparity signals into depth (Eqn. 5). The right panels show the posterior distributions (solid curves) and the disparity likelihood functions (dashed curves) for the same two situations. The posterior distributions are shifted relative to the disparity likelihood functions by different amounts depending on the convexity-depth probability distribution.

In addition, we expect that the influence of convexity on perceived depth should increase with increasing disparity. This expectation is based on two well-established findings: First, for an ideal observer, cues’ influences depend on their variances; if the variance of one cue increases, the influence of other cues will increase (Ghahramani, Wolpert, Jordan, 1997). Humans behave similarly in many perceptual tasks (Knill & Saunders, 2003; Hillis et. al. 2004; Ernst & Banks, 2002; Alais & Burr, 2004). Second, the variance of depth from disparity increases as the disparity increases; that is, discrimination thresholds increase in proportion to disparity (McKee et. al., 1990; Morgan et. al., 2000). As a result, the difference in the posteriors on the right side of Fig. 4 should increase as disparity increases: i.e., the influence of convexity on human depth judgments should increase as the disparity in the stimulus increases.

**Experiment 1**

The results from Experiment 1 are shown in Fig. 5. The PSE changes in the two panels of Fig. 5a show that to produce the same perceived depth, consistent stimuli required less disparity than neutral stimuli and inconsistent stimuli required more. For example, at a standard disparity of 15 arcmin, EKK and JIL respectively needed 1.1 and 3.9 arcmin less disparity in consistent than in inconsistent stimuli to perceive the same depth. This effect is consistent with observers using the correlation between contour convexity and depth in judging the separation of the near and far regions in our stimuli. The effect of contour shape increased systematically with increasing...
disparity. This is also expected because the reliability of depth from disparity decreases with increasing disparity (McKee, Levi, Bowne, 1990; Morgan, Watamanuik, McKee, 2000) so the influence of contour shape should increase with increasing disparity. The results thus suggest that contour shape provides metric depth information to human observers.

Fig. 5b shows that just-noticeable differences (JNDS) rose systematically with increasing disparity, as expected, but did not vary significantly across consistent, inconsistent, and neutral conditions. Interestingly, convexity had a larger influence in the observer (JIL) with higher discrimination thresholds, the expected result if both observers had internalized the same natural-scene statistics for convexity and depth.

**Figure 5.** Results from the main experiment. Upper and lower rows show data from observers EKK and JIL, respectively. Blue indicates the neutral-consistent stimulus pairing, red the neutral-inconsistent pairing, and black the neutral-neutral pairing. **a)** Point of subjective equality (PSE) minus standard disparity is plotted as a function of the disparity in the standard stimulus. This is the disparity increment of the comparison stimulus, relative to the standard disparity needed to match the perceived depth in the standard. The symbols represent the mean of the cumulative Gaussian that best fit the raw psychometric data in each condition. Error bars represent 95% confidence intervals of the mean. Dotted lines represent predictions of the non-parametric model. Solid lines represent the predictions of the power-law model. If convexity did not affect perceived depth, the data would lie on a horizontal line through zero. The PSE data show that consistent
stimuli needed less disparity and that inconsistent stimuli needed more disparity than neutral stimuli to yield the same apparent depth. \textbf{b}) Just-noticeable differences (JNDs) plotted against disparity in the standard. This is the disparity difference that was required for observers to respond that the comparison stimulus had greater depth than the standard 84% of the time. Symbols represent the standard deviation of the best-fitting cumulative Gaussian for each condition. Error bars represent 95% confidence intervals on the standard deviation. Dotted and solid lines are the predictions of the best-fitting probabilistic models (non-parametric and power law, respectively).

