Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant

David C. Knill

Center for Visual Science, University of Rochester,
Rochester, NY, USA

Most research on depth cue integration has focused on stimulus regimes in which stimuli contain the small cue conflicts that one might expect to normally arise from sensory noise. In these regimes, linear models for cue integration provide a good approximation to system performance. This article focuses on situations in which large cue conflicts can naturally occur in stimuli. We describe a Bayesian model for nonlinear cue integration that makes rational inferences about scenes across the entire range of possible cue conflicts. The model derives from the simple intuition that multiple properties of scenes or causal factors give rise to the image information associated with most cues. To make perceptual inferences about one property of a scene, an ideal observer must necessarily take into account the possible contribution of these other factors to the information provided by a cue. In the context of classical depth cues, large cue conflicts most commonly arise when one or another cue is generated by an object or scene that violates the strongest form of constraint that makes the cue informative. For example, when binocularly viewing a slanted trapezoid, the slant interpretation of the figure derived by assuming that the figure is rectangular may conflict greatly with the slant suggested by stereoscopic disparities. An optimal Bayesian estimator incorporates the possibility that different constraints might apply to objects in the world and robustly integrates cues with large conflicts by effectively switching between different internal models of the prior constraints underlying one or both cues. We performed two experiments to test the predictions of the model when applied to estimating surface slant from binocular disparities and the compression cue (the aspect ratio of figures in an image). The apparent weight that subjects gave to the compression cue decreased smoothly as a function of the conflict between the cues but did not shrink to zero; that is, subjects did not fully veto the compression cue at large cue conflicts. A Bayesian model that assumes a mixed prior distribution of figure shapes in the world, with a large proportion being very regular and a smaller proportion having random shapes, provides a good quantitative fit for subjects' performance. The best fitting model parameters are consistent with the sensory noise to be expected in measurements of figure shape, further supporting the Bayesian model as an account of robust cue integration.

Keywords: cue integration, robust estimation, Bayes, depth perception, slant perception

Citation: Knill, D. C. (2007). Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision, 7*(7):5, 1–24, http://journalofvision.org/7/7/5/, doi:10.1167/7.7.5.

Introduction

Images contain many different cues to the three-dimensional (3D) layout of objects in a scene—retinal disparity, motion, texture, figure shape, and shading. The visual system integrates these cues to estimate objects' 3D properties both for perception and to guide action. When different cues suggest similar values for a scene parameter (curvature, slant, etc.), one can reasonably approximate cue integration as a linear combination of the estimates suggested by each cue individually. A large body of contemporary research has focused on how the human visual system integrates cues when operating in this linear regime (Alais & Burr, 2004; Jacobs, 2002; Johnston, Cumming, & Landy, 1994; Johnston, Cumming, & Parker, 1993; Landy, Maloney, Johnston, & Young, 1995; Young, Landy, & Maloney, 1993). Thus, for example, research has shown that humans weight cues, both within and across sensory modalities, according to their relative reliabilities. As cue reliability changes across stimulus conditions, so do the weights that subjects give to the cues (Alais & Burr, 2004; Ernst & Banks, 2002; Hillis, Watt, Landy, & Banks, 2004; Knill & Saunders, 2003). The fact that cue weights in a local linear model of cue integration change across stimulus conditions reflects one form of global nonlinearity in how the brain integrates cues. Another potential form of nonlinearity can arise when sensory cues suggest very different estimates of a scene parameter, requiring the use of nonlinear, robust strategies for integrating cues (Landy et al., 1995). This article describes a Bayesian approach to modeling cue integration in large-conflict situations and describes two experiments designed to test a Bayesian model for integrating figural shape cues and binocular disparity cues to surface slant. The analysis provides a test of the explanatory power of the Bayesian approach for characterizing nonlinear, robust cue integration behaviors.

Pictures provide a prototypical, if somewhat artificial, example of large cue conflicts. Pictorial cues suggest the 3D layout of the photographed scene, but binocular disparities suggest a flat surface. In these cases, our brains resolve the

doi: 10.1167/7.7.5 Received July 21, 2006; published May 23, 2007 ISSN 1534-7362 © ARVO

conflict by supporting two modes of viewing—a real mode and a depicted mode. Asked to grasp the page, we would orient our hands to match the slant of the page; however, we also see the depicted surfaces as having slant, curvature, and variations in depth that are different than those of the printed page. What happens in the real world when faced with large cue conflicts, when an observer cannot use the pictorial explanation to explain away the conflict? How, for example, does the brain interpret retinal image information when the texture projected from a surface suggests an orientation very different from that suggested by binocular disparities?

One answer is that the visual system should veto one of the two cues in a process akin to outlier rejection in statistics (Landy et al., 1995). Which cue to veto could depend on which one is least reliable or, when more than two cues are available, which one is most inconsistent with the others. As with many approaches to outlier rejection in statistics, these strategies are heuristics for deciding which cue (or cues) to reject. Consideration of the situations in natural viewing that lead to large cue conflicts suggests a principled Bayesian approach to the problem. The fundamental observation underlying the approach is that most cues rely on a mixture of possible prior assumptions or constraints about objects in the world (Knill, 2003; Yuille & Bulthoff, 1996). Some constraints render cues reliable and some less so. Multiple cues can interact to effectively determine which constraints apply in a given scene. Texture information provides a prototypical example. Surface textures may be homogeneous and isotropic (have no global orientation); they may simply be homogeneous or they may be neither. At a finer level of categorization, some homogeneous textures are stochastic, whereas others are regular. The information provided by image textures depends critically on which model applies to the surface texture being viewed. Thus, for example, when stereoscopic disparities specify a slant very different from the slant suggested by texture, as interpreted using an isotropy assumption, a rational visual system might determine that the most likely interpretation is that the surface texture is not isotropic (but is perhaps homogeneous). This would appear as cue vetoing or, at least, down-weighting the texture cue (Knill, 2003). Similar observations apply to almost all monocular depth cues (e.g., motions may be rigid, elastic, etc.). In the real world, the visual system must necessarily take into account the possibility that any of these prior models might apply to an object property when interpreting a visual cue.

A normative model for robust cue integration

The information provided by a pair of cues about a surface property or set of surface properties, $\vec{S}$, is given by the posterior probability density function $p\left(\vec{S}\,|\vec{I}_a,\vec{I}_b\right)$, where $\vec{I}_a$ represents the image measurements associated with cue a and $\vec{I}_b$ represents the image information associated with cue b. When $p\left(\vec{S}\,|\vec{I}_a,\vec{I}_b\right)$ is narrowly peaked around a particular set of values of $\vec{S}$, the image information reliably determines the perceptual estimate of $\vec{S}$. Using Bayes' rule and assuming that the cues are conditionally independent (e.g., that the sensory noise associated with each cue is independent), we can write the posterior as the product of likelihood functions associated with each cue and a prior density function on $\vec{S}$,

$$p\left(\vec{S}\,|\vec{I}_a,\vec{I}_b\right)=\frac{p\left(\vec{I}_a|\vec{S}\right)p\left(\vec{I}_b|\vec{S}\right)p\left(\vec{S}\right)}{p\left(\vec{I}_a,\vec{I}_b\right)}. \qquad (1)$$

The denominator is a constant (it depends only on the given image measurements and not on $\vec{S}$); thus, we can rewrite Equation 1 as

$$p\left(\vec{S}\,|\vec{I}_a,\vec{I}_b\right)=kp\left(\vec{I}_a|\vec{S}\right)p\left(\vec{I}_b|\vec{S}\right)p\left(\vec{S}\right). \qquad (2)$$

When the likelihood functions associated with each cue and the prior are all Gaussian, both the mean and the mode of the posterior density is a weighted sum of the means (or modes) of those functions. The weights are inversely proportional to the variance of each of the functions, leading to the now well-tested hypothesis that subjects, when they integrate cues linearly, should weight sensory cues in inverse proportion to their individual uncertainty (Alais & Burr, 2004; Ernst & Banks, 2002; Hillis et al., 2004; Knill & Saunders, 2003). A little thought, however, reveals that the Gaussian model is not a good model of the true likelihood functions that should be associated with each cue. In this section, we will explore one particular feature of more naturalistic models of the likelihood functions that, when built into a Bayesian observer, gives rise to robust cue integration behavior: apparent down-weighting of one or another cue in the presence of large conflicts.

Most monocular depth cues derive their informativeness from prior constraints on *hidden* parameters describing object or scene properties that an observer is not necessarily estimating. For example, the shapes of figures in an image only provide cues to the figure's 3D orientation because of statistical constraints on the shapes of figures to be found in our environment. Because figures come in different categories (symmetric, isotropic, random, etc.), the true prior probability density function over the space of shape parameters is really a mixture of qualitatively different priors. The important consequence of this structure for cue integration is that the likelihood function associated with a cue that depends on a mixture of priors is itself an additive mixture of likelihood functions. Each component likelihood function is derived using a different prior model and then weighted by the probability that the prior model applies to the object being viewed (e.g., the probability that a figure is symmetric) and added together to form the full likelihood function for the cue. This is expressed in the equation

$$p\left(\vec{I}\,|\vec{S}\right)=\pi_1 p\left(\vec{I}\,|\vec{S},M_1\right)+\pi_2 p\left(\vec{I}\,|\vec{S},M_2\right)\ldots \qquad (3)$$

where $\vec{I}$ is a vector representing the image measurements associated with a cue, $\vec{S}$ is a vector representing the object parameters being estimated, and M_i are the different prior models used to compute the components of the mixed likelihood function. π_i are the probabilities associated with each model (e.g., the probability that a surface texture is isotropic).

In general, the likelihood functions resulting from such mixtures can be arbitrarily complex with, for example, multiple peaks for different values of $\vec{S}$. Much of the structure in the priors, however, is hierarchical, and the different prior models that could possibly apply to a given image can be arranged according to the degree to which they constrain the hidden parameters. This formally appears as a set of priors that restrict the space of allowable interpretations of the hidden parameters to lower and lower dimensional subspaces of the total parameter space. The information provided by the shapes of ellipses in the retinal image about surface slant provides a particularly simple example of this. As illustrated in Figure 1, human observers typically perceive elliptical figures in an image as slanted circles. This reflects a strong prior belief that circles are the most common form of ellipse found in our environment. Because not all ellipses in the world are circles, a reasonable prior model for ellipses in the world is that they come in two classes—randomly shaped ellipses and circles. The former would be defined by a prior density function over the range of aspect ratios. The latter would be defined by a density function that concentrates all of the probability at a single aspect ratio (one). The likelihood function for slant derived from the shape of an ellipse in the image is an additive mixture of the likelihoods associated with each of the two prior models on figure shapes in the world. The model likelihood functions depend both on the amount of noise in sensory measurements of aspect ratio and on the spread of the prior density function of aspect ratios associated with the models; therefore, the likelihood function for the circle prior will be narrower than the likelihood for the random ellipse prior. Figure 2 illustrates the calculation of this type of mixed likelihood function.

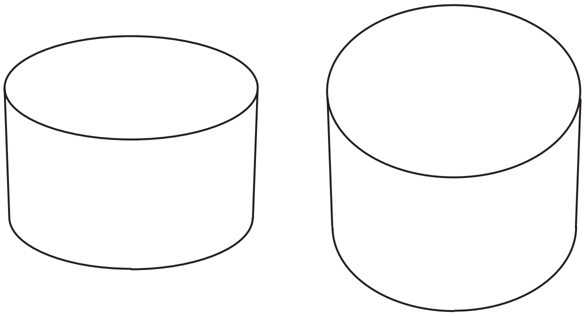

Figure 1. Both figures appear to be circular cylinders but with different orientations. The sides of the figures have the same lengths and orientations. Only the aspect ratios of the elliptical outlines at the tops and bottoms of the figures differ.

Figure 3 illustrates the behavior of a Bayesian model for estimating surface slant-from-figure shape information and stereoscopic information that incorporates a mixture of prior models for figure shape. The joint likelihood function computed from figure shape and stereopsis is the product of the mixed likelihood function for slant-from-figure shape and the one derived from stereoscopic information, which gives

$$\begin{aligned}p\left(\vec{I}_{\mathrm{s}},\vec{I}_{\mathrm{F}}|\vec{S}\right)&=p\left(\vec{I}_{\mathrm{s}}|\vec{S}\right)p\left(\vec{I}_{\mathrm{F}}|\vec{S}\right)\\&=\pi_{\mathrm{circle}}\,p\left(\vec{I}_{\mathrm{s}}|\vec{S}\right)p\left(\vec{I}_{\mathrm{F}}|\vec{S},\mathrm{circle}\right)\\&\quad+\pi_{\mathrm{ellipse}}\,p\left(\vec{I}_{\mathrm{s}}|\vec{S}\right)p\left(\vec{I}_{\mathrm{F}}|\vec{S},\mathrm{ellipse}\right), \qquad (4)\end{aligned}$$

where $\vec{I}_{\mathrm{F}}$ represents the image measurements that characterize figure shape (in our example, this would be aspect ratio) and $\vec{I}_{\mathrm{s}}$ represents the image measurements that characterize stereoscopic disparities. Which of the two terms dominates the likelihood function depends both on the prior probabilities associated with the two models for figure shape (π_{circle} and π_{ellipse}) and on whether the stereoscopic likelihood function is centered near the peak of the figure shape likelihood or is centered over one of its extended tails. When stereoscopic information suggests a slant similar to that suggested by the circle interpretation of a figure, the combined likelihood function is centered at a point that is well characterized by a weighted sum of the two. As the deviation between the two increases, the peak of the joint likelihood function shifts toward the peak of the stereoscopic likelihood function until, at high ‘‘conflicts’’, it almost perfectly aligns with the stereoscopic peak. At this point, a Bayesian estimator will appear to have nearly turned off the figure shape cue. This is because the stereoscopic information at large conflicts is not consistent with the circle model and the random ellipse model in the mixed likelihood function for figure shape dominates the combined likelihood.

