

# Slant from texture and disparity cues: Optimal cue combination

**James M. Hillis**

Department of Psychology, University of Pennsylvania,  
Philadelphia, PA, USA



**Simon J. Watt**

Department of Psychology, University of Wales,  
Bangor, Wales, UK



**Michael S. Landy**

Department of Psychology & Center for Neural Science,  
New York University, New York, NY, USA



**Martin S. Banks**

Vision Science Program, Department of Psychology, &  
Wills Neuroscience Institute, University of California,  
Berkeley, CA, USA



How does the visual system combine information from different depth cues to estimate three-dimensional scene parameters? We tested a maximum-likelihood estimation (MLE) model of cue combination for perspective (texture) and binocular disparity cues to surface slant. By factoring the reliability of each cue into the combination process, MLE provides more reliable estimates of slant than would be available from either cue alone. We measured the reliability of each cue in isolation across a range of slants and distances using a slant-discrimination task. The reliability of the texture cue increases as  $|\text{slant}|$  increases and does not change with distance. The reliability of the disparity cue decreases as distance increases and varies with slant in a way that also depends on viewing distance. The trends in the single-cue data can be understood in terms of the information available in the retinal images and issues related to solving the binocular correspondence problem. To test the MLE model, we measured perceived slant of two-cue stimuli when disparity and texture were in conflict and the reliability of slant estimation when both cues were available. Results from the two-cue study indicate, consistent with the MLE model, that observers weight each cue according to its relative reliability: Disparity weight decreased as distance and  $|\text{slant}|$  increased. We also observed the expected improvement in slant estimation when both cues were available. With few discrepancies, our data indicate that observers combine cues in a statistically optimal fashion and thereby reduce the variance of slant estimates below that which could be achieved from either cue alone. These results are consistent with other studies that quantitatively examined the MLE model of cue combination. Thus, there is a growing empirical consensus that MLE provides a good quantitative account of cue combination and that sensory information is used in a manner that maximizes the precision of perceptual estimates.

**Keywords:** depth perception, cue combination, stereopsis, Bayesian perception, texture gradient

## Introduction

The fundamental problem in depth perception is due to the geometry of perspective projection, which reduces the three-dimensional (3D) coordinates of the visual scene to the 2D coordinates of the retinal images. The third dimension of space has to be inferred from the 2D images. The visual system uses several sources of information—“depth cues” such as disparity, perspective, and motion parallax—to estimate the layout of the 3D scene. Estimates based on each individual cue are subject to error. By combining information from several depth cues, the visual system could estimate 3D layout with greater precision across a wider variety of viewing situations than it could by relying on any one cue alone. To realize this advantage, the reliability of each depth cue must be factored into the combination rule. Factoring in reliability is complicated because the

reliability of individual depth cues depends on scene parameters in different ways. Are variations in depth cue reliability with scene geometry factored into the cue-combination rule? To examine this question, we compared human slant discrimination ability based on disparity and texture cues to a model of statistically optimal cue combination. Slant estimation from texture and disparity is an interesting case to examine because the reliabilities of disparity and texture cues vary in different ways with slant and viewing distance. Knill and Saunders (2003) examined the combination of texture and disparity as a function of slant with a similar approach to what we present here. We have expanded their experiments to include surfaces slanted about a vertical axis and surfaces at multiple viewing distances (see [Discussion: Comparison to other studies](#)). We measured the reliability of slant estimates from each cue in isolation across a range of slants and distances and used an

optimal cue-combination rule to predict the appearance of two-cue stimuli and the precision of slant estimation with two-cue stimuli. We then compared these predictions to the results of two-cue slant discrimination experiments.

## Optimal cue combination

Visual estimates of slant from any depth cue are subject to error. For example, perceived slant from a given texture gradient will vary from one instance to another due to the statistical nature of slant information from texture and errors in the measurement of the gradient (Blake, Bülthoff, & Sheinberg, 1993; Cutting & Millard, 1984; Knill, 1998a). When more than one depth cue is available and informative, one can in principle reduce the uncertainty associated with any one of the cues by combining across cues (for a review and derivation of the following results, see Oruç, Maloney, & Landy, 2003).