\textbf{Experiment 2}

To make sure that the observations of Experiment 1 would generalize, we performed a shorter version of the experiment with more observers. Fig. 6a shows that the seven observers exhibited the same pattern of results as the observers in Experiment 1. Thus, the effect is observed in many people. Fig. 6b shows the results averaged across observers. On average, consistent stimuli required about 2.1 arcmin less disparity than inconsistent stimuli to produce the same apparent depth. We conducted a repeated-measures, two-factor ANOVA with the standard stimulus type (consistent or inconsistent) as one factor and the relationship between the standard and comparison stimuli as the other factor (control: conditions in which the standard and comparison were the same: consistent vs consistent and inconsistent vs inconsistent; or experimental: conditions in which the standard and comparison were different: consistent vs inconsistent and inconsistent vs consistent). The interaction between the factors was highly significant ($F_{(1,6)} = 34.8; p < 0.0011$). A multiple-comparisons test showed that the difference between consistent vs consistent and consistent vs inconsistent conditions and between inconsistent vs inconsistent and inconsistent vs consistent conditions were both significant.
Figure 6. Results from Experiments 2 and 3. a) PSEs from individual observers in Experiment 2. Four stimulus pairings were presented: consistent-consistent (CvC), consistent-inconsistent (CvI), inconsistent-inconsistent (IvI), and inconsistent-consistent (IvC). The first member of each pair is the standard stimulus; the second is the comparison. PSE minus standard disparity is plotted for each observer: the disparity increment of the comparison stimulus, relative to the standard disparity, that on average yielded the same apparent depth as the standard stimulus. Error bars represent 95% confidence intervals. The dashed horizontal lines through zero represent the expected data if region convexity did not affect perceived depth. b) PSEs from Experiment 2 averaged across observers minus the standard disparity for the four pairings. To perceive depth separation as the same, observers needed more disparity in inconsistent comparison stimuli than in consistent standard stimuli. The reverse was true for consistent comparisons and inconsistent standards. Error bars represent one standard deviation of the group mean. c) Results of Experiment 3 with a viewing distance of 325 cm. PSEs minus the standard disparity averaged across observers are plotted for the same four conditions as in a, b. d) Results of Experiment 3 with a viewing distance of 85 cm.

**Experiment 3**

Experiment 3 tested the possibility that a response bias, rather than a change in perceived depth, was responsible for the effects we observed in Experiments 1 and 2 (Gillam, Anderson, Rizwi, 2009). In a 2-interval, forced-choice procedure like ours, observers must choose a stimulus interval even if they are uncertain about which interval contained more depth. Perhaps observers in this uncertain state chose the stimulus in which the ordinal depth signaled by convexity was compatible with the depth ordering specified by disparity. Such a strategy could yield shifts in the PSEs similar to those in Fig. 5a and Fig. 6a, b. We circumvented this problem by using the method of adjustment (Gillam, Anderson, Rizwi, 2009): observers adjusted the disparity in the comparison stimulus until it appeared to have the same depth as the standard stimulus. By using this technique, we made sure that observers were really equating the perceived depth in the consistent and inconsistent stimuli. Fig. 6c shows the results for the 325-cm viewing distance. On average, consistent stimuli required about 2.4 arcmin less disparity than inconsistent stimuli to yield the same perceived depth. A two-factor, repeated-measures ANOVA with the same factors as in Experiment 2 revealed a significant interaction ($F_{1,9} = 8.68; p < 0.0163$). Eight of ten subjects required inconsistent stimuli to have more disparity than consistent stimuli. Fig. 6d shows the results for the 85-cm distance. Consistent stimuli now required 1.6 arcmin less disparity than inconsistent stimuli. A three-factor, repeated-measures ANOVA revealed a marginally significant three-way interaction between standard stimulus type, standard/comparison relationship, and viewing distance ($F_{1,9} = 4.50; p < 0.063$). The trend towards a smaller effect at the near viewing distance was expected because disparity-specified depth is less reliable at long distance so convexity-specified depth should have more influence at 325 than at 85 cm. The difference, however, was not significantly smaller than the effect size at the far distance.

Thus, we observed the same pattern of results with the method of adjustment in Experiment 3 as we had with the forced-choice procedure in Experiments 1 and 2: Stimuli in which contour shape and disparity were consistent required less disparity to yield the same perceived depth as stimuli in which contour shape and disparity were inconsistent.

**METHODS FOR MODELING OF PSYCHOPHYSICS**

The results from the three psychophysical experiments are qualitatively consistent with the behavior expected of an ideal observer that has incorporated the natural-scene statistics associated with region convexity and depth. Specifically, the results show that convexity has an effect on the amount of perceived depth.

To further examine the relationship between the psychophysical results and the results of the natural-scene analysis, we fit the observers’ responses with a probabilistic model of depth estimation to determine the probability distributions that provided the best quantitative account of those responses. To do so, we modeled the computation of the depth percept as a probabilistic process. Assuming conditional independence, Bayes’ rule states:
That is, the posterior probability of a particular depth, $\Delta$, given a measurement of disparity, $d$, across an contour bounding an image region with convexity, $c$, is equal to the product of the likelihood of the disparity measurement for a particular depth, $P(d | \Delta)$, the likelihood of the convexity measurement for a particular depth, $P(c | \Delta)$, the prior distribution of depths, $P(\Delta)$, and a normalizing constant, $P(d, c)$, for all such contours. We can combine the latter two terms in the numerator into the joint probability of $c$ and $\Delta$:

$$P(\Delta | d, c) = \frac{P(d | \Delta)P(c | \Delta)P(\Delta)}{P(d, c)}$$  \hspace{1cm} (1)$$

Rearranging yields:

$$P(\Delta | d, c) = \left( \frac{1}{P(d, c)/P(c)} \right)P(d | \Delta)P(\Delta | c)$$  \hspace{1cm} (2)$$

where $\frac{1}{P(d, c)/P(c)}$ is a normalizing constant, $P(d | \Delta)$ is the disparity likelihood, and $P(\Delta | c)$ is the convexity-depth distribution: i.e., the distribution we measured in the natural-scene luminance-range images. Note that the depth prior, $P(\Delta)$, has been absorbed by $P(\Delta | c)$.