Figure 3D shows how this behavior reflects itself in the weights that a Bayesian observer would appear to give to the compression cue as a function of the size of the cue conflict. Note that we use the term compression cue to refer to the slant suggested by the shape of the ellipse in

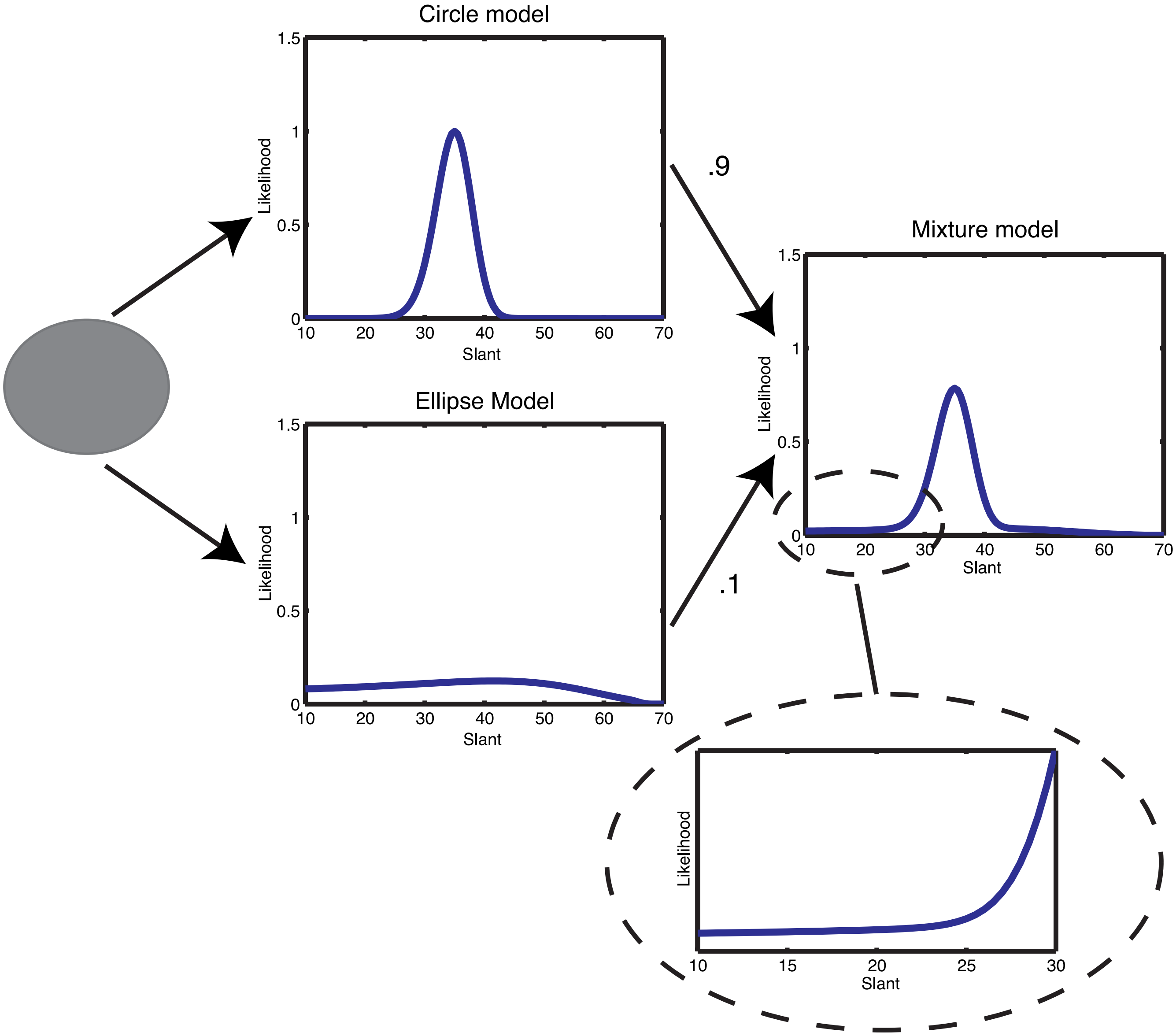

Figure 2. Given the shape of an ellipse in the retinal image, the likelihood function for slant is a mixture of likelihood functions derived from different prior models on the aspect ratios of ellipses in the world. The likelihoods shown above were derived by assuming that noise associated with sensory measurements of aspect ratio has a standard deviation of 0.03 (taken from thresholds for discriminating aspect ratios of ellipses; Regan & Hamstra, 1992), that the prior distribution of aspect ratios of randomly shaped ellipses in the world has a standard deviation of 0.25, and that 90% of ellipses in the world are circles. The mixture of narrow and broad likelihood functions creates a likelihood function with long tails, as shown in the blowup. The existence of these tails is the critical feature that supports robust cue integration.

the image under the assumption that the figure is a circle in the world. The three plots show the patterns of weights one would observe for an observer who assumes each of three prior models on figure shapes in the world. In the first model, all ellipses in the world are assumed to be circles. In the second model, all ellipses in the world are assumed to be randomly drawn from a set of ellipses with aspect ratios having a distribution that is peaked at 1 (biased toward circles) but has a standard deviation of 0.25 (circles are not a privileged category). The third model is a mixture of the first two. It reflects a world in which 90% of figures are circles, but 10% are drawn from the random set of ellipses characterized by Model 2. Note that for small cue conflicts, the Bayesian observer using the mixed model operates in a regime that is intermediate between what would be predicted from the two component models on figure shape; that is, both models contribute to the behavior of the observer. This is true even at the smallest conflicts, where the contribution of the random ellipse model decreases the apparent weight given to the compression cue. The apparent weight given to the compression cue decreases smoothly as the conflict

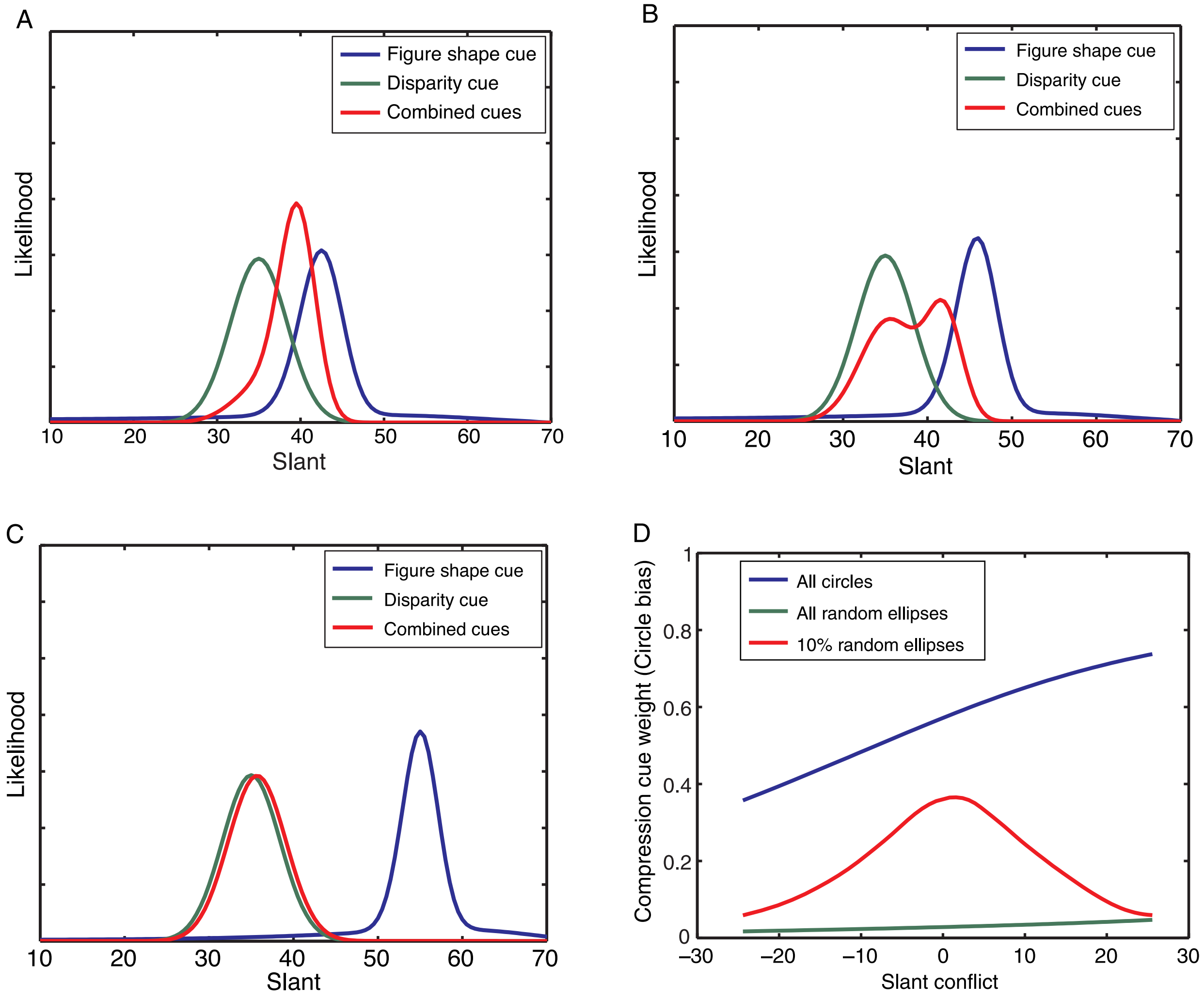

Figure 3. (A) When stereoscopically viewing an ellipse with an aspect ratio (in the world) close to 1 (in this case, 0.9), the likelihood function for the figure shape cue is shifted away from the likelihood function for the disparity cue because of the strong bias to interpret the figure as a circle. The figure shape likelihood function shown here is a long-tailed mixture of likelihoods derived using the same prior density and noise parameters used to generate the likelihood function in Figure 2. The stereoscopic likelihood function is Gaussian with a standard deviation of 3.5°, reflecting the uncertainty in slant-from-stereo discrimination judgments found experimentally (Hillis et al., 2004). The product of the likelihood functions lies intermediate between the peaks of the two cues' likelihood functions. The best estimate of slant in this case lies in between the estimates one would derive from either cue individually. (B and C) For projections of ellipses with aspect ratios very different from 1.0 (0.8 and 0.7, respectively), the combined cue likelihood function gradually shifts to become more concentrated near the slant-from-disparity likelihood function. For ellipses with aspect ratios very different from 1 (large cue conflicts), the stereoscopic likelihood function is concentrated over the tail of the figure shape likelihood. The long tail in the figure shape likelihood function causes the shift in the combined likelihood toward the stereoscopic likelihood. One can use these likelihood functions as the basis for an optimal Bayesian slant estimator by combining the joint likelihood with a prior on slant (the generic viewpoint prior is sin(σ)) and defining a cost function on errors in slant estimates. (D) Predicted compression cue weights for a Bayesian estimator that calculates the mean slant conditioned on the image information, plotted as a function of the difference between the slant suggested by the compression cue (the slant suggested by a circle interpretation of the figure) and the slant suggested by the disparity cue (assuming a stereoscopic slant of 35°). A compression cue weight of 0.5 reflects equal weighting of the compression cue and stereopsis. Compression cue weights are shown for an estimator that uses three different prior models for ellipses in the world—all circles, all random ellipses drawn from a distribution of aspect ratios (SD = 0.25), or 90% circles and 10% random ellipses drawn from the same distribution.

between cues increases until it asymptotes at the weights predicted by less constrained model of figures.

Several experimental results are consistent with a Bayesian model for robust cue integration. Knill (2003) has shown that subjects turn off the isotropy constraint for interpreting surface textures when viewing monocular images of slanted textures that have been strongly compressed in one direction (are very anisotropic). That is, as one compresses a surface texture by larger and larger amounts before projecting it into the image, subjects initially show biases to interpret the texture as isotropic, but eventually, the bias weakens and almost disappears. The result is consistent with a model that uses both texture foreshortening (which is subject to the isotropy bias) and texture scaling (which is not) as cues to slant. At large compression factors, the scaling information is strong enough to turn off the isotropy bias. The Bayesian model of robust cue integration predicts that measurements of the weights that subjects give to cues will spontaneously and smoothly vary as a function of the conflict between the cues. In particular, because stereoscopic cues do not rely on hidden prior assumptions on objects in the same way that pictorial cues do, subjects should appear to down-weight pictorial cues relative to binocular disparities as the conflict between the two increases (but see the discussion for possible violations of this behavior).

We performed two experiments to test whether subjects show this behavior when integrating the information provided by the shapes of 2D figures in the image and stereoscopic disparities to judge the slants of planar surfaces. We fit a Bayesian model to subjects' data to test whether a Bayesian account parameterized by reasonable levels of sensory noise (taken from previous psychophysical literature) is consistent with subjects' performance and to derive a model of the prior distribution on figure shape that underlay subjects' judgments. The results show that subjects did appear to down-weight the information provided by figure shape as conflicts with binocular disparity grew in magnitude but that they did not completely veto the shape cue. Rather, their behavior was well fit by a Bayesian model that assumes two categories of elliptical figures—circles and ellipses with random aspect ratios, whose probability density peaks at 1 (i.e., still shows a preference for circles).

Preview

We measured subjects' judgments of surface slant for stereoscopic images of ellipses, as depicted in Figure 4. Subjects are strongly biased to see slanted ellipses as circles. The cue to surface orientation provided by the orientation and aspect ratios of figures under an assumption of circularity (for ellipses) or isotropy (for arbitrary figures) is typically referred to as compression. The advantage of using ellipses as stimuli is that, under perspective projection, slanted ellipses project to ellipses in the retinal image, retaining the fundamental ambiguity in the percept. This means that the only monocular cue to 3D orientation provided by these stimulus images is the aspect ratio and orientation of an ellipse in the image. Furthermore, the information about surface orientation provided by ellipses is easily characterized by a prior on aspect ratio and a model of sensory noise on the measured aspect ratios and orientations of ellipses in the image. The ellipses were filled with random dots to provide a rich source of stereoscopic disparities while limiting texture information. Texture density contributes minimally to slant perception (Braunstein & Payne, 1967; Buckley, Frisby, & Blake, 1996; Cumming, Johnston, & Parker, 1993; Cutting & Millard, 1984; Knill, 1998a, 1998b), and we randomized the sizes and shapes of the texels to reduce the salience of local texture shape cues. Nevertheless, the textures potentially provided some information about slant. Because this was always consistent with the stereoscopic cues, further references to stereoscopic cues or stereoscopic slant, strictly speaking, refer to combined stereoscopic and texture cues.