One approach to optimizing cue combination is statistical: What cue-combination rule results in an estimator that is unbiased and has minimum variance? Assume that the observer has unbiased estimates  $\hat{S}_d$  and  $\hat{S}_t$  of the slant of a surface based on disparity and texture cues, respectively. Assume further that errors in these estimates are uncorrelated and have variances  $\sigma_d^2$  and  $\sigma_t^2$ . If we combine the two estimates linearly, the rule that yields the minimum-variance, unbiased estimate is a weighted average that satisfies (Cochran, 1937)

$$\hat{S} = w_d \hat{S}_d + w_t \hat{S}_t, \quad (1)$$

where

$$w_d = \frac{r_d}{r_d + r_t} \quad \text{and} \quad w_t = \frac{r_t}{r_d + r_t}, \quad (2)$$

and  $r_d$  and  $r_t$  are the *reliabilities* of the two cues (e.g.,  $r_d = 1/\sigma_d^2$ ). Furthermore, if errors associated with the individual estimators are Gaussian, no other (nonlinear) rule has lower variance.

An alternative approach is to apply Bayesian methods (for reviews, see Kersten, Mamassian, & Yuille, 2004; Mamassian, Landy, & Maloney, 2002). In the absence of any immediate consequences to an observer's actions (payoffs and penalties), the maximum a posteriori (MAP) estimate is typically employed. That is, the observer chooses a slant estimate  $\hat{S}$  that is most probable given the image data. We assume the image data can be segregated into those data  $I_d$  used to estimate slant from disparity and  $I_t$  used to estimate slant from texture. Thus, we choose the value of  $\hat{S}$  that maximizes  $p(\hat{S} | I_d, I_t)$ . Applying Bayes' rule, and assuming that the two cues are conditionally independent, we derive

$$p(\hat{S} | I_d, I_t) \propto p(I_d | \hat{S}) p(I_t | \hat{S}) p(\hat{S}). \quad (3)$$

The first two terms on the right side of the equation are the likelihood functions for each cue characterizing the probability of observing the image data if  $\hat{S}$  is the actual slant.

The last term is the prior distribution, which is the probability of observing  $\hat{S}$  in the scene, independent of the image data. If the likelihoods and prior are Gaussian, the MAP estimate has the same form as the minimum variance, linear combination estimate

$$\hat{S} = w_d \hat{S}_d + w_t \hat{S}_t + w_p \hat{S}_p, \quad (4)$$

where

$$w_d = \frac{r_d}{r_d + r_t + r_p}, \quad w_t = \frac{r_t}{r_d + r_t + r_p}, \quad (5)$$

and  $w_p = \frac{r_p}{r_d + r_t + r_p}.$

Here,  $\hat{S}_d$  and  $\hat{S}_t$  are the maximum-likelihood estimates the observer would have made from each cue in isolation (the mean of the respective Gaussian distributions), and  $\hat{S}_p$  is the mean of the prior. The  $r_i$  are the reliabilities of the respective distributions (likelihoods and prior). If the prior has large variance relative to the individual cue likelihoods, Equations 4 and 5 reduce to Equations 1 and 2, which also yields the most likely slant to have caused the current sensory data (i.e., it is the maximum-likelihood estimate or MLE). For our conditions, the variance of the individual cues is much smaller than the prior's variance (see *Ideal observer models* in Discussion), so we will use Equations 1 and 2 throughout.

By following the strategy described by Equation 2, the variance of the weighted average  $\hat{S}$  is

$$\sigma^2 = \frac{\sigma_d^2 \sigma_t^2}{\sigma_d^2 + \sigma_t^2} \quad \text{or, equivalently, } r = r_d + r_t. \quad (6)$$

The variance of  $\hat{S}$  is lower than the variance of either single-cue estimate.

Many investigations of sensory cue combination have shown that cue reliability is taken into account in the estimation process (Backus & Banks, 1999; Banks, Hooge, & Backus, 2001; Battaglia, Jacobs, & Aslin, 2003; Buckley & Frisby, 1993; Frisby, Buckley, & Horsman, 1995; Jacobs, 1999; Körding & Wolpert, 2004; Rogers & Bradshaw, 1995; van Beers, Sittig, & Denier van der Gon, 1998; van Beers, Wolpert, & Haggard, 2002; Young, Landy, & Maloney, 1993). Only five studies, however, have tested the quantitative predictions of the MLE model expressed by Equations 1, 2, and 6, to determine if sensory cue combination is statistically optimal (Alais & Burr, 2004; Ernst & Banks, 2002; Gepshtein & Banks, 2003; Knill & Saunders, 2003; Landy & Kojima, 2001). These five studies measured reliabilities of individual cues ( $r_i$  in Equations 2 and 6) and empirically tested predictions for both the appearance and discrimination thresholds for stimuli when both cues were present (provided by Equations 1, 2, and 6). All five reported that the combination is quite close to the one predicted by those equations. In these five experiments, the

