Behavioral/Systems/Cognitive

Natural-Scene Statistics Predict How the Figure–Ground Cue of Convexity Affects Human Depth Perception

Johannes Burge,[1] **Charless C. Fowlkes,**[3] **and Martin S. Banks**[2]

[1]School of Optometry and [2]Vision Science Graduate Program, University of California at Berkeley, Berkeley, California 94720-2020, and [3]University of California, Irvine, Irvine, California 92697

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 the depth information conveyed by convexity, we examined natural scenes and found that depth steps across surfaces with convex silhouettes are likely to be larger than 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 metric depth cue. They also suggest that the dichotomy between metric and nonmetric depth cues is false and that the depth information provided many cues should be evaluated with respect to natural-scene statistics. Finally, the findings provide an explanation for why figure–ground cues modulate the responses of disparity-sensitive cells in visual cortex.

Introduction

Estimating three-dimensional (3D) layout from the two-dimensional retinal images is geometrically underdetermined, but biological visual systems solve the 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 et al., 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 manner (Knill and

Saunders, 2003; Hillis et al., 2004). However, it is unclear how information from ordinal and metric cues should be combined (Landy et al., 1995). Ordinal information conveys only the sign of depth between pairs of surfaces, so either the ordinal cue is 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 perceived depth separation between the occluder and the background even when a metric depth cue, disparity, is available (Burge et al., 2005; Bertamini et al., 2008).

This counterintuitive result can be understood by considering the depth information potentially provided by both disparity and the shape of the occluding contour. Disparity can signal depth directly although uncertainty arises from noise in disparity measurement (Cormack et al., 1997) and in other signals required to interpret measured disparities (Backus et al., 1999). There is no a priori geometric reason that contour shape 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—just as other cues are statistically related (Brunswik and Kamiya, 1953; Hoiem et al., 2005; Saxena et al., 2005)—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 indeed 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.

Materials and Methods

Methods for natural-scene analysis. We investigated the hypothesis that figure–ground cues—specifically contour convexity—provide useful metric depth information by examining natural-scene statistics. Depth is defined as the difference in viewing distance for two points along a line of sight. We analyzed the figure–ground cue of convexity because convexity can be readily measured in images of natural scenes (Fowlkes et al.,2007) and because convexity is known to affect figure–ground assignment in human perception (Metzger, 1953; Kanizsa and 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 published luminance and range images (spatially coregistered RGB and time-of-flight laser-range data) (Potetz and Lee, 2003; Yang and Purves, 2003a). Figure 1*a* is an example of a luminance image and Figure 1*b* is the corresponding range image. Many of the luminance images in the published reports were underexposed (black), noisy, and/or blurred, so we selected a well exposed, sharp subset of 35 images [23 from Yang and Purves (2003a) and 12 from Potetz and Lee (2003)] that had localizable contours in a variety of indoor and outdoor scenes. To identify a representative set of region-bounding contours, five people, who were naive to the experimental hypotheses, hand selected contours in the luminance images using a previously published procedure (Martin et al., 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 selectors generally did not mark shadows or contours for image regions that subtended <3° (~20 pixels). To assess the selection consistency, two people marked each image. The agreement in their selections was measured using local consistency error (LCE) (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). We did not use the range data to segment the images because we were investigating the depth information provided by the shape of luminance contours. To use the range image for segmentation would, therefore, undermine the logic of our analysis.

For each point along a selected contour, we computed the distance and the convexity of the local image regions on either side of the contour. The local image region was defined by a circular analysis window of fixed radius (2.16°, 15 pixels) centered at each point along a selected contour. To compute depth, we determined the difference between the mean distances to each pixel in the two regions. To compute local convexity, we sampled pairs of pixels inside both regions and recorded the fraction of pairs for which a line segment connecting them lay completely within the region. We defined convexity, *c*, as the log ratio of the two fractions (Fowlkes et al., 2007). Figure 1*e* shows convexity along each selected contour in the expanded region from Figure 1*a*. Finally, we determined the frequency of different depths as a function of region convexity. This analysis (see Fig. 3) is the simplest possible in that it concerns the relationship between convexity and depth without regard to other image or scene properties. For example, we did not consider how the convexity–
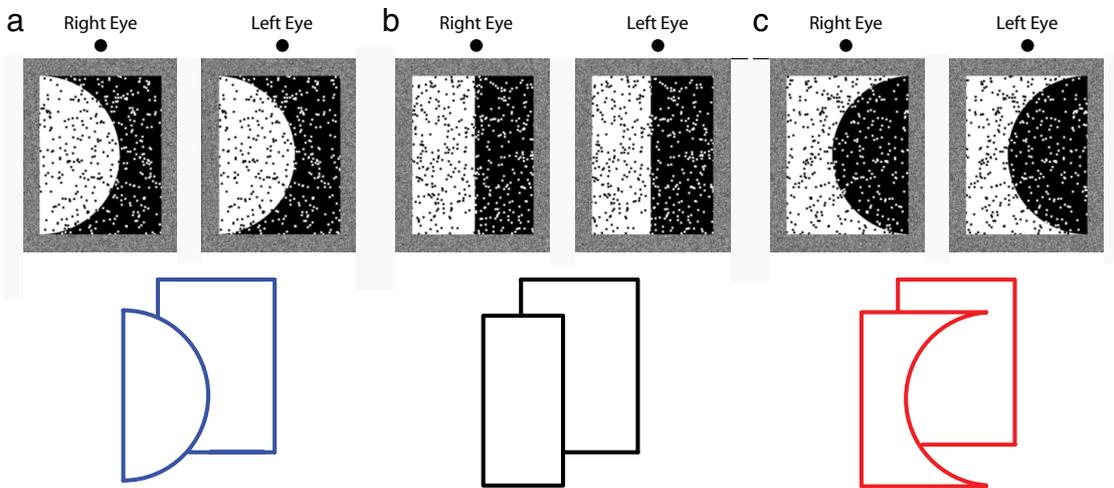

Figure 1. Luminance and range images. *a*, An example luminance image. *b*, Corresponding range image. Blue indicates nearer distances, yellow intermediate distances, and red farther distances. *c*, Close-up of the luminance image with a representative hand selection overlaid. *d*, Close-up of the associated range image with the same selection overlaid. *e*, The same image as *d*, but with convexity flags added. Flags point toward the image region that was classified as more convex. Longer flags correspond to larger convexity values.

depth relationship varies with scene properties such as the distance to a contour. Later, we consider how some low-level image properties affect the relationship (see Fig. 10).

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. All but one observer in experiment 3 were naive to the experimental hypotheses.

The stimuli were presented on a cathode ray tube monitor (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) synchronized to the display. The frame rate was 100 Hz, so each eye's image was refreshed at 50 Hz. To minimize cross talk 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 et al., 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/deg²), separated by

Figure 2. 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 and contour shape 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.

a luminance contour that was either curved (radius of curvature = 10.5 cm) or straight (Fig. 2). 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. 2*a*). For "neutral" stimuli, the contour was a vertical line (Fig. 2*b*,. In "inconsistent" stimuli, the disparity-specified near region was concave (Fig. 2*c*). 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 and Baylis, 1996), lower region (Vecera et al., 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.