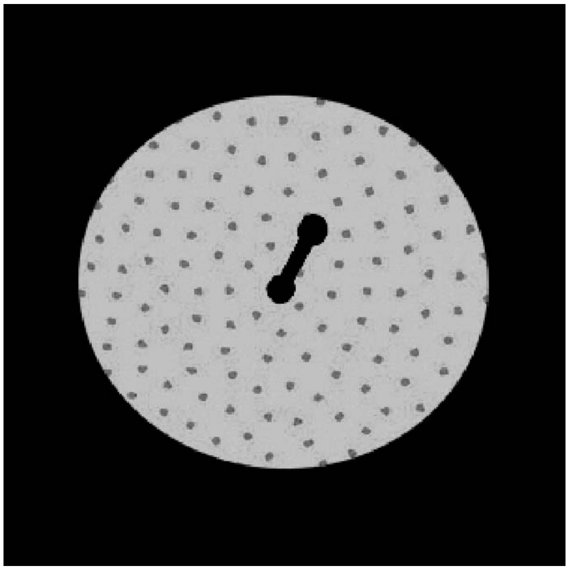

Figure 4. The stimulus used in the experiments (shown here in gray-scale for purposes of printing and reproduction). Subjects adjusted the 3D orientation of the line probe to appear perpendicular to the surface.

The experiments measured subjects' estimates of surface slant for binocularly presented ellipses with a range of aspect ratios. Images of ellipses with different aspect ratios have different degrees of conflict between the orientations suggested by the compression cue and stereoscopic disparities. All ellipses were oriented horizontally and slanted around the horizontal axis to simplify the analysis and limit the number of experimental conditions. Thus, conflicts were limited to the slant of the figure (angle away from fronto-parallel). Subjects were asked to judge the 3D orientations of the figures by adjusting a stereoscopically presented line probe to appear perpendicular to the figures, as in Figure 4. Experiment 1 measured cue weights when stereoscopic disparities specified a slant of 35°. Experiment 2 measured cue weights when stereoscopic disparities specified a slant of 55°.

We formulated the optimal Bayesian estimator for estimating slant from stereoscopic images of ellipses using a prior distribution of aspect ratios that contained a mixture of (1) a delta function at 1 (circles) and (2) a broader distribution of aspect ratios. Data from previous experiments allowed us to estimate the average of subjects' sensory uncertainty in estimating slant-from-stereo disparities. The principal free parameters in the model, therefore, were the uncertainty in sensory estimates of the aspect ratios of ellipses in the retinal image and parameters describing the prior distribution of aspect ratios. We fit this model to the data from the experiments to test whether it accurately characterized subjects' nonlinear behavior in combining the figural and binocular information in the stimulus images. Bayes' optimality predicts that the same prior model will accurately fit data from both of the slant conditions used in Experiments 1 and 2.

Experiments

Figure 4 shows an example of the stimuli used in the experiments. Both the surfaces and the line probes were presented stereoscopically. On each trial, the orientation of the probe was randomized in an annular region on the view sphere around the true (stereoscopic) orientation of the stimulus surface. Subjects used the computer mouse to adjust the 3D orientation of the line probe to appear perpendicular to the surface in the stimulus. Test stimuli consisted of stereoscopic views of an elliptical figure filled with randomly positioned dots at a fixed slant (35° for Experiment 1 and 55° for Experiment 2). Surface tilt was fixed at vertical in all stimuli. Test stimuli were given random aspect ratios by compressing or stretching a circle in the vertical direction (in the plane of the surface) so as to keep the area of the figure constant. Different subjects were used in the two experiments to keep the experiments short (two 1-hr sessions each) and to minimize potential effects of learning.

Data analysis was performed on subjects' slant settings as measured by the matching probe orientations. Because both subjects' slant estimates and their estimates of probe orientation were likely to be biased, we randomly intermixed a large number of baseline trials containing stereoscopic images of circles at slants ranging from 15° to 65°. Data from these conditions allowed us to map subjects' probe settings on test trials to equivalent slants of cue-consistent stimuli. The adjusted slants were used for analysis (e.g., to compute cue weights).

Methods

Visual stimuli

Visual displays were presented on a computer monitor viewed through a mirror (see Figure 5) using CrystalEyes shutter glasses to present different stereo views to the left and right eyes. Displays had a resolution of 1,280 × 1,024 pixels and a refresh rate of 118 Hz (59 Hz for each eye's view). Stimuli were drawn in red to take advantage of the comparatively faster red phosphor of the monitor and prevent interocular cross talk. Viewing distance to the monitor (or its virtual image behind the mirror) was approximately 50 cm, although it varied slightly from subject to subject. The viewing angle to the monitor was approximately 38°, although it, too, varied slightly from subject to subject.

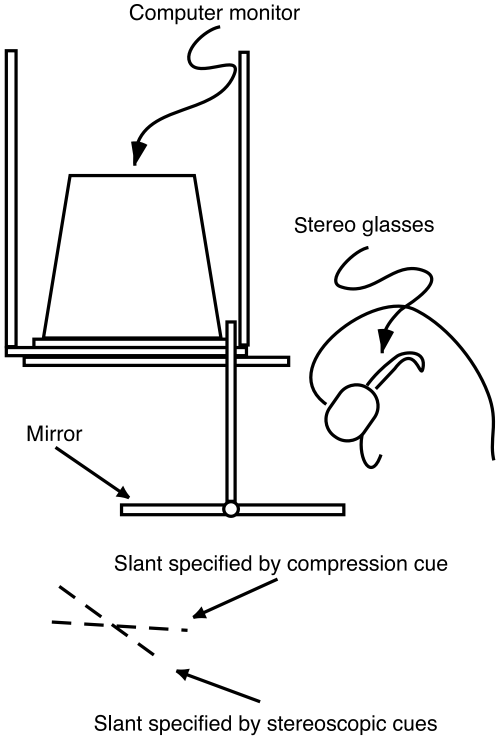
Computer monitor
Stereo glasses
Mirror
Slant specified by compression cue
Slant specified by stereoscopic cues

Figure 5. The viewing arrangement used in the experiments. Stimuli appeared as slanted ellipses floating in space behind the mirror.

A rectangular black occluder was placed on the mirror to obscure the frame of the monitor from subjects' view. Stimuli were centered on the center of the virtual image of the CRT in 3D space. Stimuli consisted of planar, elliptical disks filled with random dot textures. Disks were created from circles with a radius of 6 cm. Figures in the baseline stimuli, consisting of circles projected at slants ranging from 15° to 65° viewed from 50 cm, subtended approximately 13.8° horizontally and 5.8–13.4° vertically from the point of view of a subject. Because figures were rotated around the horizontal axis, the horizontal extent of the figures in the stimulus images changed only slightly as a function of surface slant. Test stimuli containing ellipses with aspect ratios different from 1 were created by compressing or stretching circles in a direction perpendicular to the horizontal axis. Figures were then scaled to maintain their area in the plane of the surface. Thus, for example, an ellipse with an aspect ratio of 0.8 had a horizontal radius of 7.5 cm in the virtual world. Random dot textures were created from constrained random lattices of points in the plane. The random lattices were samples from a stochastic reaction–diffusion process that effectively perturbed the positions of the points in the lattice away from a rectangular grid. The resulting lattices represented a trade-off between a completely random selection of points in the plan and a regular lattice that would have created linear perspective cues. The points in the random lattice were used to create a Voronoi pattern—a collection of polygons centered on the points in the lattice that tiled the plane. The randomly shaped polygons generated in this way were then shrunk to a width of 0.22 cm (~15 arcmin), on average, to create a set of randomly shaped ''dots''. By generating dots in the texture pattern in this way, we weakened local figure shape cues that would be provided were the dots drawn as circles. The textures were not compressed with the ellipses for the noncircular ellipse stimuli; thus, texture cues, to the extent that subjects could use them, were always consistent with the stereoscopically defined slant. Twenty different random textures were used in the experiment.

A line probe was rendered with its base at the center of the stimulus surfaces. The line itself was rendered as a cylinder with a radius of 0.25 cm. It had balls attached to the tops and bottoms to eliminate monocular cues to line orientation that would have been provided by the projections of the circular cross sections of the cylinder.

Apparatus

Figure 5 shows a schematic diagram of the viewing apparatus used in the experiment. Subjects placed their heads in a chin rest, resting against a headrest. Subjects adjusted the orientation of the line probe using a mouse placed on the table positioned under the mirror. Spatial calibration of the virtual environment required computing the positions of subjects' two eyes relative to the virtual image of the screen. These parameters were determined at the start of each experimental session using an optical matching procedure. The backing of the half-silvered mirror was temporarily removed so that subjects could see their hand and the monitor simultaneously. A test grid containing thin rods with varying heights and positions was placed on a tabletop aligned with the monitor under the mirror. Subjects aligned a crosshair on the display with the tips of rods on the test grid. A total of 23 positions subtending approximately 35° × 7° of visual angle were matched. Matches were performed monocularly in separate sequences for left and right eyes. The combined responses for both eyes were used to determine a globally optimal combination of 3D reference frame and eye position. The cyclopean reference frame used to create stimuli had its origin at the point halfway between the two eyes and had a horizontal axis defined by the vector difference between the two eyes' positions.

Procedure

Sixteen stimulus conditions were used in the experiment. Six of these were baseline conditions containing circular stimuli presented at slants of 15°, 25°, 35°, 45°, 55°, and 65°. The other 10 ''test'' conditions were ellipses presented at a slant of 35° (Experiment 1) or 55° (Experiment 2). In Experiment 1, the aspect ratios of the ellipses in the test conditions were 0.6, 0.7, 0.8, 0.85, 0.9, 0.95, 1.05, 1.1, 1.15, and 1.2. This created cue conflicts between the compression cue and stereoscopic disparities of 25.6°, 20°, 14.1°, 10.9°, 7.5°, 3.9°, −4.3°, −9.3°, −15.4°, and −24.4°. In Experiment 2, the aspect ratios of ellipses in test stimuli were 0.3, 0.5, 0.6, 0.7, 0.8, 0.9, 1.1, 1.2, 1.35, and 1.5. This gave a similar range of cue conflicts as in Experiment 1—25.1°, 18.3°, 14.9°, 11.3°, 7.7°, 3.9°, −4.1°, −8.5°, −13.2°, and −24.4°. Subjects performed two sessions on different days, each containing four blocks of trials. Each block contained 14 each of the baseline conditions and 4 each of the test conditions, giving a total of 124 trials per block. Subjects took, on average, 7 minutes to complete a block; hence, each session took approximately 45 minutes to run, including the time for calibration and breaks between blocks. Stimuli were presented with an intertrial interval of 500 ms and remained on the display until subjects pressed the mouse button to indicate a match.

Subjects

Subjects were 16 undergraduates at the University of Rochester who were naive to the goals of the experiment. There were eight subjects in Experiment 1 and eight in Experiment 2. Subjects had normal or corrected-to-normal vision and normal stereo vision.

Results

Experiment 1

Figure 6A shows slant settings for three representative subjects in the baseline conditions. Subjects' settings in the test conditions were all well within the range of their baseline settings for stimuli between 25° and 45°; therefore, we used subjects' slant settings on those baseline trials to remove biases from subjects' slant settings in the test trials. Because of significant nonlinearities in some subjects' slant judgments (see, e.g., the red curve in Figure 6A), we performed a least squares, quadratic regression to fit subjects' probe slant settings as a function of the true stimulus slant,

$$s_{\text{probe}} = as^2_{\text{stimulus}} + bs_{\text{stimulus}} + c + \text{Noise}, \tag{5}$$

where s_{probe} represents subjects' probe slant settings and s_{stimulus} represents the stimulus slant. We then computed corrected (unbiased) slant settings, s_{probe}, on test trials by inverting this equation

$$s = \frac{-b + \sqrt{b^2 - 4a(c - s_{\text{probe}})}}{2a} \tag{6}$$

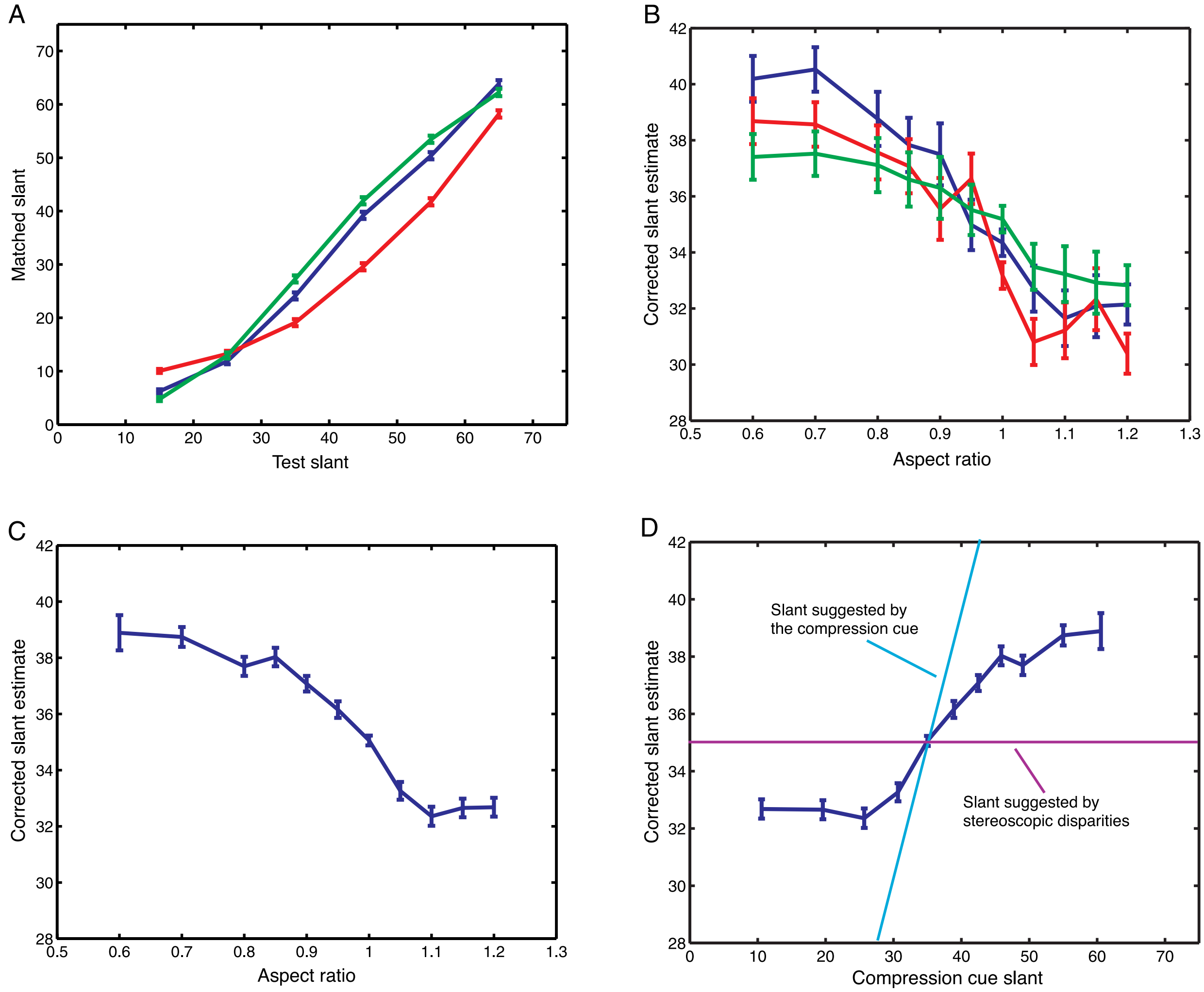

Figure 6. (A) Slant settings for three subjects on baseline trials in Experiment 1. (B) Corrected slant settings on the test stimuli for the same three subjects as a function of the aspect ratio of the ellipse. Corrected slant settings were computed by inverting the fitted quadratic function that mapped stimulus slant and subjects' settings on baseline trials at slants of 25°, 35°, and 45°. (C) Average corrected slants across all eight subjects, as a function of ellipse aspect ratio. (D) The same data replotted as a function of the slant suggested by the compression cue (the circle interpretation of the test stimuli) at each of the 10 test aspect ratios. Error bars for individual subjects (A and B) are standard errors of the mean settings for those subjects. For the grouped data (C and D), they are standard errors of the means computed across subjects.