In this model, the convexity and disparity of an image region both affect the expected perceived depth. This is illustrated schematically in Fig. 4, which shows probability distributions associated with image-region convexity and disparity (left) and the posterior distributions generated from the products of those distributions (right). The disparity signal indicates that the region on one side of the contour is nearer than the region on the other side. If the region is bounded by a contour that makes it convex (upper left), the convexity-depth distribution is skewed toward larger depths than when the region is concave (lower left). As a consequence, the estimated depth will be greater in the convex (upper right) than in the concave case (lower right). Said another way, when region convexity and disparity are consistent with one another (i.e., both indicate that the same region is nearer than the other), perceived depth should be greater than when they are inconsistent.

On each trial, observers selected the stimulus (standard or comparison) with greater perceived depth. Those responses generated the psychometric data. To incorporate the responses in the model specification, we computed posterior distributions for the standard and comparison stimuli on each trial. We then used signal-detection theory (Green & Swets, 1966) to model the discrimination and predict psychometric functions. These functions were the cumulative probability that the comparison stimulus was perceived as having more depth between its regions than the standard stimulus:

$$P(\hat{\Delta}_{\text{comp}} > \hat{\Delta}_{\text{std}}) = \int_0^\infty P(\hat{\Delta}_{\text{comp}} | d_{\text{comp}}, c_{\text{comp}}) \hat{\Delta}_{\text{comp}} P(\hat{\Delta}_{\text{std}} | d_{\text{std}}, c_{\text{std}}) d\hat{\Delta}_{\text{std}} d\hat{\Delta}_{\text{comp}}$$  \hspace{1cm} (3)$$

We thus sought distributions $P(d | \Delta)$ and $P(\Delta | c)$ that minimized the squared difference between the model predictions and the raw psychometric data for all comparison and standard disparities across all experimental conditions:

$$\min \sum_i \sum_j \sum_k N_{ijk} \left[ P(\hat{\Delta}_{\text{comp}}^i > \hat{\Delta}_{\text{std}}^j) - y_{\text{data}}(\hat{\Delta}_{\text{comp}}^i > \hat{\Delta}_{\text{std}}^j) \right]^2$$  \hspace{1cm} (4)$$

where $i$ indexes the five stimulus pairings, $j$ the eight standard disparities, $k$ the comparison disparities visited in each condition, $y_{\text{data}}$ is the proportion of times the comparison was selected at that level, and $N$ is the number of observations at that level. We fit this model to the raw
psychometric data using non-linear optimization to minimize the squared deviation between each observer’s responses and the model’s predictions across all measurements in all 40 conditions.

To make the fitting procedure tractable, we needed approximations of the probability distributions corresponding to each cue. To parameterize the probability distributions, we mapped depth, \( \Delta \), into disparity using:

\[
f(\Delta) = 2\left(\tan^{-1}(1/2v) - \tan^{-1}(1/2(v + \Delta))\right) \tag{5}
\]

where \( v \) is the viewing distance and \( l \) is the inter-pupillary separation (Howard & Rogers, 2002). Previous research on disparity distance and \( l \) is the inter-pupillary separation (Howard & Rogers, 2002). Previous research on disparity discrimination suggests that the disparity distributions can be approximated by Gaussians with standard deviations proportional to their means plus a constant (McKee et. al., 1990; Morgan et. al., 2000), so we modeled the disparity likelihood as:

\[
P(d \mid \Delta) = N(f(\Delta), \sigma(\Delta)) \tag{6}
\]

where \( f(\Delta) \) is the mean disparity signal, \( d \), for a given depth, and the standard deviation, \( \sigma(\Delta) \), is proportional to the mean. Specifically, we assumed:

\[
\sigma(\Delta) = af(\Delta) + \sigma_0 \tag{7}
\]

where \( a \) is the rate at which the standard deviation increases with increasing disparity, and \( \sigma_0 \) is a small value that corresponds to the standard deviation when the disparity is zero. Thus, the disparity likelihoods were characterized by two parameters for all experimental conditions.