Experiment 1. Two stimuli were presented sequentially on each trial: a standard and comparison. 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 s each in random order with an interstimulus period of 0.5 s. 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, neutral–inconsistent, neutral–neutral, consistent–consistent, and inconsistent–inconsistent. 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 stan-

dard disparity. Blocks consisted of 320–450 trials. We observed no systematic effect for left-versus-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 (five pairings × eight 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 84% on the psychometric function—by fitting a cumulative Gaussian to the psychometric data (Wichmann and 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. Seven observers participated in 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 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.

Methods for modeling of psychophysics. The psychophysical results suggest that human observers have incorporated the natural-scene statistics associated with region convexity and depth. We next asked whether the observed behavior is consistent with the behavior of an ideal observer that has incorporated the natural statistics associated with convexity and depth. To do so, we fit the observers' responses with a probabilistic model of depth estimation to determine the probability distributions that provided the best quantitative account of their responses. We modeled the construction of the depth percept as

a probabilistic process. Assuming conditional independence, Bayes' rule states the following:

$$P(\Delta|d,c) = \frac{P(d|\Delta)\,P(c|\Delta)\,P(\Delta)}{P(d,c)}. \qquad (1)$$

That is, the posterior probability of a particular depth, Δ, 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, $P(c|\Delta)$, and the prior distribution of depths, $P(\Delta)$, for all such contours, divided by a normalizing constant, $P(d,c)$. The product rule, $P(c|\Delta)P(\Delta) = P(c,\Delta)$, allows us to combine the latter two terms in the numerator into the joint probability of c and Δ as follows:

$$P(\Delta|d,c) = \frac{P(d|\Delta)\,P(c,\Delta)}{P(d,c)}. \qquad (2)$$

We regroup using the product rule again as follows:

$$P(\Delta|d,c) = \left(\frac{1}{P(d,c)/P(c)}\right) P(d|\Delta)\,P(\Delta|c), \qquad (3)$$

where $1/[P(d,c)/P(c)]$ is a normalizing constant, $P(d|\Delta)$ is the disparity likelihood, and $P(\Delta|c)$ is the convexity–depth distribution, which is the distribution we measured from the luminance range images (Fig. 1). Note that the depth prior, $P(\Delta)$, has been absorbed by $P(\Delta|c)$.

In this model, both the convexity of an image region and its disparity relative to the adjoining region affect the expected perceived depth. This is illustrated schematically below: Figure 4 shows probability distributions associated with 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 case (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.

The disparity likelihood, $P(d|\Delta)$, is plotted as a distribution over depth (see Fig. 4). We can do this because of the following relationship between depth and disparity:

$$d = f(\Delta) = 2(\tan^{-1}(I/2v) - \tan^{-1}(I/2(v+\Delta))), \qquad (4)$$

where v is the fixation distance and I is the interocular separation (Howard and Rogers, 2002).

In the psychophysical experiment, observers selected on each trial the stimulus (standard or comparison) with greater perceived depth. Those responses generated the psychometric data. We modeled the process that generated the responses by computing posterior distributions for the standard and comparison stimuli on each trial. We then used signal detection theory (Green and 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, as shown by the following:

$$P(\hat\Delta_{\text{comp}} > \hat\Delta_{\text{stnd}}) = \int_0^\infty P(\hat\Delta_{\text{comp}}|d_{\text{comp}}, c_{\text{comp}})$$

$$\int_0^{\hat\Delta_{\text{comp}}} P(\hat\Delta_{\text{stnd}}|d_{\text{stnd}}, c_{\text{stnd}})\, d\hat\Delta_{\text{stnd}}\, d\hat\Delta_{\text{comp}}. \quad (5)$$

We thus sought distributions $P(d|\Delta)$ and $P(\Delta|c)$ that minimized the squared difference between the model predictions and the raw psycho-

metric data for all comparison and standard disparities across all experimental conditions as follows:

$$\min\sum_{i=1}^5\sum_{j=1}^8\sum_{k=1}^n N_{ijk}\left[P(\hat\Delta_{\text{comp}}^{ik} > \hat\Delta_{\text{stnd}}^{ij}) - y_{\text{data}}\,(\hat\Delta_{\text{comp}}^{ik} > \hat\Delta_{\text{stnd}}^{ij}) \right]^2,$$

$$(6)$$

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

Previous research on disparity discrimination suggests that the disparity distributions can be approximated by Gaussians with SDs proportional to their means plus a constant (McKee et al., 1990; Morgan et al., 2000). Therefore, we modeled the disparity likelihood as follows:

$$P(d|\Delta) = N(f(\Delta),\sigma(\Delta)), \qquad (7)$$

where $f(\Delta)$ is the mean disparity signal, d, for a given depth, and the SD, $\sigma(\Delta)$, increases with the mean. Specifically, we assumed the following:

$$\sigma(\Delta) = af(\Delta) + \sigma_0, \qquad (8)$$

where a is the rate at which the SD increases with increasing disparity, and σ_0 is a small value that corresponds to the SD when the disparity is zero. Thus, the disparity likelihoods were characterized by two parameters for all experimental conditions.
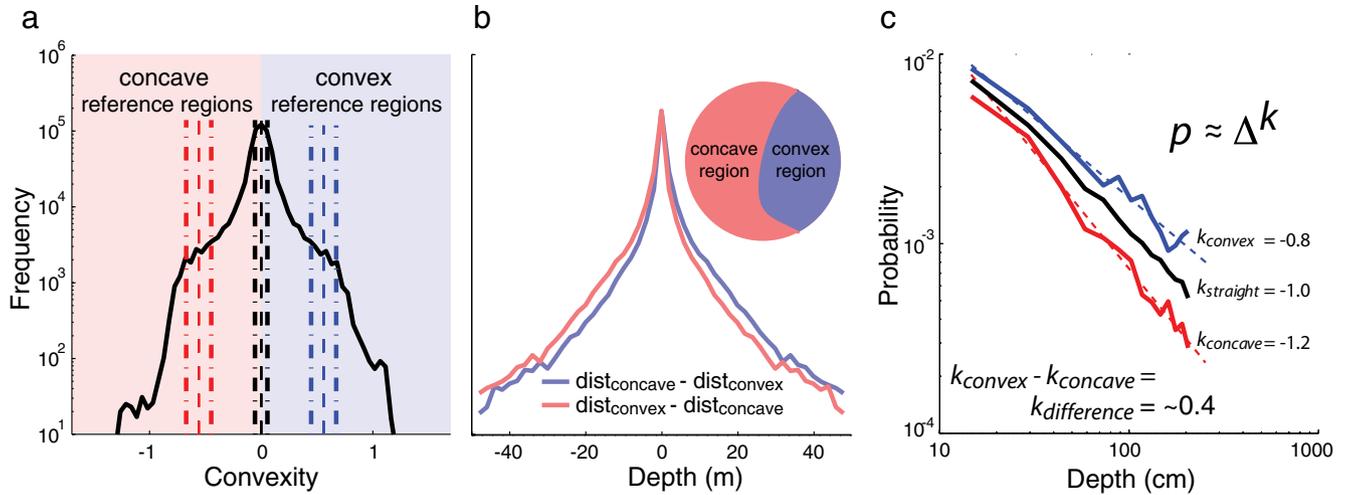
For the convexity–depth distributions, we used a nonparametric model that approximated them as piecewise log linear. We chose this model (similar to Stocker and Simoncelli, 2006) because it makes very few assumptions about the shape of the distributions. In this model, the convex, concave, and straight-contoured 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 nonparametric 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 fit a parametric model in which the convexity–depth distributions were assumed to follow a power law over depth. This has the advantage of reducing the number of free parameters necessary to model the convexity–depth distributions from 24 to 3. 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 nonparametric model.