The corrected slant settings, s, are estimates of the slants of cue-consistent stimuli (stereoscopic images of circles) that would appear to have the same slants as the test stimuli. Figure 6B shows the same three subjects' corrected slant settings on test trials, as a function of the aspect ratio of the ellipse projected into the stimulus. Figure 6C shows average, corrected slant settings across all eight subjects.

As illustrated in the figure, subjects showed an initial bias to interpret the figure as a circle, but this bias weakened at aspect ratios very different from 1. Figure 6D replots subjects' average slant settings on test trials as a function of the conflict between the slant suggested by the compression cue (the circle interpretation of a stimulus) and the stereoscopic cue (fixed at 35° for these trials). For aspect ratios close to 1, subjects behaved as if linearly combining the slant suggested by stereopsis and the slant suggested by the compression cue but appeared to down-weight the compression cue at large conflicts. A linear cue integration model approximates subjects' slant settings for a given stimulus condition as a weighted linear sum of the slants suggested by the compression cue and by stereoscopic disparities,

$$\begin{aligned} s &= w_{\text{compression}}\, s_{\text{compression}} + w_{\text{stereo}}\, s_{\text{stereo}} \\ &= w_{\text{compression}}\, s_{\text{compression}} + \left(1 - w_{\text{compression}}\right)35, \qquad (7) \end{aligned}$$

where $s_{\text{compression}}$ represents the slant suggested by the compression cue or, equivalently, the slant consistent with a circle interpretation of the projected ellipse. Rearranging terms, we arrive at an expression for the weight that subjects effectively gave to the compression cue in each stimulus condition

$$w_{\text{compression}} = \frac{s-35}{s_{\text{compression}}-35}, \tag{8}$$

For the slant and field of view used in the experiment, the slant suggested by the compression cue is given very accurately by the cosine law approximation of perspective foreshortening,

$$s_{\text{compression}} = \cos^{-1}[\alpha \cos(35^{\circ})]. \tag{9}$$

Equation 8 provides an empirical measure of the influence of the circle bias on subjects' judgments. Figure 7 shows subjects' average corrected slant estimates replotted as compression cue ''weights''. To test the statistical significance of the effect of ellipse aspect ratio (or equivalently, the cue conflict) on the effective weights that subjects gave to the compression cue, we performed a two-way ANOVA with subjects as a factor. The results showed that the effect of aspect ratio was significant, F (9, 7) = 6.73; $p < .0001$.

Experiment 2

We analyzed the results of Experiment 2 in the same way, but we used probe slant settings at baseline slants of 45°, 55°, and 65° to derive the quadratric correction for the slant settings. Results are shown in Figures 8 and 9. Again, the effect of aspect ratio on the weights that subjects gave to the compression cue was significant, F (9, 7) = 2.89; $p < .006$.

Discussion of results

Several features are notable in the results shown in Figures 7 and 9. First, the influence of the circle bias peaked for images of ellipses that were nearly circular but dropped off as ellipses became more compressed or elongated. Second, the influence of the circle bias was asymmetric as a function of the cue conflict. At negative conflicts (when the circle bias suggested a lower slant than stereopsis), the influence of the circle bias decreased monotonically with the magnitude of the conflict. At positive conflicts, the influence of the circle bias leveled off at a near-constant value. This would not have been expected from a pure form of cue vetoing. Subjects did not ''turn off'' the compression cue but appeared to downweight it. Averaging slant settings across trials and subjects as we did would have blurred out sharp transitions that might be indicative of cue vetoing. Thus, cue vetoing is not inconsistent with a gradual reduction in average cue weights. Cue vetoing, however, does predict that cue weights would have gone monotonically toward zero as the size of the cue conflict increased. Subjects' compression cue weights clearly asymptoted at large positive cue conflicts, however, suggesting that they effectively downweighted the compression cue to a smaller nonzero value.

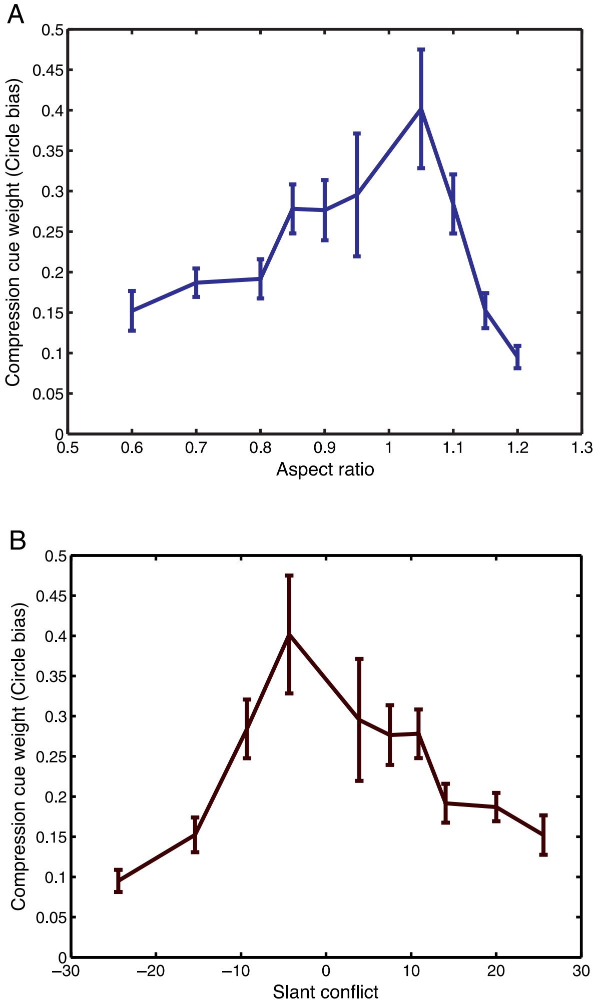

Figure 7. (A) The apparent weights that subjects give to a circle interpretation of test stimuli with aspect ratios different from 1, computed using Equation 8. (B) The same data replotted as a function of the conflict between the slant suggested by a circle interpretation of the elliptical stimuli and the slant suggested by stereoscopic disparities. Error bars are standard errors of the mean weights computed across subjects.

Before inferring from the observed changes in cue ''weights'' that subjects used a Bayesian strategy for robust cue integration based on mixtures of priors, we must consider several simpler accounts for the results. First is the possibility that subjects were biased to use stereoscopic information because the probe that they used to match the slant of the figure was presented stereoscopically. This might affect the weights that subjects gave the cues for small conflicts and might bias subjects to veto the monocular cues at large conflicts. This account is strongly argued against by previous results using the same task and almost the same stimuli (stimuli only differed in the

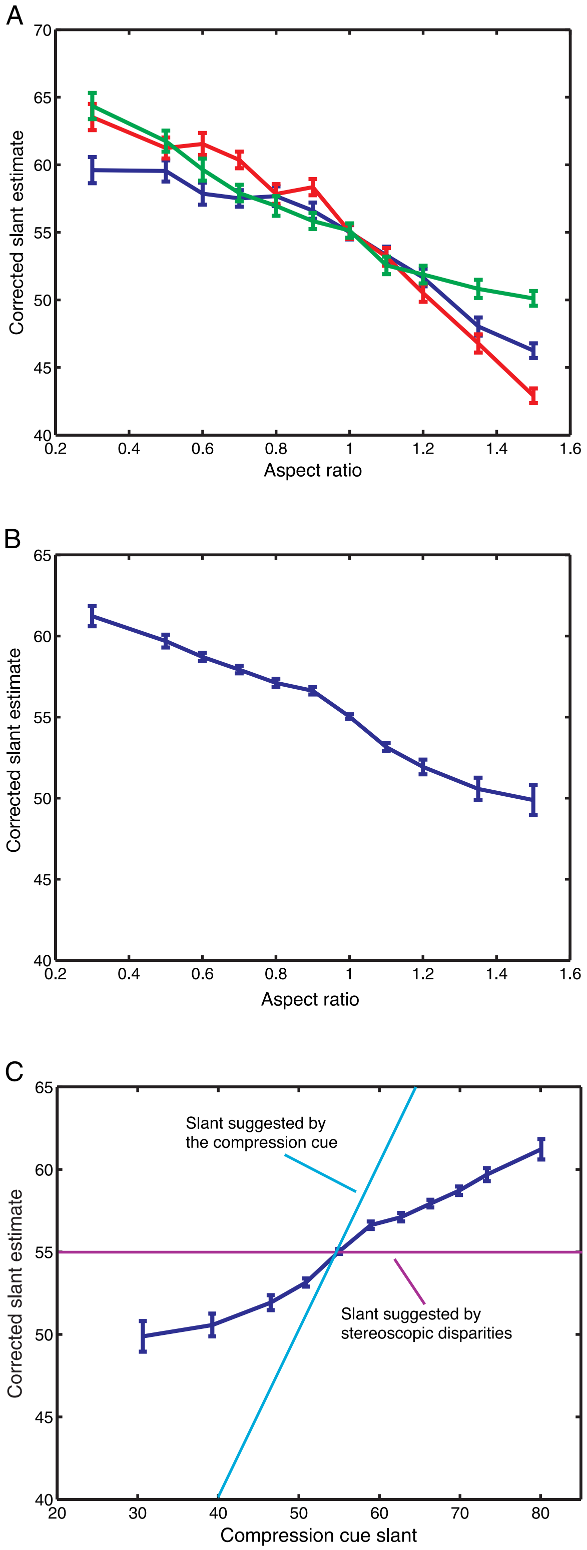

Figure 8. (A) Corrected slant settings for three subjects on test trials in Experiment 2, plotted as a function of an ellipse's aspect ratio. (B) Average corrected slants across all eight subjects, as a function of ellipse aspect ratio. (D) The same data replotted as a function of the slant suggested by the circle interpretation of test stimuli at each of the 10 test aspect ratios. Error bars are standard errors of the means.

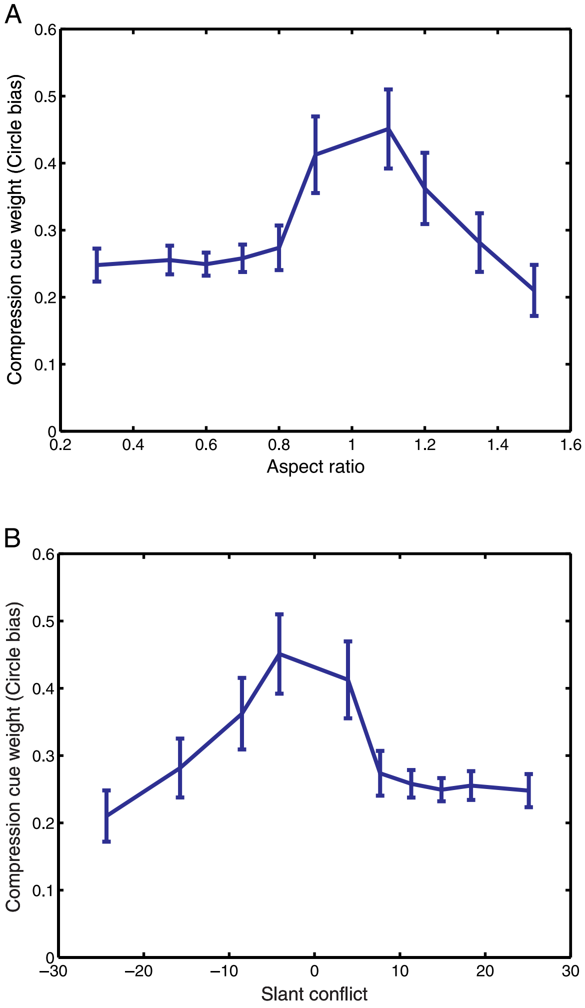

Figure 9. (A) The apparent weights that subjects give to a circle interpretation of test stimuli with aspect ratios different from 1 in Experiment 2. (B) The same data replotted as a function of the conflict between the slant suggested by a circle interpretation of the elliptical stimuli and the slant suggested by stereoscopic disparities. Error bars are standard errors of the means of the subjects' weights.

texture used to fill the ellipses), in which we found that the weights that subjects gave to monocular and stereoscopic cues were the same when the stereoscopic probe and a haptic matching task were used, in which subjects oriented an unseen cylinder to ‘‘feel’’ like it was at the same orientation as the stimulus. Moreover, the stereoscopic weights measured using the stereoscopic probe were actually lower than those measured using a motor task in which subjects placed a cylinder onto the stimulus surface (Knill, 2005).