variances of estimates derived from single cues were measured by conducting two-interval, forced-choice (2IFC) discrimination experiments when only one cue was informative. For example, Ernst and Banks (2002) conducted size-discrimination experiments for vision alone and haptics alone and then fit cumulative Gaussians to the two psychometric functions. The variance parameter of the Gaussians provided estimates of the variances of the underlying visual and haptic estimators. Equations 1, 2, and 6 were then successfully used to predict the results of two-cue (visual-haptic) experiments.

## Preview

The experiments presented here used the strategy of Ernst and Banks (2002) to ask whether texture and disparity cues to slant are combined in a statistically optimal fashion. The reliability of texture and disparity cues to slant vary with viewing geometry in different ways. First, the reliability of texture should increase with increasing slant because the image changes associated with a given change in slant increase (Blake et al., 1993; Knill, 1998a). This relationship between reliability and slant is reflected in human performance (Knill, 1998b). Theoretical and empirical analyses of the reliability of disparity as a function of slant have not been conducted, but it is unlikely that it changes significantly (Banks et al., 2001; Knill & Saunders, 2003). Second, because the magnitude of binocular disparities for a given depth difference decreases as viewing distance increases, the reliability of slant and curvature estimated from binocular disparity should decrease as viewing distance is increased. Experiments confirm that it does (Howard & Rogers, 2002; Ogle, 1950). On theoretical grounds, the reliability of texture-specified slant, for a fixed retinal-image density, should not change with distance. If a given textured surface is doubled in size and viewed from twice the distance, the retinal image is unchanged. Thus, optimal combination of disparity and texture cues to slant should involve complex changes in the weights given to the two cues depending on base slant and viewing distance.

We looked for evidence that the visual system weights the two cues appropriately across a range of slants and distances. Because the reliability of the texture cue increases with slant, we expect the texture weight to increase as slant increases. Because the reliability of disparity decreases as viewing distance increases, we expect the texture weight to increase as distance increases. As in the previous studies, we determined the reliability of the individual cues with 2IFC discrimination experiments. Then, we measured the apparent slant and slant discrimination performance for two-cue stimuli. As we shall see, the MLE cue-combination predictions based on the single-cue experiments (Equations 1, 2, and 6) were largely in accord with the data from the two-cue experiment.

## Methods

### Subjects

Four observers participated. Two were not aware of the experimental hypotheses (ACD and RM). All had normal stereopsis and did not manifest eye misalignment in normal viewing situations.

### Apparatus

All stimuli were displayed on a custom-designed stereoscope with two mirrors and two CRTs (one for each eye; see Backus, Banks, van Ee, & Crowell, 1999). Each mirror and CRT was attached to an arm that rotated about a vertical axis passing through the eye's center of rotation. With this arrangement, the eye and stereoscope arm rotate on a common axis, so when we change the vergence distance, the mapping between the stimulus array and the retina is unaltered (for fixed accommodation).

We used anti-aliasing to specify dot position to subpixel accuracy. To ensure accurate reproduction of visual direction, we spatially calibrated each CRT to eliminate distortions in the images (for details, see Backus et al., 1999).

The observer's head position was stabilized using a bite bar fastened to an adjustable mount. Each observer had a personal mount so that the vertical axes of rotation of left and right eyes were collinear with the rotation axes of the two stereoscope arms (for details, see Hillis & Banks, 2001). The optical distance between the center of rotation of each eye and the face of the CRT was 40 cm.

### Stimuli

Stimuli were virtual planes slanted about a vertical axis (i.e., tilt = 0 deg). We independently manipulated two cues to slant: disparity and texture. In single-cue measurements, we isolated one or the other of the two cues. In two-cue measurements, both cues were informative, but could have different slant values. Viewing distance was 19.1, 57.3, or 171.9 cm. Example stimuli are shown in Figure 1.