To evaluate $P(\hat\Delta_{\text{comp}} > \hat\Delta_{\text{stnd}})$ requires computing the product of the disparity likelihood and convexity–depth distributions once they are expressed in common units. We expressed the distributions in disparity. To approximate the posterior, 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 the following:

$$\ln(P(\Delta|c)) = m(c,\Delta)\Delta + b(c,\Delta), \qquad (9)$$

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

Figure 3. Natural-scene statistics computed from ∼450,000 sampled contour points. ***a***, Frequency of different convexities. A circular analysis window was centered at each point along a selected contour. One of the two regions inside the window was arbitrarily chosen as the reference region. Positive values indicate that the reference region was more convex than the other region (light blue shading); negative values indicate that the reference region was more concave (pink shading). ***b***, Joint statistics of convexity and depth. Frequency of occurrence plotted as a function of depth (distance to the region on the nonreference side of the contour minus distance to the reference region) for all convex (light blue) and concave (pink) reference regions. When the reference region is nearer ("figure"), depths are positive. When the reference region is farther ("ground"), depths are negative. The two curves are mirror images because every curved contour has a convex region on one side and a concave region on the other. ***c***, Joint statistics of convexity and depth for depth intervals (0–2 m) and convexities similar to those in the psychophysical experiments, plotted in log–log coordinates. The means of the bins defined by the dashed vertical lines in ***a*** are similar to the convexities of the three stimulus types in the psychophysical experiments: convex (blue), concave (red), and straight contoured (black). The straight bin has ∼10 times more samples than the convex and concave bins, so the black curve is smoother. Dotted lines are power-law fits to the data. Except for depths at or very near zero, the data are well fit by a power law. The difference between the convex and concave power-law exponents, $k_{difference} = k_{convex} - k_{concave}$, is ∼0.4.

parameters. The local slope of a power law in log-probability space is given by the following:

$$m(c,d) = \frac{k}{\Delta},$$ (10)

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

Assuming that the convexity–depth distribution is locally linear on a log-linear plot, we can approximate the posterior in disparity as a Gaussian as follows:

$$P(\Delta|d,c) \approx \exp\left[-\frac{(f(\Delta) - d)^2}{2\sigma^2(f^{-1}(d))} + m_d(c,d)f(\Delta) + b_d(c,d)\right],$$ (11)

with mean $d + m_d\sigma^2(f^{-1}(d))$ and SD $\sigma(f^{-1}(d))$. We found that this Gaussian approximation is quite accurate for all but the smallest depths, where $P(\Delta|c)$ has a very steep slope (results not shown). With the Gaussian approximation to the posterior in hand, we can compute $P(\hat{\Delta}_{comp} < \hat{\Delta}_{stnd})$ in closed form as a function of the parameters.

To fit the model, we optimized all parameters separately for the two observers in experiment 1 using Matlab's Nelder–Mead simplex routine. The nonparametric model had 26 parameters (24 for the convexity–depth distributions and 2 for the disparity distributions). When we modeled the distributions as power laws, the model had only five parameters (three for convexity–depth and two for disparity).

It is interesting to analyze Equation 11 in more detail. If we assume 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, then differentiating the exponent in Equation 11 with respect to d and setting it equal to zero yields the following expression for perceived depth:

$$\hat{\Delta}(d,c) \approx f^{-1}(d + m_d(c,d)\sigma^2(f^{-1}(d))).$$ (12)

The last term, $m_d(c,d)\sigma^2(f^{-1}(d))$, is the disparity bias associated with a stimulus defined by disparity and convexity signals $d = f(\Delta)$ and c, respectively. The power-law model makes a simple prediction in this regard. Substituting Equations 8 and 10 into the expression for disparity bias (again assuming the small-angle approximation) yields the following: bias $\approx k\Delta$, where k is the power-law exponent and Δ is the depth-from-disparity signal. This means that the effect of region convexity should increase in proportion to the depth specified by disparity.

Finally, we point out that only intrinsic uncertainty (e.g., uncertainty due to neural noise) contributes to stochasticity in the response of an ideal observer. Equations 4 and 5 assume that all the uncertainty in the posterior distribution over depth, $P(\Delta|d,c)$, contributes to stochasticity and thus implicitly attributes this uncertainty to intrinsic sources. This is reasonable for the disparity cue because the mapping from depth to disparity is deterministic; all the uncertainty is presumably due to intrinsic sources. However, the mapping from convexity to depth contains significant extrinsic uncertainty (Fig. 3), so Equation 5 is not strictly correct. In the supplemental material, available at www.jneurosci.org, we present a more detailed analysis that attributes sources of uncertainty correctly. This analysis gives the same result as Equation 5 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 SD of the disparity distribution determines the slope of the ideal observer's psychometric curve.

Results

Results of natural-scene analysis

Figure 3*a* shows the distribution of convexities for the contours we analyzed. Figure 3*b* shows the statistical relationship between convexity and depth, the distance between the regions on either side 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.
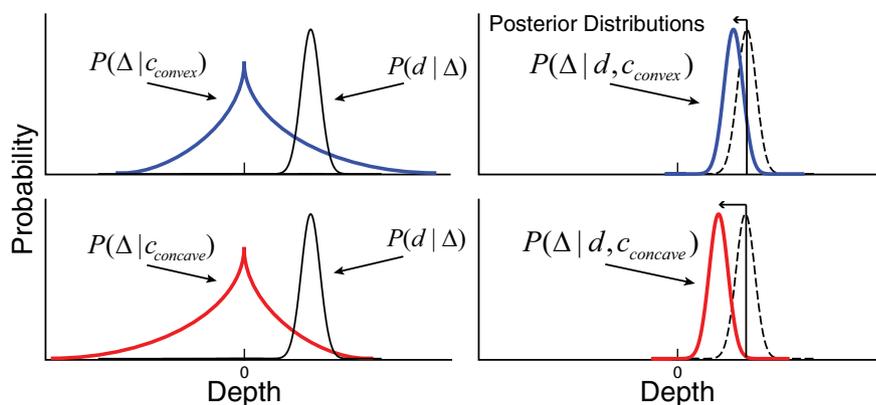
To our knowledge, this is the first evidence that a geometrically ordinal depth cue—the figure–ground cue of convexity—is statistically correlated with depth magnitude. The effect is small, but consistent. Thus, the relative convexity between two adjacent image regions can in principle provide information about the probability of different depths between visible surfaces. The cause of the statistical relationship is presumably that most objects are mostly convex. The peak of the distribution is at zero because many of the analyzed contours corresponded to changes in reflectance or lighting and not changes in depth.