Second, we must consider the already well-supported model that changes in cue weights resulted from changes in cue reliability across stimulus conditions. According to this account, the apparent changes in weight could have resulted from changes in the uncertainty attached to sensory estimates of the aspect ratios of ellipses in the image. This assumes that because the true stimulus slant was fixed within an experiment, the uncertainty in stereoscopic slant estimates remained approximately constant as a function of ellipse aspect ratio. Comparing Figures 7 and 9 provides a quick negative answer to the simple hypothesis. In Experiment 2 (Figure 9A), because ellipses at higher slants were foreshortened more by perspective, the projected aspect ratio of the ellipse whose real aspect ratio was 0.9 was essentially equal to the projected aspect ratio of the ellipse in Experiment 1 whose real aspect ratio was 0.6 (0.52 vs. 0.49). Thus, the uncertainty associated with the figure shape information should have been essentially equivalent in these two conditions. However, subjects effectively gave more weight to the compression cue in this condition in Experiment 2 than in Experiment 1—0.41 (±0.057 *SE*) versus 0.15 (±0.025 *SE*), despite the fact that slant-from-disparity is more reliable at 55° (in Experiment 2) than at 35° (in Experiment 1; Hillis et al., 2004; Knill & Saunders, 2003). The performance of the Bayesian model that assumes a simple prior on aspect ratios provides further insight into this issue. Figure 3 shows two such observers—one that assumes all ellipses are circles and one that assumes a log-Gaussian distribution of aspect ratios that is peaked at 1 (circles) and has a standard deviation of 0.25. Both observers show a monotonic increase in the apparent weight given to the compression cue as difference between the slant suggested by the compression cue and the slant suggested by stereopsis increases from large negative values to large positive values, that is, as the slant suggested by the compression cue increases from near 0° to near 60°. This behavior reflects the change in the uncertainty induced by measurement noise as a function of the aspect ratio of the retinal ellipse, itself caused by the cosine law of projective foreshortening.

Finally, we should consider what effects cues like accommodation and blur might have had on subjects' performance. In theory, were these cues to suggest a slant very different from the stereoscopic cues, they could have had a significant nonlinear impact on subjects' performance because their weights would depend on the uncertainty associated with the combined stereoscopic/figure shape cues. This would only have had a significant impact if the cues were strong. That they were not very strong is argued for by the fact that the gain in subjects' slant settings as a function of stimulus slant for cue-consistent stimuli was almost exactly 1, on average (see Figure 6A). Because cues like accommodation suggested a fixed slant (approximately 38°), any impact they would have had would have been to shrink the gain of that function. While it may be that the impact of the cues was counterbalanced by subjective biases in subjects' mapping between stimulus and probe slant, it seems unlikely that this effect was large enough to significantly impact performance. Moreover, in Experiment 1, at least, the slant suggested by uncontrolled for cues was almost the same as that of the stereoscopic cues. In this situation, the presence of the cues would not have changed the predictions of the model in any qualitative way—their contributions would only have shrunk somewhat the apparent weight that subjects gave to the compression cue.

Modeling robust cue integration: Bayesian model selection

A Bayesian model that uses a mixture of prior models on the shapes of figures in the world would seem to account for the qualitative pattern of subjects' slant judgments. To test such an account more quantitatively, we fit a Bayesian model to subjects' data. The Bayesian model has four free parameters. Two parameters describe subjects' prior distributions on aspect ratios of ellipses in the world—the proportion of circles in the world and the spread of the distribution of aspect ratios for ellipses that are not circles. The other two characterize the uncertainty in sensory measurements of the aspect ratios of ellipses in the retinal image and of slant-from-disparity. Preliminary simulations showed that for the 55° stimuli in Experiment 2, the Bayesian model was slightly more skewed toward negative slant conflicts than subjects' data would suggest but otherwise could fit the shape of the weight function well. We therefore added a fifth free parameter to the model that represented possible biases in subjects' estimates of slant-from-stereo. This could arise from known biases in depth estimates from stereopsis or from small biases in the calibration procedure (e.g., of the interocular distance).

The structure of the model

We modeled subjects' prior beliefs about the aspect ratios of ellipses in the world, α, as a mixture of a delta function at $\alpha = 1$ (all ellipses in the category are circles) and a log-Gaussian distribution over aspect ratios. The log-Gaussian distribution ensures that the probability density function for $1/\alpha$ is equal to the density function for α (necessary to make the prior invariant to rotations). The prior density function was given by

$$\begin{aligned} p(\alpha) &= \pi_{\text{circle}}\, p_{\text{circle}}(\alpha) + \pi_{\text{ellipse}}\, p_{\text{ellipse}}(\alpha) \\ &= \pi_{\text{circle}}\, \delta(\alpha - 1) \\ &\quad + \pi_{\text{ellipse}} \left\{ \frac{1}{\alpha} \frac{1}{\sqrt{2\pi}\sigma_\alpha} \exp\left[-(\log \alpha)^2 / 2\sigma_\alpha^2 \right] \right\}, \end{aligned} \tag{10}$$

where we have labeled the categories of figures as circles or ellipses. The prior model has two free parameters—the relative probability of figures being circles, π_{circle} ($\pi_{\text{ellipse}} = 1 - \pi_{\text{circle}}$), and the standard deviation of the log-Gaussian distribution of aspect ratios in the ellipse category, σ_α.

We modeled the aspect ratio of the ellipse in a stimulus image, A, as the aspect ratio of the projected ellipse corrupted by random Gaussian noise,

$$A \approx \alpha \cos(S) + \Omega_A, \tag{11}$$

where Ω_A is a random noise variable that is normally distributed with mean 0 and standard deviation, σ_A, and S is the slant of the surface. For the field of view used in the experiment, the cosine foreshortening law is within 1% of the true perspective foreshortening. Finally, we modeled the disparity cue as an estimate of slant corrupted by Gaussian noise

$$S_{\text{stereo}} = S + \Omega_{\text{stereo}}, \tag{12}$$

where Ω_{stereo} is a random noise variable that is normally distributed with a standard deviation, σ_{stereo}. We left the mean of the noise process as a free parameter to model biases in perceived slant-from-stereo.

The likelihood function for the compression cue is a weighted sum of likelihood functions derived from the circle and ellipse priors on shapes in the world,

$$\begin{aligned} p(A|S) &= \pi_{\text{circle}}\, p_{\text{circle}}(A|S) + \pi_{\text{ellipse}}\, p_{\text{ellipse}}(A|S) \\ &= \pi_{\text{circle}} \int_0^\infty p(A|S,\alpha)\, p_{\text{circle}}(\alpha)\mathrm{d}\alpha \\ &\quad + \pi_{\text{ellipse}} \int_0^\infty p(A|S,\alpha)\, p_{\text{ellipse}}(\alpha)\mathrm{d}\alpha \\ &= \frac{1}{\sqrt{2\pi}\sigma_A} \left\{ \begin{array}{l} \pi_{\text{circle}} \exp\left[-(A-\cos(S))^2/2\sigma_A^2\right] + \\ \pi_{\text{ellipse}} \int_0^\infty \exp\left[-(A-\alpha\cos(S))^2/2\sigma_A^2\right] p_{\text{ellipse}}(\alpha)\mathrm{d}\alpha \end{array} \right\}, \end{aligned} \tag{13}$$

where $p_{\text{ellipse}}(\alpha)$ is taken from Equation 10. The first term in the mixture is an integral of the product of a Gaussian density function with a delta function on α. This results in a Gaussian function with α replaced by the value that sets the argument of the delta function equal to 0 (in this case, $\alpha = 1$). This is equivalent to the likelihood one would obtain by simply setting $\alpha = 1$, rather than integrating over a delta function prior on α. The second term in the mixture is an integral over the possible aspect ratios in the ellipse model. This has the effect of shrinking the magnitude of that component of the likelihood function. The likelihood function shown in Figure 2 was computed using this model.

The likelihood function for stereoscopic disparities is simply a Gaussian with standard deviation, σ_{stereo}, and mean equal to the true slant plus the bias term. The posterior distribution of slant, given the measured aspect ratio of an ellipse in the image and the measured stereoscopic disparities, is given by the product of this Gaussian with the likelihood function for compression and the prior on slant, which, assuming a generic viewpoint on a surface, is given by sin(S). In the simulations described below, we used a minimum mean square error estimator. The estimator selected as its best estimate of the slant on a given trial the expected value of slant computed from the normalized joint likelihood function. The results were essentially the same when we used a maximum a posteriori (MAP) estimator that selected the mode of the posterior distribution on each trial.

Fitting the model to the data

The model has five free parameters—two determine the sensory uncertainty associated with each cue, two describe the prior distribution on aspect ratios assumed to characterize the world, and the fifth is a bias term that accommodates relative biases in observers' estimates of slant-from-stereopsis. The data from the experiments do not support fitting absolute measures of sensory noise (which determine variability in subjects judgments), in part because subjects' variability is confounded with uncertainty in the probe slant settings and in part because their variability is increased by attentional lapses, decision noise, and so forth. We dealt with this difficulty by fixing the standard deviation of the slant-from-disparity noise based on data from other, more sensitive experiments (see the Appendix for details). The resulting values for σ_{stereo} were 3.5° and 2.59° for the 35° and 55° slant conditions, respectively. The difference reflects the fact that disparity cues to slant improve slightly as a function of increasing slant (Hillis et al., 2004).

Because we have access only to an estimate of the average variance in subjects' slant-from-stereo estimates, we fit the Bayesian model to the compression cue weights averaged across subjects, as plotted in Figures 7 and 9. The Appendix gives details of the model fitting procedure. Table 1 lists the model parameters for the best fits to the data from Experiment 1 and 2, respectively, and Figure 10 plots the predicted compression cue weights for the best fitting models along with subjects' data. Although not exactly equivalent, the parameters characterizing the best-fitting prior distributions were similar for the two experimental conditions. Given the approximate methods used to set the noise parameters for slant-from-stereo in the two experimental conditions, we cannot expect exact equivalence in our model fits across the two conditions. The estimate of the sensory noise in subjects' visual estimates of aspect ratio is in fairly close accord with published data on aspect ratio discrimination, which range from 0.02 to 0.04 (Regan & Hamstra, 1992). We should note, however, that published data for aspect ratio discrimination are for ellipses that subtend only 1° of visual angle.

General discussion

The impact of the subjects' biases to interpret an elliptical figure as a circle shrank as the conflict between

Slant	Proportion of ellipses (π_{ellipse})	Standard deviation of aspect ratios in the ellipse model (σ_{α})	Standard deviation of aspect ratio measurements (σ_{A})	Bias in visual estimates of slant-from-stereo
35°	0.124 (±0.037)	0.127 (±0.0085)	0.024 (±0.0086)	−0.489 (±0.4511)
55°	0.039 (±0.025)	0.104 (±0.011)	0.036 (±0.0048)	−3.49 (±1.07)

Table 1. Best fitting model parameters for Experiment 1 and 2. Standard errors in parentheses were derived from the Hessian of the log-likelihood function for the model fits.

this interpretation and stereoscopic cues to slant grew, but it never disappeared. In the standard terminology of cue integration, subjects appeared to down-weight but not to veto the compression cue to slant at large conflicts with stereoscopic cues. Subjects' behavior was consistent with a Bayesian model of robust cue integration that accommodates the possibility of interpreting pictorial cues according to either a strong prior (circles) or a weak prior (ellipses with constrained aspect ratios). The model does not explicitly change cue weights—it computes an estimate based on the likelihood function computed by multiplying a likelihood function for slant-from-disparity with a mixed likelihood function for slant-from-figure shape. It gradually shifts from being dominated by the component of the slant-from-figure shape likelihood function derived from the circle prior to the component derived from the random ellipse prior.

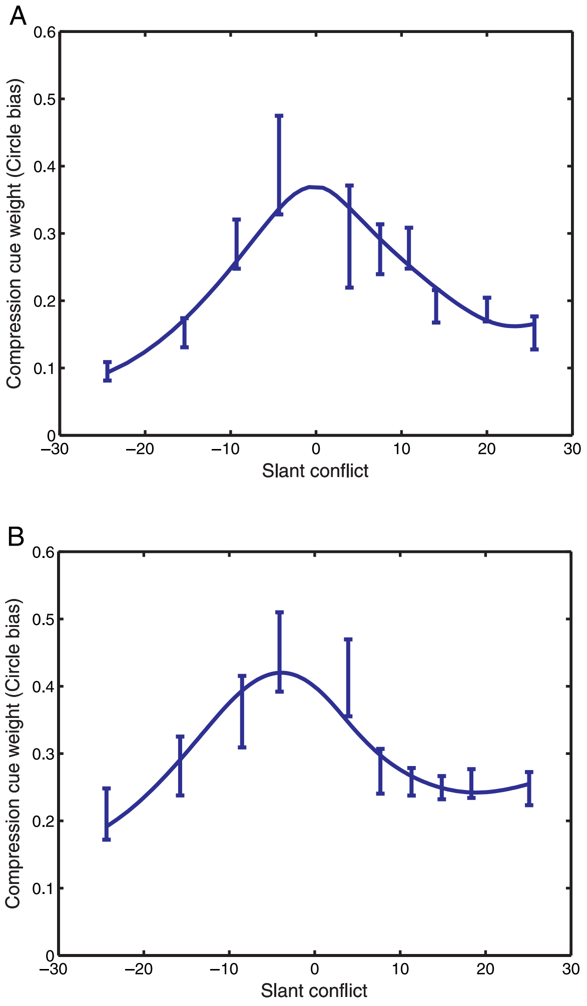

Figure 10. Expected compression cue weights for the best fitting models for (A) Experiment 1 and (B) Experiment 2 (solid curves) plotted along with subjects' average compression cue weights in the two experiments. The expected model weights were computed using the same analysis applied to subjects' data—first removing estimator bias (due to the slant-from-stereopsis bias term) from slant estimates using a quadratic fit and then applying Equation 8 to compute compression cue weights. Note the shift in the peak for the model weights for Experiment 2. Error bars are the standard errors of the mean weights computed across subjects.

Comparisons with other approaches to robust cue integration

Landy et al. (1995) were the first to formally connect the problem of integrating depth cues to statistical approaches to robust estimation. They did not propose a specific model for robustly integrating discrepant depth cues, but they did suggest that methods from robust statistics should be applied. By and large, these methods apply to problems somewhat different in kind from the cue integration problem. Specifically, they deal with problems in which many data points are available to estimate some parameter. The methods are nonlinear techniques that either determine outlier points to discard from analysis or adjust how points are weighted to estimate a parameter (Wonnacott & Wonnacott, 1990). On the face of it, these methods appear like they would be applicable to cue integration; however, they rely on having a large number of data points available. Consider the example of the trimmed mean technique for estimating a population mean (Wonnacott & Wonnacott, 1990). This method excludes some percentage (e.g., 5%) of the points in the two tails of the data histogram before calculating a standard sample average to estimate the mean. By analogy, this would suggest a strategy whereby the visual system vetoes a cue when it disagrees by a large amount from another set of cues.