For the convexity-depth distributions, we used a non-parametric model that approximated them as piecewise log-linear. We chose this model (similar to Stocker & Simoncelli, 2006) because it makes very few assumptions about the shape of the distributions. In this model, the convex, concave and straight-contoured convexity-depth distributions were approximated with eight segments that were piecewise log-linear as a function of disparity. We used one segment for each disparity of the standard stimulus. The shapes of the three convexity-depth distributions were therefore specified by 24 parameters. The non-parametric model requires one of the 24 local slopes to be set by the modeler because the relative shift between posteriors at a given disparity is determined by the difference between the local slopes at that disparity; any two local slopes with the same difference yield the same relative shift. We set the slope of the straight-contoured distribution for the smallest standard disparity to ensure that the curves were integrable and ranged between \( 10^{-1} \) and \( 10^{10} \).

We also modeled the convexity-depth distributions as power laws. This has the advantage of reducing the number of free parameters necessary to model the convexity-depth distributions from 24 to three. In the power-law model, we assumed that the convexity-depth distribution followed a power law over depth. Three parameters, the exponents for each of the three convexity-depth distributions, determined the shapes of those distributions. For the power-law model, we set the power of the straight-contoured distribution based on the best-fit, straight-contoured distribution estimated with the non-parametric model.

To approximate the posterior, we evaluated \( P(\hat{\Delta}_{\text{comp}} > \hat{\Delta}_{\text{std}}) \) which requires evaluating the product of the disparity-likelihood and convexity-depth distributions. The distributions must be expressed over common units. We expressed the distributions over disparity for computational convenience. We modeled the log of the convexity-depth distribution as locally linear in the region of a given disparity likelihood, which is justified if the distribution changes slowly across that region. Log probability is described locally by:

\[
\ln\left(P(\Delta \mid c)\right) = m(c, \Delta)\Delta + b(c, \Delta) \tag{8}
\]

where \( c \) is the convexity cue, \( m \) is the local slope, and \( b \) is a scaling factor that ensures that the distribution is continuous and integrates to one. For the piecewise log-linear model, the local slope
is simply read out from the parameters. The local slope of a power law in log-probability space is given by:

$$m(c, \Delta) = \frac{k}{\Lambda}$$  \hspace{1cm} (9)

where $c$ is the convexity cue, $m$ is the local slope, $k$ is the power-law exponent, and $\Delta$ is depth. To find the local slope in disparity space, $m_d(c, f(\Delta))$, we mapped the approximation of the convexity-depth distribution into disparity and used the chain rule to find the derivative.

Assuming local linearity, we can approximate the posterior in disparity as a Gaussian:

$$P(\Delta \mid d, c) \approx \exp \left[ -\frac{(f(\Delta) - d)^2}{2\sigma^2(\Delta)} + m_d(c, f(\Delta))f(\Delta) + b(c, \Delta) \right]$$  \hspace{1cm} (10)

with mean $f(\Delta) + m_d\sigma^2(\Delta)$ and standard deviation $\sigma(\Delta)$. We found that this Gaussian approximation is quite accurate for all but the smallest depths where $P(\Delta \mid c)$ has a very steep slope (results not shown). With the Gaussian approximation to the posterior in hand, we can compute $P(\hat{\Delta}_{\text{comp}} > \hat{\Delta}_{\text{strd}})$ in closed form as a function of the parameters.

To fit the model, we optimized all parameters separately for the two observers in the main experiment using Matlab’s Nelder-Mead simplex routine. The non-parametric model had 26 parameters (24 for the convexity-depth distributions and two for the disparity distributions). When we modeled those distributions as power laws, the model had only five parameters (three for convexity-depth and two for disparity).

The effect of convexity should increase monotonically as the depth specified by disparity increases. We assumed that perceived depth is given by the maximum-a-posteriori (MAP) estimate, and that the disparities are small enough for the small-angle approximation to apply. Differentiating the exponent in Eqn. (10) with respect to $d$ and setting it equal to zero yields the following for perceived depth:

$$\hat{\Delta}(d, c) = f(\Delta) + m_d(c, f(\Delta))\sigma^2(\Delta)$$  \hspace{1cm} (11)

where the last term, $m_d(c, f(\Delta))\sigma^2(\Delta)$, is the disparity bias associated with a stimulus defined by disparity and convexity signals $d = f(\Delta)$ and $c$, respectively. We can convert this bias into depth.