To examine conditions similar to those in the psychophysical experiments we describe below, we analyzed the scene statistics further. The depths specified in the experiments were not large, so we focused the scene analysis on depths between 0 and 2 m and on image regions whose convexities were very nearly equal (e.g., straight contours) or were significantly different, as they were in the psychophysical experiments and as defined in Figure 3a. The distributions conditioned on these three convexities at depths of 0–2 m are shown in Figure 3c. The distributions are well described by power laws (linear on log–log plots) at all depths except ~0. The power-law relationship is robust across subsets of images and human contour selectors (supplemental Fig. S1, available at www.jneurosci.org as supplemental material). We will take advantage of the power-law property and henceforth describe distributions by the difference in the exponents k of the best-fitting functions: $k_{\text{difference}} = k_{\text{convex}} - k_{\text{concave}}$.

Results of psychophysical experiments

Figure 4 illustrates the expected influence of contour shape on depth judgments. The upper row depicts the probability distributions associated with consistent stimuli (Fig. 2a) in which convexity and depth both indicate that a particular 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. 2c). 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) even when the disparity-specified depth is the same.

We also expect the influence of convexity on perceived depth to 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 et al., 1997). Humans behave similarly in many perceptual tasks (Ernst and Banks, 2002; Knill and Saunders, 2003; Alais and Burr, 2004; Hillis et al., 2004). Second, the variance of depth from disparity increases as the disparity increases; that is, discrimina-

Figure 4. Predicted depth percepts for different combinations of convexity and disparity. The stimulus is composed of two regions separated by a contour (Fig. 2). The left panels show probability distributions associated with convexity and disparity expressed over depth (Δ). 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 represent the probability distributions derived from the natural-scene statistics (Fig. 3) for convex and concave reference regions, respectively. The black curves represent the distribution over depths derived from a noisy disparity signal that specifies that the reference region is nearer than the other region. The right panels show the disparity likelihood functions (dashed curves) as in the left panels and the posterior distributions (solid 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. More perceived depth is predicted with a convex than a concave reference region.

tion 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 Figure 4 should increase as disparity increases: i.e., the influence of convexity on depth judgments should increase as the disparity in the stimulus increases.
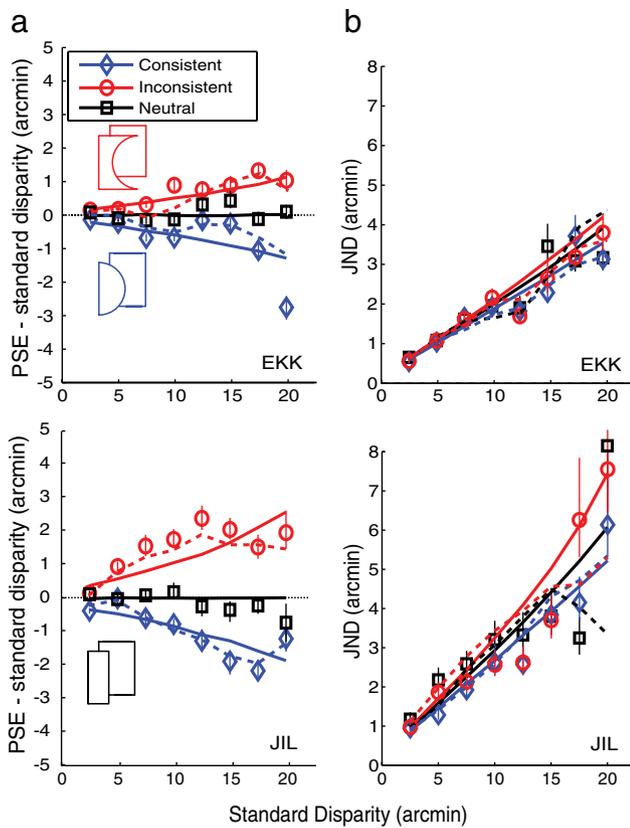
Experiment 1

The results from experiment 1 are shown in Figure 5. The PSE changes in the two panels of Figure 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, observers EKK and JIL respectively needed 1.1 and 3.9 arcmin less disparity in consistent than in inconsistent stimuli to perceive the same depth (i.e., $\text{PSE}_{\text{consistent}} - \text{PSE}_{\text{inconsistent}} = 1.1$ and 3.9 arcmin). This effect is consistent with observers using the relationship 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 et al., 1990; Morgan et al., 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.

Figure 5b shows that 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 observer JIL who had higher discrimination thresholds than EKK, the expected result if both observers had internalized the same natural-scene statistics for convexity and depth.

Experiment 2

To make sure that the observations of experiment 1 would generalize, we performed a shorter version of the experiment with more observers. Figure 6a shows that the seven observers exhibited the same pattern of results as the observers in experiment 1. Figure 6b shows the results averaged across observers. On aver-

Figure 5. Results from experiment 1. Upper and lower rows show data from observers EKK and JIL, respectively. **a**, Effects of contour shape and disparity on perceived depth. The abscissa is the disparity of the standard stimulus. The ordinate is the disparity of the comparison stimulus that on average had the same perceived depth as the standard (the PSE) minus the disparity of the standard. This is the disparity increment of the comparison stimulus relative to the standard disparity needed to match the perceived depth in the standard. Blue indicates the neutral–consistent stimulus pairing, red the neutral–inconsistent pairing, and black the neutral–neutral pairing. The symbols represent the mean of the cumulative Gaussian that best fit the raw psychometric data in each condition. Error bars represent bootstrapped 95% confidence intervals of the mean. 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. Dotted lines represent predictions of the nonparametric 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. **b**, JNDs plotted against the disparity of 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. The three sets of data are for neutral vs neutral (black), consistent vs consistent (blue), and inconsistent vs inconsistent (red). Symbols represent the SD of the best-fitting cumulative Gaussian for each condition. Error bars represent 95% confidence intervals on the SD. Dotted and solid lines are the predictions of the best-fitting probabilistic models (nonparametric and power law, respectively).

age, consistent stimuli required approximately 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 versus consistent and inconsistent versus inconsistent; or experimental: conditions in which the standard and comparison were different: consistent versus inconsistent and inconsistent versus 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 versus consistent and consistent versus inconsistent conditions and between inconsistent

versus inconsistent and inconsistent versus consistent conditions were both significant. The results from experiment 2 show that the effect of contour convexity is generally observed.

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 et al., 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 Figures 5a and 6a,b. We circumvented this problem by using the method of adjustment (Gillam et al., 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. Figure 6c shows the results for the 325 cm viewing distance. On average, consistent stimuli required approximately 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. Figure 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 toward 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.