The problem with the analogy is that the trimmed mean technique is only effective when a sizable number of samples are available. In vision, the number of available cues is relatively small (perhaps two to six). The methods do not easily apply to a situation like the one described here in which only two cues are available; hence, we cannot directly translate standard methods for robust statistical estimation to the problem of visual cue integration. Because of this, there are no extant models for robust visual cue integration that are well specified enough to make quantifiably testable predictions of human performance.

The alternative model that would seem to compete with the Bayesian account is that the visual system vetoes or down-weights one of a set of grossly discrepant cues. To be tested experimentally, however, this proposal requires a specification of a principled method for determining which cue to veto or down-weight—and, in the latter case, how to down-weight it. In the next two sections, we will describe how both of these models can arise naturally as *implementations* of an optimal Bayesian observer who uses mixed priors to interpret a cue. The hard computational problem in such systems is determining how to reweight cues or which cue to veto. The Bayesian model, in this context, can be seen as characterizing the optimal way to perform either of these functions. The cost function that an estimator is designed to minimize will determine whether the estimator behaves as a linear integrator with graded reweighting of cues or as a system that vetoes one or another cue as a function of cue uncertainty and the size of cue conflict.

Comparing Bayesian cue integration with reweighting

When the likelihood functions for different cues and the prior density function are Gaussians or mixtures of Gaussians, a robust Bayesian estimator that uses the mean of the posterior density function as its estimate of a surface property can be expressed as a linear combination of individual estimates of that property. Rather than having a single estimate for each cue, however, the system linearly combines estimates derived using each of the prior models that can be used to interpret that cue. Figure 11 illustrates a linear system that effectively implements the optimal Bayesian estimator for the stereopsis/figure shape integration problem under the approximation that the likelihood function for slant-from-figure shape is a mixture of Gaussians. It is a cascade of linear processes. In the first stage, the figure shape cue is interpreted by two different estimators, each of which relies on a different prior model for ellipses in the world—that all ellipses are circles or that ellipses are drawn from a random ensemble of ellipses. These estimates are linearly combined with the estimate of slant-from-stereopsis and the slant suggested by an observer's prior model. The weights in this stage of cue combination are specified in the usual way for linear combination; that is, they are in inverse proportion to the

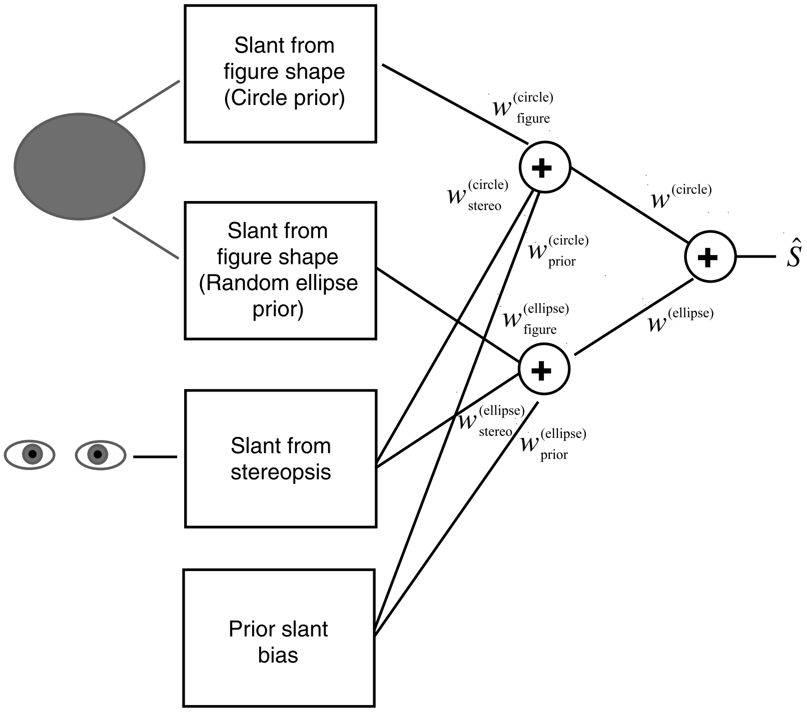
Slant from figure shape (Circle prior)
Slant from figure shape (Random ellipse prior)
Slant from stereopsis
Prior slant bias
$w^{(\text{circle})}_{\text{figure}}$
$w^{(\text{circle})}_{\text{stereo}}$
$w^{(\text{circle})}_{\text{prior}}$
$w^{(\text{ellipse})}_{\text{figure}}$
$w^{(\text{ellipse})}_{\text{stereo}}$
$w^{(\text{ellipse})}_{\text{prior}}$
$w^{(\text{circle})}$
$w^{(\text{ellipse})}$
+
+
+
$\hat{S}$

Figure 11. When the likelihood functions associated with each of the prior models that can be used to interpret one cue are Gaussian, the likelihood function for another cue is Gaussian, and the prior density function for the parameter being estimated is Gaussian, one can model the optimal Bayesian estimator as a cascade of linear integrators. The integrators compute weighted sums of estimates derived from each of three processes—an estimator that uses one of the prior models to interpret the first cue (here, we use the figure cue as an example), an estimator that uses a different prior model to interpret the first cue, an estimator that uses the second cue (here, we use stereopsis as an example), and a prior estimate of *S* (see text for discussion of the weights in each stage).

variance of the associated likelihood functions (or of the prior density function). In the second stage, the slant estimates derived by integrating stereopsis with each of the estimates derived from the figure's shape are linearly combined to derive a final estimate of slant. The weights in the second stage are determined by several factors—the probabilities of each of the prior models for Cue 1 being true in the world, the *heights* of the joint likelihood functions computed from combining Cue 2 with Cue 1 under each prior model, and the reliabilities of the estimates derived from each estimator (the spread of the associated likelihood functions).

To see this, we expand the posterior density function for slant conditioned on figure shape and stereoscopic information as a weighted sum of the posteriors derived from the circle and random ellipse models of figures

$$p\left(S|\alpha,\vec{d}\right) = p\left(M=\text{circle}|\alpha,\vec{d}\right)p_{\text{circle}}\left(S|\alpha,\vec{d}\right) + p\left(M=\text{ellipse}|\alpha,\vec{d}\right)p_{\text{ellipse}}\left(S|\alpha,\vec{d}\right), \tag{14}$$

where α is the aspect ratio of the figure in the image and $\vec{d}$ is a vector of disparities that represents the information provided by stereopsis. The expected value (mean) of the posterior density function is a weighted average of the means of the posterior density functions derived form each mode

$$\overline{S} = E\left[S|\alpha,\vec{d}\right] = p\left(M=\text{circle}|\alpha,\vec{d}\right)\overline{S}_{\text{circle}} + p\left(M=\text{ellipse}|\alpha,\vec{d}\right)\overline{S}_{\text{ellipse}}, \tag{15}$$

where $\overline{S}_{\text{circle}}$ is the mean of the posterior density function derived from the circle model of figures and $\overline{S}_{\text{ellipse}}$ is the mean of the posterior density function derived from the random ellipse model. When the likelihood functions and the prior density are Gaussian, these means are the weighted sums of slant estimates derived from the different cues and the prior as in the standard linear model of cue integration (Landy et al., 1995).

$$\begin{aligned}\overline{S} &= E\left[S|\alpha,\vec{d}\right] \\ &= p\left(M=\text{circle}|\alpha,\vec{d}\right) \\ &\times\left(w_{\text{figure}}^{(\text{circle})}\overline{S}_{\text{circle}} + w_{\text{stereo}}^{(\text{circle})}\overline{S}_{\text{stereo}} + w_{\text{prior}}^{(\text{circle})}\overline{S}_{\text{prior}}\right) \\ &+ p\left(M=\text{ellipse}|\alpha,\vec{d}\right) \\ &\times\left(w_{\text{figure}}^{(\text{ellipse})}\overline{S}_{\text{ellipse}} + w_{\text{stereo}}^{(\text{ellipse})}\overline{S}_{\text{stereo}} + w_{\text{prior}}^{(\text{ellipse})}\overline{S}_{\text{prior}}\right),\end{aligned} \tag{16}$$

where the weights are in inverse proportion to the variances of the associated likelihoods and priors and the weights within each term sum to 1. The weights in the first term are given by

$$\begin{aligned}w_{\text{figure}}^{(\text{circle})} &= \frac{1/\sigma_{\text{circle}}^2}{1/\sigma_{\text{circle}}^2 + 1/\sigma_{\text{stereo}}^2 + 1/\sigma_{\text{prior}}^2} \\ w_{\text{stereo}}^{(\text{circle})} &= \frac{1/\sigma_{\text{stereo}}^2}{1/\sigma_{\text{circle}}^2 + 1/\sigma_{\text{stereo}}^2 + 1/\sigma_{\text{prior}}^2} \\ w_{\text{prior}}^{(\text{circle})} &= \frac{1/\sigma_{\text{prior}}^2}{1/\sigma_{\text{circle}}^2 + 1/\sigma_{\text{stereo}}^2 + 1/\sigma_{\text{prior}}^2}\end{aligned} \tag{17}$$

where σ_{circle}^2 is the variance of the likelihood function for slant-from-figure shape given that the figure in the world is a circle, σ_{stereo}^2 is the variance of the likelihood function for slant-from-stereopsis, and σ_{prior}^2 is the variance of the prior density function for slant. The weights in the second term are similarly given by

$$\begin{aligned}w_{\text{figure}}^{(\text{ellipse})} &= \frac{1/\sigma_{\text{ellipse}}^2}{1/\sigma_{\text{ellipse}}^2 + 1/\sigma_{\text{stereo}}^2 + 1/\sigma_{\text{prior}}^2} \\ w_{\text{stereo}}^{(\text{ellipse})} &= \frac{1/\sigma_{\text{stereo}}^2}{1/\sigma_{\text{ellipse}}^2 + 1/\sigma_{\text{stereo}}^2 + 1/\sigma_{\text{prior}}^2}. \\ w_{\text{prior}}^{(\text{ellipse})} &= \frac{1/\sigma_{\text{prior}}^2}{1/\sigma_{\text{ellipse}}^2 + 1/\sigma_{\text{stereo}}^2 + 1/\sigma_{\text{prior}}^2}\end{aligned} \tag{18}$$

The only difference between the weights in the two terms is the replacement of σ_{circle}^2 with $\sigma_{\text{ellipse}}^2$, the variance of the likelihood function for slant-from-figure shape given that the figure in the world is taken form a random ensemble of ellipses.

The weights in the second stage are given by the posterior probability of each model for the figures given all of the image data. These weights depend on many factors—the variance of the likelihood functions associated with the figure cue and each prior model, the variance of the stereoscopic cue, the conflict between the interpretation suggested by the stereoscopic cues and the interpretations suggested by the figure cue using each of the two prior models, and the prior probabilities of the figure being drawn from each model. These effects are summarized below.

- Size of conflict: The probability of a particular model given the image data decreases as the size of the conflict between the estimate derived using that model and the interpretations derived from other cues increases. This effect depends on the size of the conflict relative to the variances of the associated likelihood functions. A less constrained prior model is *less* affected by conflict size than a more constrained prior model because the latter leads to a lower variance likelihood function for interpreting a cue.

This is the reason why a Bayesian observer switches to a less constrained model at large cue conflicts.

- Occam's razor: The number of parameters that are free to vary in a model and thus that need to be marginalized over to calculate that model's likelihood function determines in part the magnitude of the likelihood of the mode—the greater the number of these parameters, the smaller the likelihood of the model. In this article's example, the random ellipse model has one free parameter—the figure's aspect ratio, but in general, it has two—aspect ratio and orientation. The circle model has no free parameters. This biases the Bayesian estimator toward the more constrained model, until the previous factor overcomes this effect. For other cues, for example, texture cues provided by images of lattice textures, the random texture model can have a large number of free parameters. When viewing a perspective image of a square lattice, the likelihood of the random texture model is extremely low because of this Occam's razor effect, whereas the likelihood of the square lattice interpretation is high—assuming the prior probability of viewing a square lattice is not exceedingly low.
- The prior probability of a model: This is a simple multiplicative factor that appears in the posterior probability for a model given the available image data. Higher probability models are more likely given the image data.

Because the linear formulation can implement a robust Bayesian cue integrator in some situations, one way to view the Bayesian model is that it provides a rational way to determine the weights that one should assign cues as a function of the conflict between them. Calculating the weights, however, is a nontrivial computation. Moreover, taking this perspective on robust cue integration has two dangers. First, it only truly applies when the likelihood functions are Gaussian. Second, it obscures the conceptual power of the approach, which is to reconceptualize robust cue integration as parameter estimation in a more complex world—one that has the type of categorical structure that exists in our environment.

Comparing Bayesian integration with cue vetoing

The linear formulation of the optimal Bayesian integrator derives from assuming a particular cost function for specifying the optimal estimate; in particular, a mean-squared error cost function. This results in an optimal estimator that selects the mean of the posterior density function on the scene parameter being estimated. A MAP estimator (which picks as its estimate the peak of the posterior density function), for example, cannot be exactly implemented by such a linear scheme. Both of these estimators assume that an observer implicitly considers only errors in one variable (e.g., slant) as contributing to the cost. An estimator that imposes a large cost on determining the right model to use for making its inference will behave quite differently. Such an estimator can be thought of as estimating two variables (at least) —a continuous object parameter like its slant and a discrete parameter specifying the category within which the object falls (e.g., circle vs. ellipse).

An observer who assumes a high cost for errors in the categorical judgment and who enforces a constraint that the categorical judgment and the continuous estimate be consistent with one another will operate in two steps. The observer will first determine the most likely model to use for interpreting a cue and then use that model and only that model when integrating the cue with other cues to estimate the continuous variable. Using the example of elliptical figures, such an observer will first calculate the a posteriori most probable shape category for the figure (circle or ellipse) and then use only the likelihood function derived for that model when integrating the figure shape cue with stereopsis. If the prior on figure shape in the random ellipse model is very broad, such an observer will appear to effectively veto the compression cue at large cue conflicts, because in those stimulus conditions, the random ellipse interpretation is more likely than the circle interpretation and the broad prior on the random ellipse model renders it useless as a slant cue.