The power-law model makes an interesting prediction in this regard. Substituting Eqns. 7 and 9 into the expression for disparity bias (again assuming the small-angle approximation) yields:

$$\text{bias} \approx k\Lambda$$

where $k$ is the power-law exponent and $\Lambda$ is the depth-from-disparity signal. This means that the effect of region convexity should increase in proportion to the depth specified by disparity.

Note that only intrinsic uncertainty (e.g., uncertainty due to neural noise) contributes to stochasticity in the response of an ideal observer. Eqns. 3 and 4 assume that all the uncertainty in the depth distribution $P(\Delta \mid d, c)$ contributes to stochasticity and thus implicitly attributes this uncertainty to intrinsic sources. This is reasonable for the disparity cue because the mapping from disparity to depth is deterministic; all the uncertainty is probably due to intrinsic sources. However, the mapping from convexity to depth contains significant extrinsic uncertainty (Fig. 2), so Eqn. 3 is not strictly correct. In Supplementary Material, we present a more detailed analysis that attributes sources of uncertainty correctly. This analysis gives the same result as Eqn. 3 for the case under consideration (i.e., when the convexity-depth distribution has much greater variance than the disparity-depth distribution). Consequently, in the fitting procedure, the standard deviation of the disparity distribution determines the slope of the ideal observer’s psychometric curve.
RESULTS FOR MODELING OF PSYCHOPHYSICS

We first examined how the disparity-likelihood distributions we estimated from the psychophysical data compared with previous findings in the literature. Fig. 7 shows the standard deviations of the estimated disparity distributions plotted along with previously reported results (McKee et. al., 1990). The agreement is good. This replication is important because it shows that our analysis recovered reasonable values for the disparity distributions, and it supports our assumption that disparity and convexity provide conditionally independent depth estimates.

Figure 7. Disparity increment thresholds. Solid and dashed lines were derived from the data from Experiment 1. The solid lines represent the estimated standard deviations of the disparity likelihoods with increasing disparity when the data were fit with the non-parametric model. Dotted lines represent the estimated standard deviations when the power-law model was used. Estimates for both observers are shown. Symbols represent disparity increment thresholds for different subjects and conditions in a previously published report (McKee et al, 1990). In that report, threshold was defined as the 75% point on the psychometric function. We transformed the data to match our 84% criterion.

We now turn to the estimation of the convexity-depth distributions that observers used in Experiment 1. Fig. 8a shows the distributions estimated from psychophysics when we used the non-parametric model, a model that makes very few assumptions about the shapes of the distributions. The estimated distributions are similar in many respects to the convexity-depth distributions recovered from the natural-scene statistics (Fig. 2): 1) the distributions are skewed such that large depths are more probable across contours that bound convex regions than straight-contoured or concave regions; 2) all three distributions have much heavier tails than Gaussians; 3) the estimated convexity-depth distributions are roughly linear in a log-log plot, which means that they are similar to power laws. Fig. 8b shows the distributions that were recovered from the psychophysical data when we when we fit them using the power-law model. By fitting only the power-law exponents, the number of free parameters was reduced from 26 to five (two for disparity and three for convexity-depth distributions).
Figure 8. Convexity-depth distributions estimated from Experiment 1. Blue represents the distribution for convex reference regions, red the distribution for concave reference regions, and black the distribution for straight-contoured reference regions. a) Convexity-depth distributions recovered from fitting the psychometric data with the non-parametric model. The distributions are approximately linear in these log-log plots, suggesting that they were well described by a power law. b) Convexity-depth distributions that were recovered by re-fitting the psychometric data with the power-law model. Note the similarity to the non-parametric distributions in a. One free parameter had to be set to uniquely determine these fits (see Methods). The differences between the estimated distributions were unaffected by the value of that parameter. The standard errors for the best-fitting parameters were estimated by re-doing the fitting procedure 30 times with bootstrapped sets of psychometric data. The standard error intervals were about the same as the line widths in a and b, so they are not plotted.