The texture cue was the perspective projection of planar patches textured with Voronoi patterns with an average of 64 Voronoi cells per patch (de Berg, van Kreveld, Overmars, & Schwarzkopf, 2000; Figure 1, bottom panel). The actual number of cells varied from trial to trial depending on the randomly selected width of the patch (i.e., cells with a constant average area filled the area of the elliptical patch). Voronoi patterns were generated from a jittered grid of dots. On a frontoparallel plane, a regular grid of points was defined. Then, each point on the dot grid was perturbed horizontally and vertically (uniform distribution from  $-0.3$  to  $0.3$  deg). The Voronoi pattern defined by these points was then computed. Finally, the resulting textured plane was rotated by an amount equal to the texture-defined slant. To isolate the texture cue, the stimuli were viewed monocularly. The visible portion of the plane was

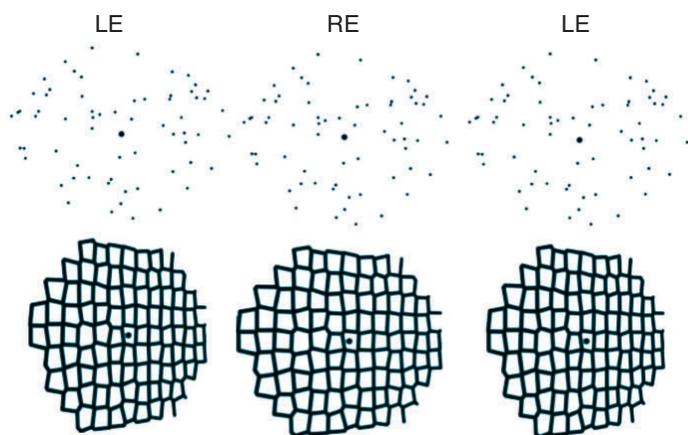


Figure 1. Examples of the stimuli. Cross fuse or divergently fuse to see the appropriate slants. The upper stimulus is an example of the disparity-alone stimulus. It has a negative slant (right side near). The lower row provides examples of the texture-alone stimulus when viewed monocularly and the disparity-texture stimulus when viewed binocularly. The disparity- and texture-specified slants are positive (right side far).

elliptical with a height of 15 deg. The width on each presentation was randomly chosen from a uniform distribution from 15 to 20 deg when the stimulus was frontoparallel. The stimulus was then rotated to the appropriate slant. Thus, the retinal shape of the stimulus outline was an unreliable cue to slant.

The disparity cue to slant was the difference between left- and right-eye projections (calculated for each observer's inter pupillary distance). To isolate the disparity cue, the stimulus was defined by sparse random dots (Figure 1, top panel). Each stimulus consisted of 64 dots, with positions

randomly drawn from a uniform distribution (note that the texture gradient specified by the dots was therefore consistent with a frontoparallel plane). Dot density was  $\sim 0.3$  dots/deg<sup>2</sup>.

When both cues were present, disparity and texture could be consistent ( $S_d = S_t$ ) or they could be in conflict. In the no-conflict case, homogeneous Voronoi-textured surfaces were projected directly to the two eyes. In cue-conflict cases, we first calculated a perspective projection of the texture with slant  $S_t$  at the Cyclopean eye (Figure 2, left panel). We then found the intersections of rays through this Cyclopean projection with a surface patch at the disparity-specified slant  $S_d$  (Figure 2, middle panel). The markings on this latter surface were then projected to the left and right eyes to form the two monocular images (Figure 2, right panel).

## Control experiments and procedures to validate single-cue measurements

We went to some lengths to ensure that the single-cue experiments measured the variances of the disparity and texture estimators in a fashion appropriate for making two-cue predictions. In this section, we describe control experiments and methodological procedures used to achieve that goal.

### 1. Are disparity-alone measurements affected by monocular slant signals?

To make sure that only binocular information determined slant discrimination in the disparity-alone case, we conducted two control experiments.