A cost function that gives rise to behavior somewhere between the ''linear-with-reweighting'' scheme and pure cue vetoing is a local mass cost function, in which the cost of errors in an observer's estimates grows with the square of the error (as in the mean square error cost function) up to a point, at which it remains fixed. When applied to a simple problem like estimating the mean of a set of sample data points, this cost function gives rise to an estimator much like the trimmed mean referred to earlier. The logic behind using this cost function to derive an optimal estimator is that beyond some magnitude, all errors are equally costly. Because large errors caused by outliers are not penalized as heavily as they would be with a quadratic cost function, estimators like this are robust to outliers. In the context of cue integration, such a cost function results in an estimator that is unaffected by the tails of a likelihood function far away from its peak. Because these tails are dominated by the less likely of the different models that could be used to interpret a cue, an estimator that uses such a cost function behaves in many stimulus regimes as if it has vetoed those cues; that is, as if it has selected one model to interpret a cue and ignored the others. The behavior is not exactly equivalent to vetoing, and such an estimator will still show a graded transition between states in which one or another model is dominant, but the transition will be sharper than with the linear-with-reweighting model.

Which of these cost functions was implicitly used by observers in the experiment described here is difficult to determine. As noted earlier, trial-by-trial variation in the decision whether or not to veto the compression cue would certainly lead to what appears as a smooth transition to lower cue weights at large cue conflicts, as seen here. However, one would expect a system that employed a discrete process of cue vetoing, even were the data smoothed by trial-to-trial and subject-to-subject variance, to have compression cue weights that monotonically decreased to zero as a function of the magnitude of cue conflict. Subjects' weights, however, asymptoted at a nonzero value at large positive cue conflicts. In Bayesian terms, observers behaved at extreme conflicts as if they had switched models for interpreting the figure shape cue to a less constrained model of figure shape, but the unconstrained model is still constrained enough that observers gave some weight to the compression cue when operating in ''random ellipse'' mode.

Interpretation of the model

As the previous discussion suggests, the Bayesian model for robust cue integration and either the cue reweighting or cue vetoing models should not properly be considered competing models. Rather, the models are targeted at different levels of explanation of the problem. To use Marr's (1982) terminology, the cue reweighting and cue vetoing models would be, were they fleshed out to detail the mechanisms that determine cue weights or to decide which cue to veto, models of the algorithm used by the visual system to integrate cues. This is true even if these mechanisms do not exactly implement the Bayesian calculations needed to optimally reweight or veto cues, but rather implement some heuristics that approximate optimal performance. The Bayesian model itself is a computational model of human performance. It provides an explanation of performance in terms of the computations that are effectively implemented by the system. It does so by modeling human behavior as if they were optimal observers in a particular world. This world is defined primarily by two sets of parameters—those characterizing low-level visual noise associated with each of several cues and those describing the statistical structure of the world.

In the case of integrating the shapes of retinal ellipses with stereoscopic information about surface slant, the sensory noise parameter that was left free to vary when fitting the model was the noise on measurements of ellipse aspect ratio. The model parameters estimated for subjects' data accord well with published data on human subjects' ability to discriminate the aspect ratios of ellipses in the image plane. These data suggest an effective standard deviation between 0.02 and 0.04 in subjects' estimates of aspect ratio (Regan & Hamstra, 1992). The model parameters fit to subjects' data were within this range (0.024 and 0.036 in Experiments 1 and 2, respectively). Even considering only the small conflict conditions, in which the parameters for the prior density on aspect ratios do not interact strongly with the sensory noise estimates, this suggests that subjects are near optimal in how they integrate the compression cue with stereopsis.

The prior model parameters fit to the data from each experiment further suggest that subjects were behaving as if in a very regular world. The proportion of random ellipses is small and the standard deviation of ellipses among the random ellipse set is also apparently small. Figure 12 illustrates the fitted prior distributions on aspect ratios in the random ellipse category. Below the graph are drawn ellipses with aspect ratios at the 95% bounds on aspect ratios derived from Experiment 1. It is notable, however, that although the range 0.75–1.333 seems small numerically, it appears much larger visually, when comparing the perceived shapes of the corresponding ellipses. Moreover, the small proportion of circles in the model that was fit to the data is enough to give rise to apparent down-weighting of the compression cue at large cue conflicts with stereopsis.

The effects of different perceptual factors on nonlinear cue integration behavior

As is traditionally done, we have quantified subjects' cue integration behavior using linear weights calculated by regressing subjects' slant settings against the slants suggested by each of a pair of cues. In the context of the nonlinear Bayesian model, however, it should be clear that this is strictly a means of quantifying subjects' average behavior. Because the weights are linear functions of subjects' corrected slant settings (eliminating quadratic biases between physical slant and subjects' slant settings), we could have as easily fit the model parameters to subjects' corrected slant settings. Calculating cue weights

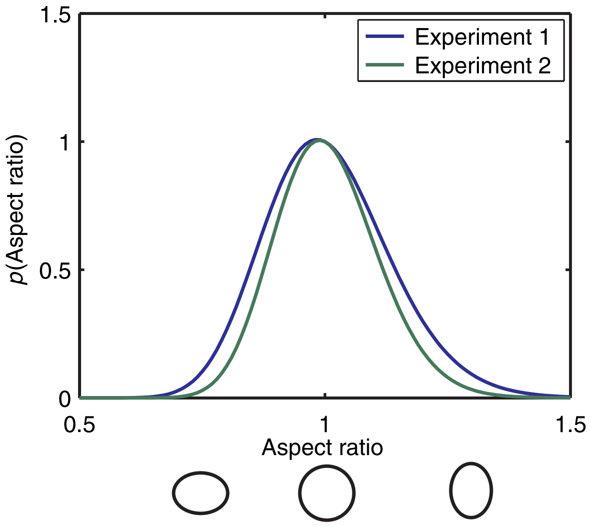

Figure 12. The probability density functions on aspect ratio for the ``random ellipse'' component of the mixed prior on figure shape, as fit to the data for the two experiments.

has the advantage of providing a picture of the relative importance of the cues to subjects' judgments; however, they should not be thought of as characterizing a model of the mechanism that implements cue integration.

Looking at how the ideal observers' cue weights change as a function of the model parameters also provides insight into the functional role of distinct parts of the perceptual system, particularly sensory noise and prior assumptions. Figure 13 shows how the model behaves for a number of variations on model parameters (holding the others fixed) for the 35° slant condition used in Experiment 1. Not surprisingly, changing the standard deviation of subjects' sensory estimates of aspect ratio in the retinal image has the effect of decreasing the weight that the model gives to the compression cue. This change is localized to the small cue-conflict conditions, because in these conditions, the circle model dominates subjects' estimates of slant from the figural cue, so that the primary source of uncertainty is in the sensory noise on aspect ratio measurements. At large cue conflicts, the effect of changing the sensory noise is negligible, because the random ellipse model dominates judgments and the sensory noise is small relative to the uncertainty in the prior distribution of aspect ratios in the world.

Changing the proportion of random ellipses assumed to be in the environment has a very similar effect on the weight function. Because the total likelihood function for the figure shape cue is an average of the likelihood functions for circles and random ellipses, weighted by the relative proportions of the two types of figures, the effective weight of the compression cue depends on that proportion. This effect is apparent at small cue conflicts but not at larger conflicts. This is because the random ellipse model has a prior peaked at $\alpha = 1$, as does the circle model. At large cue conflicts, the tails of likelihood function for the random ellipse model are proportionally so much larger than the tails of the likelihood function for the circle model that the random ellipse model overwhelms the circle model when stereoscopic cues conflict by a great deal with the circle interpretation of slant, even when the proportion of circles in the environment is high. Finally, changing the standard deviation of the random ellipse model has a somewhat complicated effect on performance—it lowers the compression cue weights at large cue conflicts but slightly raises compression cue weights at small conflicts.

Particularly interesting in these data are the effects of the two different parts of the prior model on subjects'

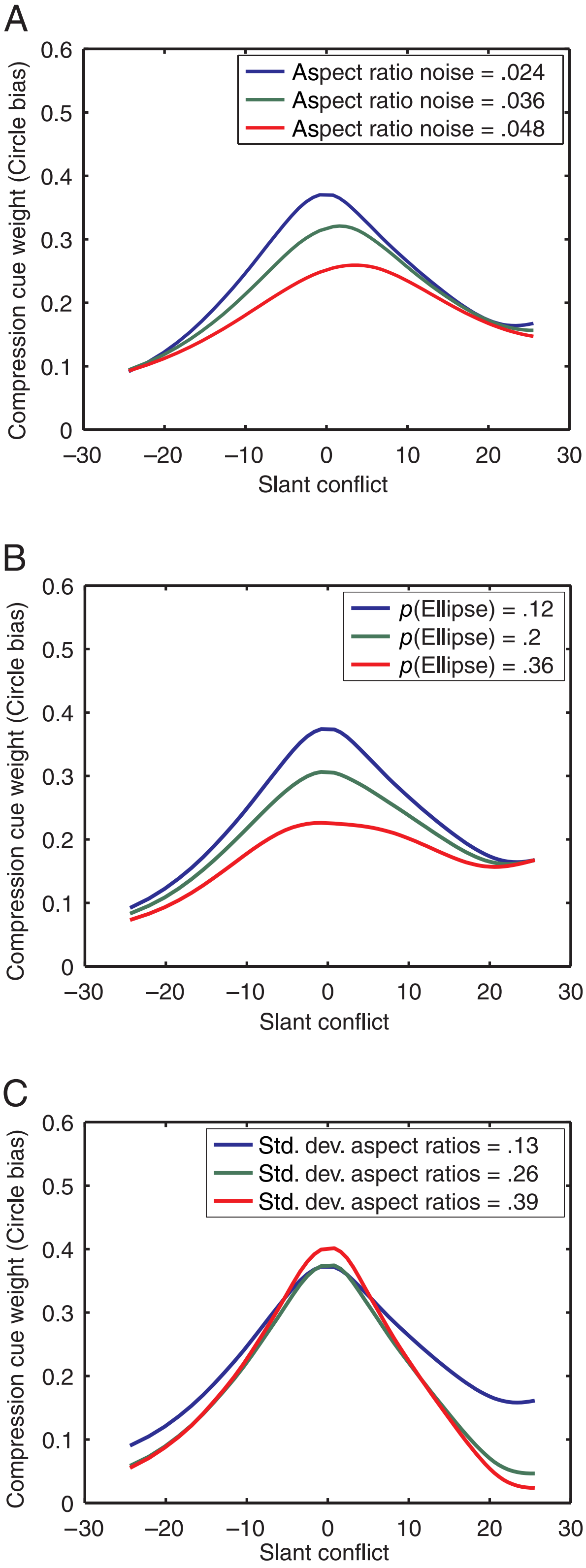

Figure 13. The effects of different model parameters on performance. In all cases, we have fixed all but one of the model parameters to those fit to subjects' data in Experiment 1, in which stereoscopic cues indicated a 35° slant. Different parameters (indicated in the legends) were varied for the simulations represented in the three graphs. The three graphs show performance for (A) different levels of sensory estimates of aspect ratio, represented as the standard deviation in aspect ratio measurements; (B) different proportions of random ellipses/circles in the world; and (C) different standard deviations on the range of ellipse aspect ratios included in the random ellipse category.

performance. First, increasing the proportion of random ellipses in the prior model shrinks the apparent weight of the compression cue at low cue-conflict levels. This derives from the fact that the likelihood function for slant-from-figure shape is a mixture of the functions derived for circles and random ellipses. Both priors are peaked at aspect ratios = 1 (circles) but have different spreads. Thus, both induce a bias toward a circle interpretation. When combined with the likelihood function from stereo disparities, the result is a combined likelihood function that is skewed toward the stereo slant interpretation (see Figure 2). The degree of skew depends on the weighting of the two component functions that make up the slant-from-figure shape likelihood. The skew shifts the expected value of slant derived from the joint cue likelihood function toward the peak of the slant-from-stereo likelihood.

The effect of the spread of ellipse aspect ratios in the random ellipse model is more complicated and less intuitive. That the compression cue weights at large cue conflicts shrink with increasing spread of this distribution simply reflects the fact that performance at large cue conflicts is dominated by the random ellipse model, which is less reliable when the ellipse category contains a broader range of shapes. The slight increase in compression cue weight at low conflicts derives from a more subtle behavior of the likelihood functions derived for different models. While the increased spread of the component likelihood function associated with the random ellipse category will tend to reduce the weight of the compression cue at small conflicts, the relative contribution of this component to overall performance depends on the absolute height of the function. This decreases with increasing spread of the prior on aspect ratios, which tends to down-weight its contribution to performance at low conflict conditions. The overall pattern is the result of this trade-off.

Is the prior on figure shape really categorical?

We have presented the categorical nature of the prior model as its critical feature—the one that drives robust performance at large cue conflicts. Mathematically, however, the behavior of the model fundamentally derives from the long tails of the prior distribution on aspect ratios—tails that do not go to zero as fast as a Gaussian. It is worth asking, therefore, how useful the mixture model is for characterizing human perceptual performance. The first answer to this question is empirical. The large weights that subjects give to the compression cue at small cue conflicts suggest a very strong peak in the prior density function at 1. If we assume that subjects behave near-optimally, we can assume that this weight is largely determined by the relative spreads of the slant-from-stereopsis and slant-from-figure shape likelihood functions near their peaks. We have fixed the spread on the slant-from-stereopsis likelihood function based on previous psychophysical data. The spread of the slant-from-figure shape likelihood function is a function of the spread in the prior near 1 and the sensory noise on aspect ratio measurements. As noted above, the sensory noise levels fit to the data are very near those estimated from psychophysical experiments on aspect ratio discrimination (Regan & Hamstra, 1992). This leaves little room for uncertainty induced by spread in the prior distribution on aspect ratios.

The second argument for a mixture model is largely conceptual. It seems appropriate for many of the prior models that we use to characterize objects in the world. Circles are ''special'' in our environment. Similarly, symmetry is ubiquitous in both artifactual and natural environments. Most objects are rigid (not simply biased toward being rigid). The list can go on. Furthermore, it matches our phenomenal experience. As one example of this, subjects who have participated in experiments like the one described here, in which we ran subjects 1 day using all circle stimuli or stimuli with small cue conflicts and then on another day used randomly shaped ellipses, spontaneously commented on the second day that we had changed the stimuli from all circles to circles combined with ellipses. Finally, mixture models with modal priors provide a natural constraint on constructing model priors with tight peaks and long tails.