Fig. 9 shows that the power-law model provides nearly as good a fit to the data as the non-parametric model, even though the power-law model had ~1/5 the number of free parameters. To assess the goodness of fit of a given model, we computed the sum-squared error (SSE) between the model predictions and the psychometric data, took the inverse of the SSE, and normalized such that 1 corresponded to the inverse error obtained by fitting a cumulative Gaussian to each experimental condition independently (80 free parameters). Thus, a value of 1 represents an upper bound on goodness of fit. We computed the normalized inverse SSE for four models: a random coin-flipping model, a model in which the convexity-depth distributions were identical (equivalent to assuming convexity is not used in depth estimation), and the aforementioned power-law and non-parametric models (Fig. 9). Because the power-law model provided nearly as good a fit to the data as the non-parametric model, we conclude that power laws are excellent descriptions of the internal convexity-depth distributions.
Figure 9. Model comparison. The sum of squared error (SSE) between model predictions and the psychometric data was calculated. The inverse of those values was then computed and normalized such that 1 represents the normalized inverse error obtained by fitting a cumulative Gaussian to each experimental condition independently (80 free parameters). Larger values represent better fits (lower SSE). Normalized inverse error is shown for four probabilistic models fit to the data: i) Coin-flipping model (0 free parameters); ii) a model that assumed that the convexity-depth distributions were the same for convex, concave, and straight-contoured regions; this is equivalent to assuming that convexity was not used when observers estimated perceived depth (3 free parameters); iii) power-law model (5 free parameters); iv) non-parametric model (26 free parameters). We also measured the fits when the difference between the convex and concave distributions was set equal to the difference recovered from the natural-scene statistics. The fits were nearly as poor as the model in which convexity is not used (relative fit error: EKK = 0.62; JIL = 0.67). Finally, we measured the fit errors with Gaussians instead of power laws and found poor fits (relative fit error: EKK = 0.7; JIL = 0.82); JIL’s Gaussian fit is comparable to the power-law fit, but that fit required extremely low-variance disparity likelihoods, which were inconsistent with the literature (McKee et al., 1990; Morgan et al., 2000). We conclude the power-law model is a good description of the internal convexity-depth distributions.

While the convexity-depth distributions estimated from the psychophysical experiment and natural-scene measurements were similar in many ways, they did differ in one interesting respect. The effect of convexity was much larger in the psychophysically estimated distributions (see Fig. 9 caption): $k_{\text{difference}}$, the difference between the exponents of the best-fitting power laws from the experiment was $\sim 4.4$, while $k_{\text{difference}}$ from the natural statistics was 0.1 to 0.6 depending on which set of contours was included (Fig. 2c,d,e).

There is a plausible explanation for this discrepancy. The contours presented in the psychophysical experiments were different from the great majority of contours analyzed in the natural scenes. In the experiments, the contours were high contrast, and the regions they bounded were circular so they had consistent convexity across scale (e.g., the silhouette of a basketball). In the natural-scene dataset, many contours were low contrast, and the regions they bounded were convex at some spatial scales and concave at others (e.g., the silhouette of a tree). To examine this possibility, we re-computed the natural-scene statistics, focusing on high-contrast, regular contours.
Figs. 2d and 2e show the convexity-depth distributions for natural-scene contours after selection with more or less restrictive criteria. When the selected contours are most similar to those in the psychophysical experiments (i.e., a narrow range of convexities or consistent classification as convex or concave across scale), \( k_{\text{difference}} \) becomes more similar quantitatively to the value recovered from the psychophysical experiments. Thus, conditioning on additional visual features that restrict the contours under consideration to be more similar to the experimental stimuli yields natural-scene distributions that are quantitatively more similar to the distributions recovered from the psychophysics. Unfortunately, we cannot further refine the selection of contours from the database on other visual features such as contrast because doing so yields too few samples to calculate meaningful statistics. As more natural-scene data become available, one could pursue this further.

**DISCUSSION**

*Natural-scene statistics*

The importance of natural-scene statistics in perceptual tasks was first articulated by Brunswik (Brunswik & Kamiya, 1953) who argued that Gestalt cues could and should be ecologically validated. The role of natural-scene statistics has since been investigated in relation to several perceptual tasks: contour grouping (Geisler, Perry, Super, Gallogly, 2001; Elder & Goldberg, 2002; Ren & Malik, 2002), figure-ground assignment (Fowlkes et al., 2007) and length estimation (Howe & Purves, 2002). It is useful to distinguish this line of thinking from a much larger literature, starting with Barlow (Barlow, 1961), on how the statistics of natural images relate to neural coding. Current thinking is that processing in early visual pathways evolved to code and transmit as much information about retinal images as efficiently as possible (Barlow, 1961; Simoncelli & Olshausen, 2001). To determine the encoding efficiency, one needs to know the statistics of the image properties to be encoded, so recent work has focused on measuring natural-image statistics and determining whether or not visual-cortical neurons are constructed to exploit those statistics (Olshausen & Field, 1996; Vinje & Gallant, 2000). It is unclear, however, that efficient encoding is the primary task of early visual processing.

In order to make strong claims about the role of natural statistics in perception, we believe that it is important to study tasks that are known to be critical for the biological system under study. By knowing that the task is critical, one can determine that a failure to exploit the task-relevant statistics is due to sub-optimal performance in a necessary task, rather than to a mistaken hypothesis about what the system does. Surely one of the most important perceptual tasks is to estimate the 3D layout of the environment because such estimation is required for guidance of visuomotor behavior in that environment (Geisler, 2008). For this reason, we chose to focus on the task of estimating 3D layout.