First, to make sure that the stimulus did not provide a monocular cue to slant, we measured monocular slant-

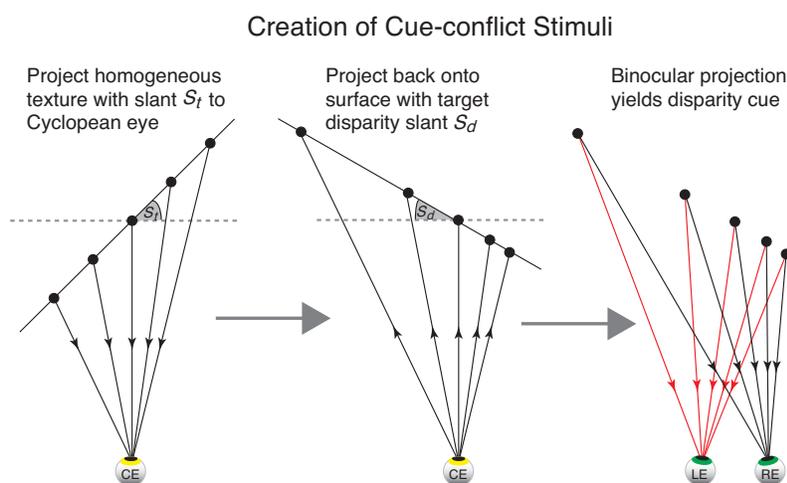


Figure 2. Creation of the cue-conflict stimuli. Left: Perspective projection of a homogeneously textured surface with the Cyclopean eye as the center of projection. This projection creates the texture-specified slant,  $S_t$ . The rays from the surface toward the eye are used in the next step. Middle: A virtual surface with the disparity-specified slant,  $S_d$ , is created. The rays from the first step are back-projected from the Cyclopean eye to find their intersections with the disparity-defined surface. They are marked in the diagram with black points. Right: Viewing the black points binocularly yields the cue-conflict stimulus containing the texture-specified slant in the left panel and the disparity-specified slant in the middle panel.

discrimination thresholds at various slants for the 64-dot stimulus. Observers could not reliably discriminate anything but large slant changes, and those changes were at least a factor of 10 larger than the thresholds in the disparity-alone experiment. We conclude that there is no useful monocular slant information in the 64-dot random-dot stimulus.

Second, we wanted to make sure that we presented enough dots in the display for disparity-based thresholds to be as low as possible while still isolating the disparity estimator. The details of this control experiment and the results are provided in [Appendix A](#). We found that threshold decreased as dot number increased from 2 to 32 and then leveled off beyond 32 dots. With 64 dots, disparity-based thresholds were as low as they could be. The results were simpler for observer JMH than for ACD: ACD may have given some weight to the texture signal at base slants different from 0 deg. We will return to this point when we discuss her two-cue data (in [Discussion: Summary of results](#)).

## **2. Are disparity-alone measurements based on perceived slant or on only the disparity gradient?**

To combine two cues for slant, the cues must be promoted to the same units. Disparity signals alone do not provide a slant estimate because they must be scaled or normalized for distance (Gårding, Porrill, Mayhew, & Frisby, 1995). We were concerned that observers might perform the slant-discrimination task in the single-cue, disparity-alone case by comparing only the disparity gradients in the two stimulus intervals. Said another way, they might perform the task without normalizing the disparity signals into slant estimates. To test the MLE model, we must acquire valid measures of the reliabilities of single-cue slant estimates. In the disparity-alone condition, this means our measure must reflect the process of scaling the disparity signal into units of slant. If the task in the disparity-alone condition were done without normalizing the disparity signal, the psychometric data would not reflect errors introduced by the scaling process (which, within the framework of weighted-linear cue combination, is essential for combining disparity and texture signals), and we would underestimate the variance of disparity-based slant estimates. We therefore looked for evidence that observers scale the disparity signal for distance in a discrimination task with our disparity-alone stimuli. Observers performed the slant-discrimination task with the disparity-alone stimulus, but with the comparison stimuli appearing at different distances relative to the standard stimulus. We found that most observers (importantly, JMH and ACD) take distance into account when performing the slant-discrimination task; that is, they do not perform the task by only comparing the disparity gradients in the two stimulus intervals. The details of the experiment and results are described in [Appendix B](#). The results of this control experiment support the assumption that the disparity-alone measurements pro-

vide valid estimates of the reliability of the disparity-based slant estimator.