Results for modeling of psychophysics

We first examined how the disparity likelihood distributions estimated from the psychophysical data compared with previous findings in the literature. Figure 7 shows that the agreement between the two was quite 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.
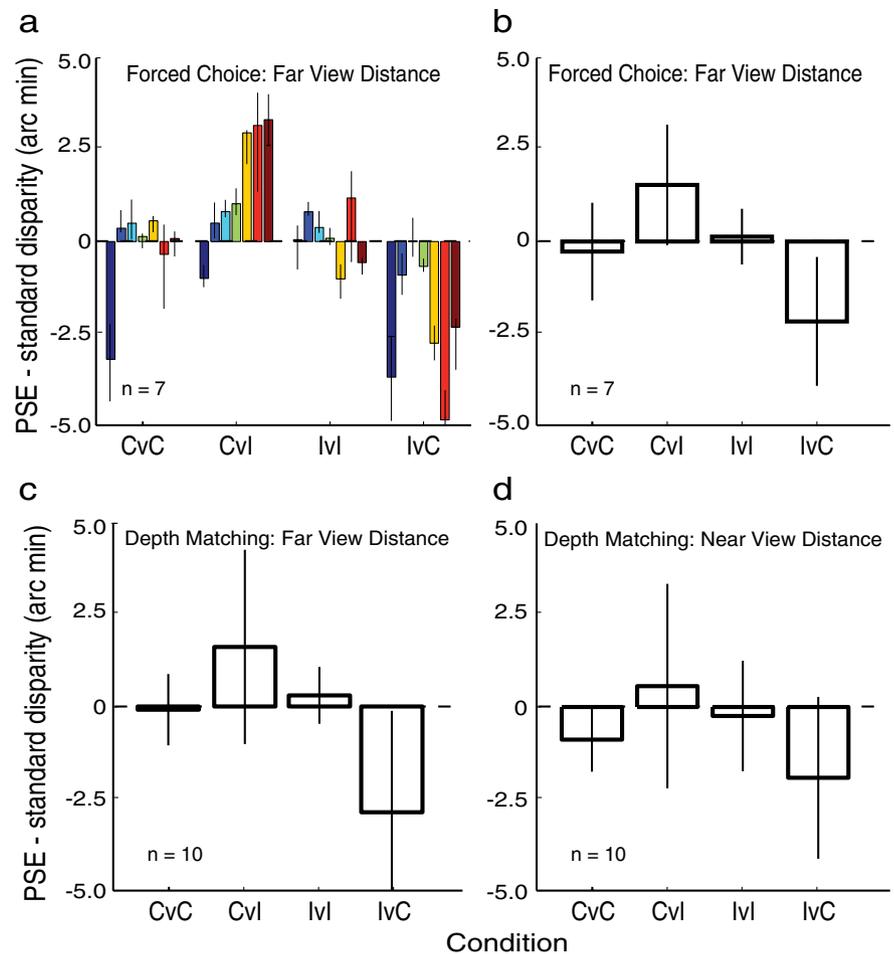
Figure 8 shows the convexity–depth distributions we estimated from the psychophysical data in experiment 1. Figure 8a shows the estimated distributions when we used the nonparametric model, a model that makes very few assumptions about the shapes of the distributions, but has 26 free parameters. The estimated distributions are similar in many respects to the convexity–depth distributions recovered from the natural-scene statistics (Fig. 3): (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; and (3) the estimated distributions are approximately linear in a log–log plot, which means that they are similar to power laws. Figure 8b shows the estimated distributions when we used the power-law model, which has only five free parameters (two for disparity and three for convexity–depth distributions).

Figure 9 shows that the power-law model provides nearly as good a fit to the data as the nonparametric model, even though the power-law model had many fewer 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. We computed an upper bound on goodness of fit by fitting separate cumulative Gaussians to each experimental condition. This Gaussian-fit model has 80 free parameters. We then normalized the SSEs so that a goodness-of-fit value of 1 represents an upper bound: specifically, our measure of goodness of fit was $SSE_{Gaussian}/SSE_{model}$. We computed the normalized 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 nonparametric models (Fig. 9). The nonparametric and power-law models provided excellent and nearly equal fits to the data. We conclude that power laws are excellent descriptions 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 differed in one interesting respect: the effect of convexity was larger in the psychophysically estimated distributions). $k_{difference}$, the difference between the exponents of the best-fitting power laws from the experiment was ~4.4 (Fig. 8b), while $k_{difference}$ from the natural statistics was ~0.4 (Fig. 3c).
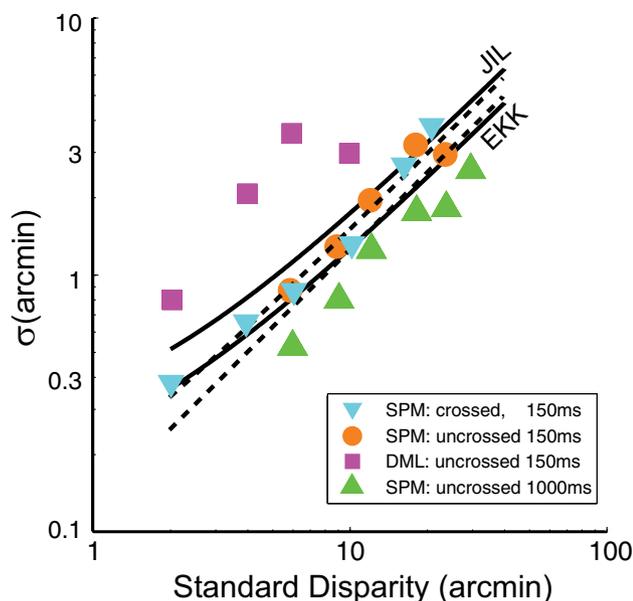
There is a plausible explanation for this discrepancy. The contours presented in the psychophysical experiments were different from the great majority of contours analyzed in the natural scenes. The contours in the experiments were circular and high contrast. Many contours in the natural-scene dataset had varying curvatures (i.e., convex at some spatial scales and concave at others, like the silhouette of a tree) and low contrast. Perhaps the differences in contour properties were the cause of the discrepancy between $k_{difference}$ for the natural-scene and psychophysical data. To test this idea, we recomputed the natural-scene statistics for contours with greater contrast, more consistent convexity, and greater average convexity. Figure 10a plots $k_{difference}$ as a function of the lowest contrast among the contours included in the analysis. As minimum contrast increases (i.e., as average con-

Figure 6. Results from experiments 2 and 3. **a**, Effect of contour shape for the seven observers in experiment 2. The abscissa represents the four stimulus pairings: 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. The ordinate represents the disparity of the comparison stimulus that on average yielded the same perceived depth of the standard (the PSE) minus the disparity of the standard. 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**, Data from experiment 2 averaged across observers. The abscissa and ordinate are the same as in **a**. 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 1 SD of the group mean. **c**, Results of experiment 3 with a viewing distance of 325 cm. The abscissa and ordinate are the same as in **b**. Data are averaged across observers. **d**, Results of experiment 3 with a viewing distance of 85 cm. Data are averaged across observers.

trast increases), $k_{difference}$ increases monotonically. Figure 10b plots $k_{difference}$ as a function of the consistency of convexity across spatial scale. As convexity becomes more consistent (e.g., contours become more circular), $k_{difference}$ increases monotonically. Figure 10c plots $k_{difference}$ as a function of the convexity value. When convexity is near zero (e.g., straight contours), there is little difference in the convexities of image regions on either side of the contour and $k_{difference}$ is low. As convexity increases, the analyzed contours become more similar to the contours presented in the psychophysical experiment, and $k_{difference}$ increases. We would like to have been able to select contours on all three criteria combined, but doing so leads to too few samples to compute meaningful statistics.