*Closing the loop in probabilistic models of perception*

Many perceptual effects cannot be explained from *a priori* geometric constraints alone. For example, perceived speed decreases when the contrast of the moving stimulus is reduced (Stocker & Simoncelli, 2006; Weiss, Simoncelli, Adelson, 2002). However, this effect can be easily understood in a probabilistic framework in which sensory measurements are combined with statistical information about the environment: in this case, a prior for slow motion. With decreased contrast, the variance of sensory measurement increases and the prior has greater influence, pulling perceived speed toward zero. In that work, the prior for speed was estimated from psychophysical data (Stocker & Simoncelli, 2006). The analysis makes sense, but the estimated probability distribution cannot be compared to “ground truth” because the speed prior for the natural environment is not known. Thus, it cannot be argued from that analysis alone that the visual system has internalized statistical properties of the environment.
The work presented here attempts to close this gap. We showed that humans behave as if they use an asymmetric convexity-depth distribution when making depth judgments in the presence of a region-bounding contour. Importantly, the distribution they appear to be using is similar to the one in the natural environment. Our experiments thus demonstrate the ecological validity of convexity as a cue to depth and thus explain its usefulness to the human visual system.

**The depth information in figure-ground cues**

Figure-ground cues are stimulus features that bias observers to see one region as occluding another (Palmer, 1999). Such cues—e.g., convexity, lower region, size, symmetry, familiarity, and contrast—help determine to which region a contour belongs; the figure region appears shaped by the contour and closer to the viewer. Image segmentation is a fundamental part of vision, so the role of figure-ground cues in biological and machine vision has been extensively investigated (Palmer, 1999; Burge et al., 2005; Bertamini et al., 2008; Fowlkes et al., 2007; Rubin, 1921; Driver & Baylis, 1996; Vecera et al., 2002; Peterson, Harvey & Weidenbacher, 1991; Qiu & von der Heydt, 2005; Sugita, 1999; Bakin, Nakayama, Gilbert, 2000), but no comprehensive theory of how figure-ground cues affect contour assignment and depth perception has emerged. Our analysis of the figure-ground cue of convexity may be the beginning of such a theory. The natural-scene statistics reveal why convexity is a cue to occlusion and to metric depth. The probabilistic model shows how the convexity cue might be integrated into the visual percept and the psychophysical results are consistent with this form of integration.

Many figure-ground cues could affect metric depth percepts in similar fashion. Consider, for example, the figure-ground cue of lower region (Vecera et al., 2002). The ground plane is present in most views of natural scenes, so positions low in the visual field are likely to be nearer to the viewer than positions high in the field (Potetz & Lee, 2003; Huang, Lee, Mumford, 2000; Yang & Purves, 2003b). In our analysis of natural-scene statistics, we observed that the asymmetry in the probability distribution of depth conditioned on elevation in the visual field is greater than the asymmetry in the distribution conditioned on convexity. If the visual system has incorporated the statistics associated with lower region, we would therefore expect generally larger perceptual effects from lower region than from convexity. Psychophysical experiments confirm this expectation. Lower region and convexity determine figural assignment ~70% and ~60% of the time, respectively (Vecera et al., 2002; Peterson & Skow, 2008). Lower region is also a better predictor than convexity of depth-ordering judgments in images of natural scenes (Fowlkes et al., 2007).

Two other phenomena are consistent with our claim that figure-ground cues provide metric depth information. First, a monocularly viewed sequence of stacking disks generates a vivid sense of movement toward the viewer (Engel, Remus, Sainath, 2006). By the traditional taxonomy, the depth cues in this stimulus provide only distance-ordering information: T-junctions (Palmer, 1999) and the figure-ground cues of surroundedness and small area (Rubin, 1921). When the stacking sequence ceases, an aftereffect of movement away from the viewer occurs (Engel et al., 2006). Second, a monocularly viewed moving disk that occludes a background of vertical bars when translating leftward and is occluded by the bars when moving rightward is perceived as moving elliptically in depth; it also elicits vergence eye movements consistent with an elliptical path (Ringach, DL. PhD Dissertation, New York University. 1996, AAT 9621827). Standard models of motion estimation cannot explain these motion-in-depth percepts, nor the vergence eye movements induced by the second stimulus, if cues to occlusion provide only distance-ordering information. However, if occlusion cues provide metric information, as proposed here, the percepts and accompanying eye movements are readily understood.