## **3. Are the estimates of the single-cue reliabilities valid for the two-cue experiment?**

The single-cue data were used to specify the model's parameters (single-cue variances) and the model was then used to predict the two-cue data. An important assumption is that the appropriate variances are being measured in the single-cue experiments. One might question this assumption for the measurements of the disparity estimator's reliability because a different type of stimulus was used in the single-cue, disparity-alone condition (random-dot stimuli) than in the two-cue condition (Voronoi stimuli). This concern cannot be addressed by using Voronoi stimuli in the disparity-alone condition because such stimuli provide salient texture cues to slant. We can, however, check the validity of using random-dot stimuli by comparing single-cue thresholds with those stimuli to two-cue thresholds when the texture weight is expected to be approximately zero. This check, described in [Results: Just-noticeable differences](#), confirmed the validity of our assumption for observer JMH (the only observer for whom the required data were available).

A similar concern can be raised about using monocular stimuli in the texture-alone condition to measure texture reliability in the two-cue experiment. The stimulus in the two-cue experiment is binocular, so the visual system receives two samples while there is only one sample in the texture-alone experiment. The two samples will not be the same because the texture-specified slant at the left eye necessarily differs from the texture slant at the right eye (see [Appendix D](#)). Thus, the visual system must integrate two texture-gradient signals into one binocular estimate before combining with the slant estimated from disparity. The presence of two samples in the two-cue experiment might reduce the uncertainty associated with the texture cue in a fashion similar to the reduction in contrast threshold with binocular viewing (Legge, 1984). We can check this by comparing the monocular single-cue thresholds to binocular two-cue thresholds when the texture weight is expected to be approximately one. This check, described in [Results: Just-noticeable differences](#), confirmed the validity of our assumption that the monocular measurements were a valid estimate of the variance of the texture estimator in the two-cue experiment.

## **4. Do unmodeled slant cues affect responses?**

We wanted to isolate the cues of disparity and texture, so we had to consider whether other slant cues might be present in the display. Three cues—the blur gradient, accommodation, and the phosphor grid of the CRTs—always signaled a slant of 0 deg. If the observer failed to ignore those conflicting cues, the variances we measured would be higher than the true variances associated with disparity and texture. To reduce the salience of all three cues, we placed diffusers on the faces of the CRTs to blur the stimuli

slightly. Blurring the stimuli decreases the probability that the observers used the blur gradient because the blur gradient is a less reliable depth cue with blurred as opposed to sharply focused stimuli (Mather & Smith, 2002). Blurring should also decrease the probability that accommodation was used as a depth cue because humans accommodate inaccurately if at all to blurred stimuli (Heath, 1956). The diffusers also made the phosphor grid invisible.

### Procedure: Single-cue conditions

To estimate the reliabilities of the texture- and disparity-based slant estimates, we obtained psychometric functions for texture and disparity presented in isolation at several base slants ( $\pm 70$ ,  $\pm 60$ ,  $\pm 45$ ,  $\pm 30$ ,  $\pm 15$ , and  $0$  deg) and distances (19.1, 57.3, and 171.9 cm) for two observers (ACD and JMH). The other two observers participated in a subset of these conditions. We used a 2IFC task with no feedback. On each trial, the observer indicated which of two stimuli—one at the base slant and the other at the base slant  $\pm \Delta S$ —had the greater apparent slant. The stimuli were displayed for 1.5 s with a 0.3-s interstimulus interval. We used staircases to control the value of  $\Delta S$  and four reversal rules—3-down/1-up, 1-down/3-up, 2-down/1-up, and 1-down/2-up—to sample points along the entire psychometric function. At least eight staircases were employed for each psychometric function for ACD and JMH, which corresponds to approximately 350-450 trials per function (each staircase was terminated after 12 reversals). At least two, but typically more, staircases were employed for RM and MSB. In each session, at least four interleaved staircases were run: two base slants (one positive and one negative to avoid adaptation) with two staircases each. Viewing distance was fixed in each session.