The analysis in Figure 10 shows that restricting the contours in the natural-scene dataset to be more similar to the contours in the psychophysical experiment makes the natural-scene convexity–depth distribution more similar to the psychophysically recovered distributions. Unfortunately, we cannot further restrict the

Figure 7. Disparity increment thresholds from the literature and recovered from our modeling. The curves represent the SDs of the disparity likelihood distributions estimated from experiment 1. They are plotted as a function of the disparity of the standard stimulus. Solid and dashed curves represent the estimated values when the data were fit with the nonparametric and power-law models, respectively. There are two pairs of curves for each observer. Symbols represent disparity increment thresholds for different subjects and conditions in McKee et al. (1990). In that report, threshold was defined as the 75% point on the psychometric function, so we transformed the data to match our 84% criterion.
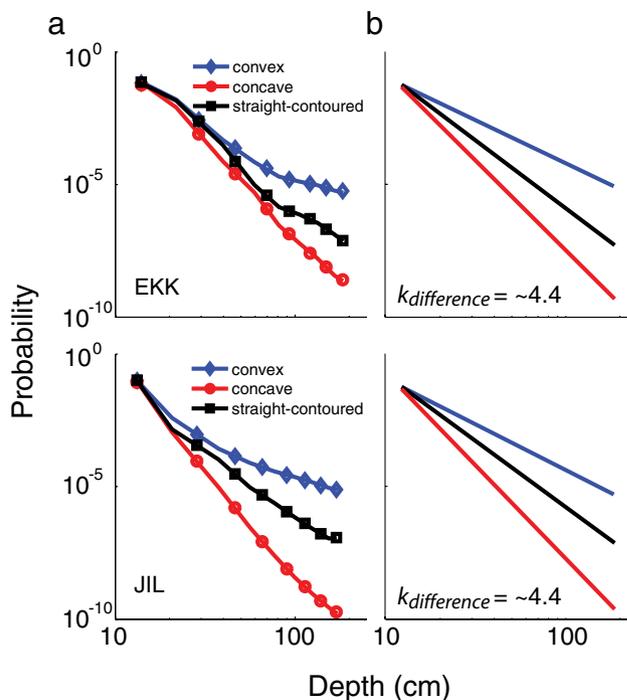
contours selected from the scene database to make them even more similar to the experimental stimuli because doing so yields too few samples to calculate meaningful statistics. As more natural-scene data become available, one could pursue this further.

Discussion

The depth information in figure–ground cues

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

Many other figure–ground cues could affect metric depth percepts in similar manner. Consider the cue of lower region (Vecera et al., 2002). Because objects typically rest on the ground plane, positions low in the visual field are likely to be nearer to the viewer than positions high in the field (Huang et al., 2000; Potetz and Lee, 2003; Yang and Purves, 2003b,. In our analysis of natural-scene statistics, we observed that the asymmetry in the distribution of depth was affected more by elevation in the visual

Figure 8. Convexity–depth distributions estimated from experiment 1. Probability is plotted as a function of the separation in depth between the adjacent regions. 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 estimated with the nonparametric 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 estimated with the power-law model. Note the similarity to the nonparametric distributions in **a**. One free parameter had to be set to uniquely determine these fits (see Materials and Methods), but the differences between the estimated distributions were unaffected by the value of that parameter. The SEs for the best-fitting parameters were estimated by redoing the fitting procedure 30 times with bootstrapped sets of psychometric data. The SE intervals were about the same as the line widths in **a** and **b**, so they are not plotted.

field than by contour convexity. We therefore expect 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 and 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 et al., 2006). By the traditional taxonomy, the depth cues in this stimulus—T-junctions (Palmer, 1999), surroundedness, and size (Rubin, 1921)—provide only ordinal information. 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, 1996). If cues to occlusion provide only distance-ordering information, standard models cannot explain these motion-in-depth percepts, nor the vergence eye movements induced by the second stimulus. However, if occlusion cues provide metric information, as proposed here, the percepts and eye movements are readily understood.

Other depth cues

Many other depth cues such as aerial perspective (Troscianko et al., 1991), shading (Koenderink and van Doorn, 1992), and blur (Mather and Smith, 2000) are regarded as nonmetric because from the cue value alone, one cannot uniquely determine depth. We propose that this taxonomy is unnecessary and that the visual system uses the information in those cues probabilistically to refine estimates of metric depth. By capitalizing on the statistical relationship between images and the environment to which our visual systems have been exposed, the probabilistic approach will yield a richer understanding of how we perceive 3D layout.
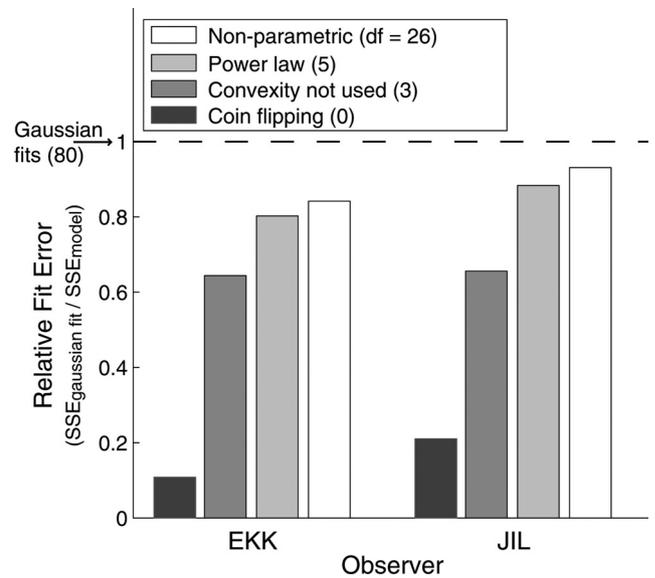
The usefulness of treating depth cues in a probabilistic framework is further evidenced by recent results in computer vision (Hoiem et al., 2005; Saxena et al., 2005) in which machine-learning techniques were used to combine statistical information about depth and surface orientation provided by a diverse set of image features. From the information in these features, the algorithms could generate reasonably accurate estimates of 3D scene layout. These results confirm that useful metric information is available from image features that have traditionally been considered nonmetric depth cues. The results also reveal that useful depth information is available from image features that have not yet been identified as depth cues.

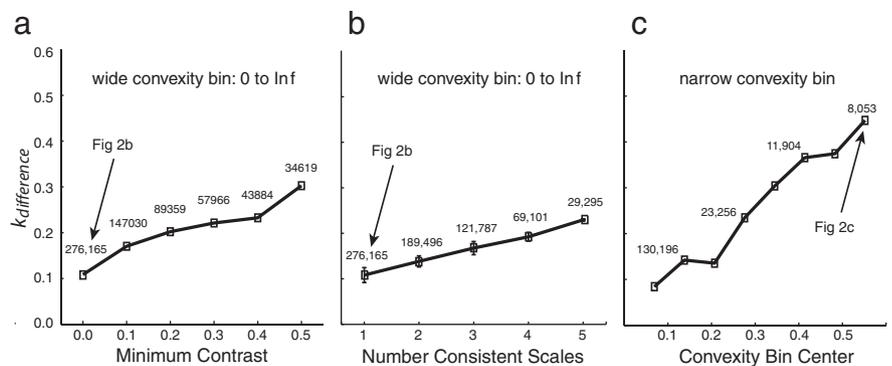## Neurophysiology and figure–ground cues

Figure–ground cues affect the responses of many neurons in early visual areas. For example, the responses of V2 and V4 neurons to the figure–ground cues of surroundedness and size have been studied. A small square of one luminance— defined by surroundedness and size—was presented on a uniform background of another luminance (Zhou et al., 2000). Even though the information specifying "figure" was outside the classical receptive field, many neurons responded more to the contour when the square was on the preferred side of the contour. Other research has found that many V2 neurons selective for disparity-defined contours specifying that one side of the contour is near, also respond to figure–ground-defined contours indicating that the same side is near. Importantly, these neurons respond more vigorously when disparity and figure–ground cues are consistent than when they are inconsistent (Qiu and von der Heydt, 2005). It would be interesting to investigate whether convexity modulates responses in similar manner.