Generalizations of the model

The Bayesian model presented here is a special case of hierarchical Bayesian inference (Tenenbaum, Griffiths, & Kemp, 2006). Here, we have only considered one aspect of the generative model that gives rise to the image data associated with sensory cues—the different prior models that might be applied to interpret a cue. This can be thought of as a discrete variable on which the likelihood function for the cue depends. Hierarchical Bayesian inference models can be applied to a wider range of nonlinear cue integration behavior. We describe a few of these generalizations here.

Multisensory integration

The problem of how the brain integrates information from multiple sensory modalities to infer the properties of objects (position, size, etc.) has recently garnered much attention in the psychophysics literature. Like earlier research on visual depth cue integration, this work has focused on the question of whether or not the brain combines multimodal information in a statistically optimal way when operating in a linear regime (it does; Alais & Burr, 2004; Battaglia, Jacobs, & Aslin, 2003;

Ernst & Banks, 2002). One might well ask how an observer should behave when faced with large conflicts between modalities, for example, when the direction of a sound signaled by audition is very different from the direction signaled by vision. The notion that different priors might be used to interpret one cue or another does not seem applicable in this situation. A very similar observation, however, does, that is, that the auditory and visual signals could arise from the same or different physical sources. Formulating the idea that sensory signals from different modalities could arise from the same or different causal events in the world in a Bayesian framework leads to essentially the same model presented here—a mixture of different models that could have generated the sensory data. Kording et al. (2007) have shown that this model can explain a range of nonlinear data in subjective judgments of auditory and visual source localization when the two signals are presented simultaneously.

Multiple sources of noise

The model that we applied to explain figural and stereoscopic cue integration assumes a special place for stereopsis, namely, that its likelihood is localized in the slant domain. This predicts that in the presence of large cue conflicts between stereopsis and monocular depth cues that rely on mixtures of priors, stereoscopic information will dominate in the sense that it will lead to a switch in the interpretation of the monocular cues. In special situations, an observer can attribute different causal events to the stereoscopic cues and the monocular cues. This is what happens when viewing a picture. The pictorial cues in a picture are attributed to the 3D layout of the scene rendered in the picture, whereas the stereoscopic cues are attributed to the paper on a scene that is rendered. This may also apply to our percepts of stereographic displays presented on a computer monitor and almost certainly does when care is not taken to eliminate many of the cues indicating that a display is flat. In the real world, such explanations of large cue conflicts are difficult to conceive. It remains possible, however, that the stereoscopic system is corrupted by qualitatively different noise sources that lead to different likelihood functions for depth or shape from disparity. Landy et al. (1995) acknowledge this as a motivating factor for robust integration schemes by noting that stereoscopic noise can be ‘‘local’’, as in simple Gaussian noise on disparity measures, or more global, as in noise caused by mismatches in the solution of the correspondence problem.

It is conceivable that for some stimuli, monocular cues can serve to down-weight stereoscopic cues in a rational way. For example, a perspective image of a slanted, square grid provides very reliable evidence that the texture is, in fact, a slanted square grid. The likelihood for a nonsquare interpretation of the grid is considerably lower than that for the square grid because of the Occam's razor effect alluded to previously—a form of the generic view argument for why we see perspective images of regular patterns so reliably rather than frontal views of irregular patterns that happen to mimic perspective. Such patterns would be highly accidental if drawn from an ensemble of random patterns. When the unconstrained interpretation of a monocular cue has a likelihood that is low enough, it could be lower than the likelihood that the noise in the stereoscopic system comes from an outlier process, leading to apparent down-weighting of the stereoscopic cue or even bimodal perceptual effects like those described by van Ee, Adams, and Mamassian (2003) for these types of stimuli.

Adaptation and recalibration

One way in which the brain can resolve large conflicts between cues from different modalities is to recalibrate how it interprets one of the cues. This happens when the conflicts maintain a particular size or sign over time, as in prism adaptation. The Bayesian account of this type of fast adaptation is that the calibration parameters (e.g., the gain between vergence angle and depth) can change over time due to a mixture of possible causes—slow drifts over time and catastrophic, sudden changes due to disease or injury. Smith, Ghazizadeh, and Shadmehr (2006) have applied this notion to model visuomotor adaptation and recovery from adaptation. They have modeled these effects as resulting from a Bayesian recalibration scheme that assumes that a mixture of processes could have caused the system's calibration parameters (the mapping between visual location and movement amplitude) to drift over different timescales. Similar ideas could be applied to adaptation processes that affect depth perception such as adaptation of the vergence signal used to calibrate depth-from-disparity estimates or the vestibular signal used to calibrate depth-from-motion parallax.

Conclusion

Subjects appeared to spontaneously down-weight the information about surface slant provided by figure shape relative to stereoscopic cues as the conflict between the cues grew. Their behavior was well fit by a Bayesian model that assumes a mixed prior on figure shapes that include categories for circles and random ellipses. Similar models apply to most other monocular depth cues, which rely on some form of strong prior constraints on objects in the world because those constraints do not apply to all objects. Understanding the particular patterns of nonlinear cue integration exhibited by different combinations of cues will require fully modeling these priors and how they combine with sensory noise to constrain the information provided by the cues. The

Bayesian framework described here provides a powerful tool for building parametric, predictive models of nonlinear cue integration performance.

Appendix A

We assume that for each trial, an observer estimates slant from a noisy measurement of the aspect ratio of the ellipse in the retinal image and noisy measurements of disparity. Rather than model the slant-from-disparity explicitly, we simplify it by representing the information provided by disparities as an estimate of slant corrupted by Gaussian noise. The Bayesian observer, therefore, computes a posterior probability density function on surface slant conditioned on these two measurements. For a particular combination of noisy measurements, A and S_{stereo}, the posterior distribution is given by the product of likelihood functions associated with the two measurements and a prior on surface slant,

$$p(S|A, S_{\text{stereo}}) = p(A|S)p(S_{\text{stereo}}|S)p(S). \tag{A1}$$

Using Equation 10 for $p(A \mid S)$ and assuming that surfaces are viewed from a uniform distribution on the view sphere, Equation A1 becomes

$$\begin{aligned} p(S|A, S_{\text{stereo}}) &= \frac{1}{\sqrt{2\pi}\sigma_A} \\ &\times \left\{ \begin{array}{l} \pi_{\text{circle}} \exp\left[-(A - \cos(S))^2/2\sigma_A^2\right] + \\ \pi_{\text{ellipse}} \displaystyle\int_0^{\infty} \exp\left[-(A - \alpha\cos(S))^2/2\sigma_A^2\right] p_{\text{ellipse}}(\alpha)\mathrm{d}\alpha \end{array} \right\} \\ &\times \exp\left[-\left(S_{\text{stereo}} - (S + \text{bias})^2\right)/2\sigma_{\text{stereo}}^2\right]\sin(S), \end{aligned} \tag{A2}$$

where σ_A is the standard deviation of the noise on measurements of aspect ratio in the image and σ_{stereo} is the standard deviation of the noise on slant-from-disparity measurements. The bias term allows us to model relative biases in subjects' estimates of slant-from-stereo and slant-from-figure shape. The first likelihood function in Equation A2 is a mixture of a likelihood function computed with the assumption that a figure is a circle and a likelihood function computed with the assumption that a figure is a randomly shaped ellipse with an aspect ratio drawn from the distribution $p_{\text{ellipse}}(\alpha)$. As described in the text, we model this as a log-Gaussian distribution,

$$p(\alpha) = \frac{1}{\alpha}\frac{1}{\sqrt{2\pi}\sigma_\alpha}\exp\left[-(\log\alpha)^2/2\sigma_\alpha^2\right]. \tag{A3}$$

We assume that on each trial, the observer sees an aspect ratio A that is a random sample from a Gaussian distribution with mean, $\alpha_{\text{stimulus}}\cos(S_{\text{stimulus}})$, and standard deviation, σ_A, and an estimate of slant-from-disparity that is a random sample from a Gaussian distribution with mean, S_{stimulus} + bias, and standard deviation, σ_{stereo}. The bias term allows us to incorporate into the model potential relative biases in subjects' estimates of slant-from-figure shape and slant-from-disparity. The Bayesian observer computes as its estimate of the slant the expected value of the posterior distribution,

$$\begin{aligned} S = E[S|A, S_{\text{stereo}}] \\ = \int_0^{\pi/2} S\, p(S|A, S_{\text{stereo}})\mathrm{d}S. \end{aligned} \tag{A4}$$

This choice of estimator minimizes the squared error of an observer's estimates. We have simulated estimators that use other criteria, for example, a MAP estimator that selects as its slant estimate the peak of the posterior density function. Simulation results were very similar using the different estimators.

The free parameters in the model are σ_A, σ_{stereo}, σ_α (which parameterizes the spread of $p_{\text{ellipse}}(\alpha)$—see Equation 7), and the relative slant bias. To fit the model to subjects' data, for candidate settings of the model's parameters, we applied the same analysis used to analyze subjects' data to the outputs of the model observer averaged over many noise samples of aspect ratio and slant-from-disparity for each stimulus condition. For each stimulus condition, represented as a combination of figure aspect ratio in the world α_i and slant S_i, we computed the expected slant estimate for the model over many trials as the integral of $p(S \mid A, S_{\text{stereo}})$ over all possible noisy values of A and S_{stereo} for that condition,

$$\begin{aligned} E[S|\alpha_i, S_i] = k\int_0^{\pi/2}\int_0^{\infty}\int_0^{\pi/2} S\, p(S|A, S_{\text{stereo}}) \\ \times \exp\left[-(A - \alpha_i\cos(S_i))^2/2\sigma_A^2\right] \\ \times \exp\left[-(S_{\text{stereo}} - (S_i + \text{bias}))^2/2\sigma_A^2\right]\mathrm{d}S\mathrm{d}A\mathrm{d}S_{\text{stereo}}, \end{aligned} \tag{A5}$$

where k is a normalizing constant that guarantees that the exponential distributions for sensory noise inside the integral integrate to 1 (because the range of integration is bounded on at least one side, the noise distributions are not, strictly speaking, Gaussian, although the bounds are many standard deviations away form the means).

The goodness of fit of the model to subjects' data was computed in two steps. First, we applied the same quadratic regression on the expected slant estimates for cue-consistent stimuli as we applied to subjects to remove the bias caused by the bias term in the model. We then computed the expected slant estimates for the test conditions in the experiment, corrected them using the quadratic fit to remove the bias created by the biased slant-from-stereo estimate, and used the corrected slant estimates to compute compression cue weights for each of the test conditions (see Equation 4). Second, we computed a χ^2 statistic from the difference between the models' expected compression cue weights and the average of subjects' weights in each of the test conditions,

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(w_i^{\text{model}} - \overline{w}_i^{\text{subjects}}\right)^2}{\sigma_i^2}, \tag{A6}$$

where N is the number of test conditions and σ_i is the standard error on the mean of subjects' compression cue weights in condition i. We fit the model parameters by minimizing Equation A6 using a simplex algorithm.

Acknowledgments

This research was supported by research grant NIH-EY013389 from the National Institutes of Health.

Commercial relationships: none.

Corresponding author: David C. Knill.
Email: knill@cvs.rochester.edu.
Address: Center for Visual Science, University of Rochester, Rochester, NY 14627, USA.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14,* 257–262. [PubMed] [Article]

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A, Optics, image science, and vision, 20,* 1391–1397. [PubMed]

Braunstein, M. L., & Payne, J. (1967). The effect of pattern and texture gradient on slant and shape judgments. *Perception & Psychophysics, 81,* 584–590.

Buckley, D., Frisby, J. P., & Blake, A. (1996). Does the human visual system implement an ideal observer theory of slant from texture? *Vision Research, 36,* 1163–1176. [PubMed]

Cumming, B. G., Johnston, E. B., & Parker, A. J. (1993). Effects of different texture cues on curved surfaces viewed stereoscopically. *Vision Research, 33,* 827–838. [PubMed]

Cutting, J. E., & Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General, 113,* 198–216. [PubMed]

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415,* 429–433. [PubMed]

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision, 4*(12):1, 967–992, http://journalofvision.org/4/12/1/, doi:10.1167/4.12.1. [PubMed] [Article]

Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences, 6,* 345–350. [PubMed]

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research, 34,* 2259–2275. [PubMed]

Johnston, E. B., Cumming, B. G., & Parker, A. J. (1993). Integration of depth modules: Stereopsis and texture. *Vision Research, 33,* 813–826. [PubMed]

Knill, D. C. (1998a). Discrimination of planar surface slant from texture: Human and ideal observers compared. *Vision Research, 38,* 1683–1711. [PubMed]

Knill, D. C. (1998b). Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Research, 38,* 2635–2656. [PubMed]

Knill, D. C. (2003). Mixture models and the probabilistic structure of depth cues. *Vision Research, 43,* 831–854. [PubMed]

Knill, D. C. (2005). Reaching for visual cues to depth: The brain combines depth cues differently for motor control and perception. *Journal of Vision, 5*(2):2, 103–115, http://journalofvision.org/5/2/2/, doi:10.1167/5.2.2. [PubMed] [Article]

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research, 43,* 2539–2558. [PubMed]

Kording, K. P., Beierholm, U., et al. (2007). *Causal inference in multisensory perception.* Manuscript submitted for publication.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research, 35,* 389–412. [PubMed]

Marr, D. (1982). *Vision.* Cambridge: MIT Press.

Regan, D., & Hamstra, S. J. (1992). Shape discrimination and the judgment of perfect symmetry: Dissociation of shape from size. *Vision Research, 32,* 1845–1864. [PubMed]

Smith, M. A., Ghazizadeh, A., & Shadmehr, R. (2006). Interacting adaptive processes with different time-scales underlie short-term motor learning. *Public Library of Science: Biology, 4,* e179. [PubMed] [Article]

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences, 10,* 309–318. [PubMed]

van Ee, R., Adams, W. J., & Mamassian, P. (2003). Bayesian modeling of cue interaction: Bistability in stereoscopic slant perception. *Journal of the Optical Society of America A, Optics, image science, and vision, 20,* 1398–1406. [PubMed]

Wonnacott, T. H., & Wonnacott, R. J. (1990). *Introductory statistics*. New York: John Wiley and Sons.

Young, M. J., Landy, M. S., & Maloney, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research, 33,* 2685–2696. [PubMed]

Yuille, A., & Bulthoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian inference*. Cambridge, England: Cambridge University Press.