**Other depth cues**
In addition to the figure-ground cues, many depth cues such as aerial perspective (Trosclair et. al., 1991), shading (Koenderink & van Doorn, 1992), and blur (Mather & Smith, 2000) are regarded as non-metric because from the cue value alone, one cannot uniquely determine depths in the scene. This view stems from the geometric relationship between retinal-image features associated with the cue and depth: with these cues there is no deterministic relationship between relevant image features and depth in the scene. We propose instead that the visual system uses the information in those cues probabilistically, much as it uses convexity. From this proposal, it follows that all depth cues have the potential to affect metric depth estimates so long as there exists a non-uniform statistical relationship between the cue value and depth in the environment.

Indeed, all depth cues should probably be conceptualized in a probabilistic framework. Such an approach has recently been explored in computer vision (Hoiem et. al., 2005; Saxena et. al., 2005) where machine-learning techniques were used to combine statistical information about depth and surface orientation provided by a diverse set of image features. Some of these features were similar to known monocular depth cues; others were not. From the information contained in a large collection of these features, the algorithms were able to generate reasonably accurate estimates of 3D scene layout. These results show that useful metric information is available from image features that have traditionally been considered non-metric depth cues. Interestingly, the results also show that useful depth information is available from image features that have not yet been identified as depth cues.

Our results and the above-mentioned computer-vision results indicate that the conventional, geometry-based taxonomy that classifies depth cues according to the type of distance information they can provide is unnecessary. By capitalizing on the statistical relationship between images and the environment to which our visual systems have been exposed, the probabilistic approach used here will yield a richer understanding of how we perceive 3D layout.

**Neurophysiology and figure-ground cues**

Neurons in visual cortex exhibit properties that might be closely related to the perceptual effects reported here. Figure-ground cues affect the responses of many disparity- and orientation-selective neurons in early visual areas. For example, many V2 neurons respond to contours defined by disparity alone and to contours defined by figure-ground cues alone (Qiu & von der Heydt, 2005). If disparity and figure-ground cues are both present, these neurons generally respond more vigorously when disparity and figure-ground cues are consistent (both indicate that the region on one side of a contour is nearer than the region on the other side) than when the cues are inconsistent. The increased response when the cues are consistent might be interpreted by other neural centers as signaling greater depth.

Sensitivity to the consistency of disparity and figure-ground cues has been demonstrated in other ways. Most V1 and V2 neurons respond vigorously to extended bars of the preferred orientation. When a small patch covers the classical receptive field but only partially obscures the stimulus bar, the response of some V1 and many V2 neurons depends strongly on the patch’s disparity relative to the bar (Sugita, 1999; Bakin et. al., 2000). The neurons barely respond when disparity specifies that the patch is in the plane of the bar or behind it. But when disparity specifies that the patch is in front of the bar, as occurs with real occluders in natural scenes, the neurons respond as vigorously as when the patch was not present.

Such results show that links between disparity and cues to occlusion occur early in the visual pathways. These neurons may partially mediate the perceptual effects reported here.

**Summary**

In the analysis of natural-scene statistics, we found that convexity provides information about depth in natural scenes. The convexity of an image region can, therefore, provide information about the probability of different depths across the contour bounding that region. We constructed a
probabilistic model of how that information would be used to maximize the accuracy of depth estimates. In psychophysical experiments, we showed that convexity affects depth estimation in a manner consistent with such a model. Our work thus establishes the ecological validity of the figure-ground cue of convexity and its usefulness to the human viewers. The increasing availability of natural-scene datasets of the environmental properties the visual system seeks to estimate—depth, velocity, surface orientation, object identity—should allow similar undertakings for many cues and tasks and guide the study of neural mechanisms that underlie the relationship between scene statistics and perceptual estimates.

ACKNOWLEDGEMENTS
Supported by AOF’s William C. Ezell Fellowship to JB and NIH Research Grant R01-EY12851 to MSB. The authors thank Ahna Girshick, Mike Landy, Jitendra Malik, Steve Palmer, and Mary Peterson for helpful discussion, and Zhiyong Yang, Dale Purves, Brian Potetz, and Tai Sing Lee for making their natural-scene data sets available.
REFERENCES


Fowlkes, C, Martin, D, Malik, J (2007). Local Figure/Ground Cues are Valid for Natural Images. *Journal of Vision*, 7(8), 1-9.


Qiu, FT, von der Heydt, R (2005) Figure and ground in the visual cortex: V2 combines stereoscopic cues with Gestalt rules. *Neuron*, 47, 155-166


Rubin, E. (1921) *Visuell Wahrgenommene Figuren*. Kobenhaven: Glydenalske Boghandel