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 depend strongly on the patch's disparity relative to the bar: They barely respond when disparity specifies that the patch is in the plane of the bar or behind it, but respond vigorously when disparity specifies that the patch is in front (Sugita, 1999; Bakin et al., 2000).

These neurophysiological results are consistent with the view that a representation of cue-invariant object boundaries emerges in processing from V1 through V2 to V4 (Orban, 2008). The neurons in-

Figure 9. Goodness of fit for four models. The SSE between model predictions and the psychometric data was calculated. Goodness of fit was indexed by $SSE_{Gaussian}/SSE_{model}$ so that 1 represents an upper bound. 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 error is shown for four models fit to the data: (1) Coin-flipping model (0 free parameters); (2) 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); (3) power-law model (5 free parameters); (4) nonparametric 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 McKee et al. (1990) and Morgan et al. (2000). We conclude the power-law model is a good description of the internal convexity– depth distributions.

Figure 10. The effect of contour properties on convexity– depth statistics. Each panel plots $k_{difference}$—the difference between the convex and concave power-law exponents (Fig. 3c)—for different subsets of the contours in the natural-scene dataset. **a**, $k_{difference}$ as a function of contrast. Contours with different ranges of contrast were analyzed. The abscissa is the lowest contrast in the tested range (the highest contrast was always the same), so increasing abscissa values corresponds to higher average contrast. $k_{difference}$ increases monotonically with increasing contrast. The arrow shows the conditions plotted in Figure 3b. **b**, $k_{difference}$ as a function of the number of spatial scales for which the sign of region convexity is consistent. With increasing consistency, contours become more circular. $k_{difference}$ increases monotonically as the consistency of convexity increases. **c**, $k_{difference}$ as a function of the mean convexity of a narrow range included in the analysis. The width of the narrow bin is shown in Figure 3a. The arrow shows the conditions plotted in Figure 3c. As the mean convexity increases and more closely matches the convexity presented in psychophysical experiment, $k_{difference}$ increases.

volved in this representation may mediate the perceptual effects reported here. A given disparity stimulates numerous cortical neurons with different preferred disparities (Poggio et al., 1988). Presumably, perceived depth is determined by a read-out from the noisy population response. When a figure–ground cue is added specifying a skewed distribution of depths, the cue may increase the response of neurons with disparity preferences consistent with the cue and/or decrease the responses of neurons with inconsistent disparity preferences. This could yield an increase in the perceived metric depth.

Natural-scene statistics

The importance of natural-scene statistics in perceptual tasks was first articulated by Brunswik (Brunswik and 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 et al., 2001; Elder and Goldberg, 2002; Ren and Malik, 2002), figure–ground assignment (Fowlkes et al., 2007), and length estimation (Howe and Purves, 2002). It is useful to distinguish this work from a much larger literature on the relationship between the statistics of natural images and the efficiency of neural coding in early visual pathways (Barlow, 1961; Simoncelli and Olshausen, 2001). To determine encoding efficiency, one must know the statistics of the image properties to be encoded, so recent work has focused on measuring natural-image statistics and determining if cortical neurons exploit those statistics (Olshausen and Field, 1996; Vinje and Gallant, 2000).

It is unclear, however, that efficient encoding is the primary task of early visual processing. To make strong claims about the role of natural statistics in perception, we believe it is important to study tasks that are known to be critical for the biological system under study. This makes it possible to distinguish suboptimal performance from a mistaken hypothesis about what the system does. Surely, estimating the 3D layout of the environment is a crucial task because such estimation is required for guidance of visuomotor behavior (Geisler, 2008). It is plausible, therefore, that natural selection would work toward the incorporation of 3D information contained in natural-scene statistics.

Closing the loop in probabilistic models of perception

A key idea in the probabilistic characterization of perception is that perceptual systems have internalized the statistical properties of their sensors and the natural environment, and that the systems use those properties efficiently to make optimal perceptual estimates. Many observations in the literature are compatible with this framework. These include the findings that the variances of multiple sources of sensory information are integrated optimally (Ernst and Banks, 2002; Knill and Saunders, 2003; Alais and Burr, 2004). There has also been work motivated by the idea that prior information affects perception optimally (Mamassian and Goutcher, 2001; Weiss et al., 2002). However, previous work has not demonstrated that perceptual systems do in fact behave optimally.

For example, perceived speed decreases when the contrast of the moving stimulus is reduced (Weiss et al., 2002; Stocker and Simoncelli, 2006). This effect can be understood if we assume that the visual system has a prior for slow motion. With lower contrast, the variance of sensory measurement increases so the prior is expected to have greater influence; it pulls perceived speed toward zero. By varying contrast, Stocker and Simoncelli (2006) estimated the speed prior from psychophysical data. Although the estimated prior is consistent with several phenomena in mo-

tion perception (Weiss et al., 2002), there is no evidence that the estimated prior actually matches the distributions of speeds encountered in natural viewing. Thus, it cannot be argued from psychophysical analysis alone that the visual system has accurately internalized statistical properties of the environment.

The work described here attempts to close this gap by measuring the statistics from natural scenes and seeing whether observers have internalized those same statistics. We showed that people behave as if they use an asymmetric convexity–depth distribution when making depth judgments in the presence of a region-bounding contour. The convexity–depth distribution that best explains their behavior is qualitatively similar to the distribution for natural scenes. Our experiment thus demonstrates the ecological validity of convexity as a cue to metric depth and explains its usefulness to the visual system. But we too have not yet closed the above-mentioned gap because the distributions estimated from the psychophysical data were not quantitatively the same as the natural-scene distributions. Thus, with the possible exception of findings in categorical perception (Geisler et al., 2001; Elder and Goldberg, 2002; Ren and Malik, 2002, Fowlkes et al., 2007), the field still awaits evidence that the internalized statistics really do match the statistics of natural scenes.

Summary

In the analysis of natural-scene statistics, we found that the convexity of an image region provides information about the probability of different depths across the contour bounding that region. We constructed a probabilistic model of how this information can 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.

References

Alais D, Burr D (2004) The ventriloquist effect results from near-optimal cross-modal integration. Curr Biol 14:257–262.

Backus BT, Banks MS, van Ee R, Crowell JA (1999) Horizontal and vertical disparity, eye position, and stereoscopic slant perception. Vision Res 39:1143–1170.

Bakin JS, Nakayama K, Gilbert CD (2000) Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. J Neurosci 20:8188–8198.

Barlow HB (1961) Possible principles underlying the transformation of sensory messages. In: Sensory communication (Rosenblith WA, ed), pp 217–234. Cambridge, MA: MIT.

Berkeley G 1709 An essay toward a new theory of vision. Dublin: Pepyat.

Bertamini M, Martinovic J, Wuerger SM (2008) Integration of ordinal and metric cues in depth processing. J Vis 8:1–12.

Bruce V, Green PR, Georgeson MA (2003) Visual perception: Physiology, psychology, and ecology. New York: Psychology.

Brunswik E, Kamiya J (1953) Ecological cue validity of "Proximity" and of other gestalt factors. Am J Psychol 66:20–32.

Burge J, Peterson MA, Palmer SE (2005) Ordinal configural cues combine with metric disparity in depth perception. J Vis 5:534–542.

Cormack LK, Landers DD, Ramakrishnan S (1997) Element density and the efficiency of binocular matching. J Opt Soc Am 14:723–730.

Driver J, Baylis GC (1996) Edge-assignment and figure-ground organization in short term visual matching. Cogn Psychol 31:248–306.

Elder JH, Goldberg RM (2002) Ecological statistics of Gestalt laws for the perceptual organization of contours. J Vis 2:324–353.

Engel SA, Remus DA, Sainath R (2006) Motion from occlusion. J Vis 6:649–652.

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433.

Fowlkes CC, Martin D, Malik J (2007) Local figure/ground cues are valid for natural images. J Vis 7:1–9.

Geisler WS (2008) Visual perception and the statistical properties of natural scenes. Annu Rev Psychol 59:10.1–10.26.

Geisler WS, Perry JS, Super BJ, Gallogly DP (2001) Edge co-occurrence in natural images predicts contour grouping performance. Vision Res 41:711–724.

Ghahramani Z, Wolpert DM, Jordan MI (1997) Computational models of sensorimotor integration. In: Self-organization, computational maps and motor control (Morasso PG, Sanguineti V, eds), pp 117–147. Amsterdam: Elsevier.

Gibson BS, Peterson MA (1994) Does orientation-independent object recognition precede orientation-dependent recognition? Evidence form a cuing paradigm. J Exp Psychol 28:299–316.

Gillam BJ, Anderson BL, Rizwi F (2009) Failure of facial configural cues to alter metric stereoscopic depth. J Vis 9:3,1–5.

Green DM, Swets JA (1966) Signal detection theory and psychophysics. New York: Wiley.

Hillis JM, Watt SJ, Landy MS, Banks MS (2004) Slant from texture and disparity cues: optimal cue combination. J Vis 4:967–992.

Hoiem D, Efros AA, Hebert M (2005) Geometric context from a single image. IEEE International Conference on Computer Vision. Beijing, October.

Howard IP, Rogers BJ (2002) Seeing in depth. Vol 2. Depth perception. New York: Oxford UP.

Howe CQ, Purves D (2002) Range image statistics can explain the anomalous perception of length. Proc Natl Acad Sci U S A 99:13184–13188.

Huang J, Lee AB, Mumford D (2000) Statistics of range images. Proceedings of the IEEE Conference on Computational Vision and Pattern Recognition, Vol 1, pp 324–331. IEEE: Hilton Head Island, SC.

Kanizsa G, Gerbino W (1976) Convexity and symmetry in figure-ground organization. In: Vision and artifact (Henle M, ed), pp 25–32. New York: Springer.

Knill DC, Saunders JA (2003) Do humans optimally integrate stereo and texture information to slant? Vision Res 43:2539–2558.

Koenderink JJ, van Doorn AJ (1992) Surface shape and curvatures scales. Image Vis Comput 10:557–565.

Landy MS, Maloney LT, Johnston EB, Young M (1995) Measurement and modeling of depth cue combination: in defense of weak fusion. Vision Res 35:389–412.

Levitt H (1971) Transformed up-down methods in psychoacoustics. J Opt Soc Am 49:467–477.

Mamassian P, Goutcher R (2001) Prior knowledge on the illumination position. Cognition 81:B1–B9.

Martin D, Fowlkes CC, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Vancouver: ICCV.

Mather G, Smith DR (2000) Depth cue integration: stereopsis and image blur. Vision Res 40:3501–3506.

McKee SP, Levi DM, Bowne SF (1990) The imprecision of stereopsis. Vision Res 30:1763–1779.

Metzger F (1953) Gesetze des Sehens. Frankfurt: Waldemar Kramer.

Morgan MJ, Watamaniuk SN, McKee SP (2000) The use of an implicit standard for measuring discrimination thresholds. Vision Res 40:2341–2349.

Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381:607–609.

Orban GA (2008) Higher order visual processing in macaque extrastriate cortex. Physiol Rev 88:59–89.

Palmer SE (1999) Vision science: photons to phenomenology. Cambridge, MA: Bradford Books, MIT.

Peterson MA, Skow E (2008) Inhibitory competition between shape properties in figure-ground perception. J Exp Psychol 32:251–267.

Peterson MA, Harvey EM, Weidenbacher HJ (1991) Shape recognition contributions to figure-ground organization: which route counts? J Exp Psychol Hum Percept Perform 17:1075–1089.

Poggio GF, Gonzalez F, Krause F (1988) Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. J Neurosci 8:4531–4550.

Potetz B, Lee TS (2003) Statistical correlations between 2D images and 3D structures in natural scenes. J Opt Soc Am 20:1292–1303.

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.

Ren X, Malik J (2002) A probabilistic multi-scale model for contour completion based on image statistics. Vol 1, pp 312–327. Copenhagen: ECCV.

Ringach DL (1996) Binocular eye movements caused by the perception of three-dimensional structure from motion. PhD dissertation, New York University.

Rubin E (1921) Visuell wahrgenommene Figuren. Copenhagen: Glydenalske Boghandel.

Saxena A, Chung SH, Ng A (2005) Learning depth from single monocular images. Neural information processing systems. In: Proceedings of Conference on Neural Information Processing Systems (NIPS), pp 1161–1168. Cambridge, MA: MIT.

Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. Annu Rev Neurosci 24:1193–1216.

Stocker AA, Simoncelli EP (2006) Noise characteristics and prior expectations in human visual speed perception. Nat Neurosci 9:578–585.

Sugita Y (1999) Grouping of image fragments in primary visual cortex. Nature 401:269–272.

Troscianko T, Montagnon R, Le Clerc J, Malbert E, Chanteau PL (1991) The role of color as a monocular depth cue. Vision Res 31:1923–1929.

Vecera SP, Vogel EK, Woodman GF (2002) Lower region: a new cue for figure-ground assignment. J Exp Psychol 131:194–205.

Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287:1273–1276.

von Helmholtz H (1867) Handbuch der physiologischen Optik. Leipzig: Leopold Voss.

Watt SJ, Akeley K, Ernst MO, Banks MS (2005) Focus cues affect perceived depth. J Vis 5:834–862.

Weiss Y, Simoncelli EP, Adelson EH (2002) Motion illusions as optimal percepts. Nat Neurosci 5:598–604.

Wichmann FA, Hill NJ (2001) The psychometric function: I. Fitting, sampling and goodness-of-fit. Percept Psychophys 63:1293–1313.

Yang Z, Purves D (2003a) Image/source statistics in natural scenes. Network: Computation in Neural Systems 14:371–390.

Yang Z, Purves D (2003b) A statistical explanation of visual space. Nat Neurosci 6:632–640.

Zhou H, Friedman HS, von der Heydt R (2000) Coding of border ownership in monkey visual cortex. J Neurosci 20:6594–6611.