### Procedure: Two-cue conditions

The procedure in the two-cue conditions was the same as in the single-cue conditions except that a no-conflict stimulus (disparity- and texture-specified slants equal to one another) and a conflict stimulus (disparity and texture slants not necessarily equal) were presented on each trial. Figure 3 depicts the disparity- and texture-defined slants of the no-conflict and conflict stimuli. In both panels, the slant specified by the disparity cue is plotted on the abscissa and the slant specified by the texture cue on the ordinate. The conflict stimulus had one cue set to a base slant ( $S_b = \pm 60$ ,  $\pm 30$ , or  $0$  deg) and the other cue was perturbed. The left panel depicts the conflict stimuli when disparity was perturbed and the right panel shows the stimuli when texture was perturbed. The perturbed cue had incremental slants of  $\pm 10$ ,  $\pm 5$ , or  $0$  deg relative to the unperturbed cue, so the conflict was always small. In previous work with quite similar stimuli, a difference of 10 deg between the disparity- and texture-specified slants was generally not detectable (Hillis, Ernst, Banks, & Landy, 2002). The five possible perturbed cue values are represented along the abscissa and ordinate in the left and right panels, respec-

tively. On each trial, a conflict stimulus and a no-conflict stimulus were presented and the observer indicated the one containing the apparently greater slant. No feedback was given. The value of the no-conflict stimulus was varied according to staircase procedures to map out the psychometric function. At least four staircases were run per experimental session: two conflict conditions for each of two base slants.

Figure 3 also shows how we determined the point of subjective equality (PSE), the value of the no-conflict stimulus with the same average perceived slant as the conflict stimulus. The enlarged bold symbols—the dark-blue circle in the left panel and black diamond in the right—represent two particular conflict stimuli.  $\Delta$  represents the incremental slant of the perturbed cue in the conflict stimulus, and  $\delta$  represents the increment given to the no-conflict stimulus as the staircase procedure varies its slant. As  $\delta$  is increased and thereby the slant of the no-conflict stimulus is increased (represented in the figure by displacement up along the main diagonal where  $S_t = S_d$ ), the observer will be increasingly likely to report that it had greater slant than the conflict stimulus. At some value of  $\delta$ , the no-conflict stimulus will on average have the same apparent slant as the conflict stimulus; this is the PSE. If the cue weights are constant across small variations in slant, we can determine the weights from this value of  $\delta$ .

Consider first the conflict stimulus. From Equations 1-2 and the fact that disparity-defined stimulus slant

#### Stimulus Variation in Two-cue Experiment

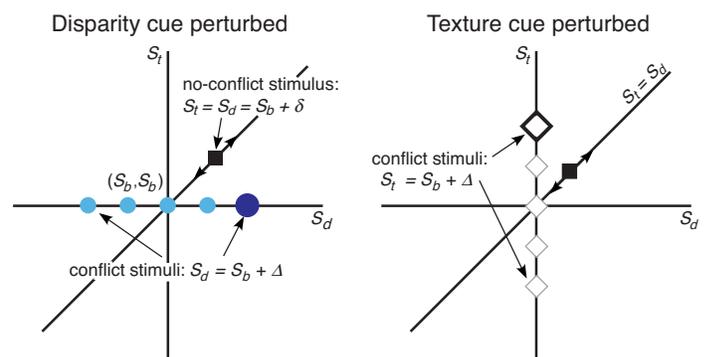


Figure 3. Depiction of the stimulus values and their manipulation in the two-cue experiment. Both graphs plot the disparity-specified slant ( $S_d$ ) on the abscissa and the texture-specified slant ( $S_t$ ) on the ordinate for the conflict and no-conflict stimuli. The base slant is represented by the origins ( $S_b, S_b$ ). The conflict stimulus either had the disparity-specified slant perturbed from the base slant by  $\Delta$  (depicted in the left panel) or the texture-specified slant perturbed by  $\Delta$  (right panel). Five different conflicts were presented for each base slant and those are represented by the blue circles in the left panel and gray diamonds in the right panel. In the no-conflict stimulus, the disparity- and texture-specified slants were equal to one another. The staircase procedure varied the increments added to the base slant:  $S_d = S_t = S_b + \delta$ .

$S_d = S_t + \Delta = S_b + \Delta$  (where  $S_b$  is the base slant), the expected value of the estimated slant of the conflict stimulus is

$$S_c = w_d(S_b + \Delta) + (1 - w_d)S_b = S_b + w_d\Delta. \tag{7}$$

Now consider the no-conflict stimulus. From Equations 1-2 and the fact that  $S_d = S_t = S_b + \delta$ , the expected value of the estimated slant of the no-conflict stimulus is

$$S_{nc} = w_d(S_b + \delta) + (1 - w_d)(S_b + \delta) = S_b + \delta. \tag{8}$$

The conflict and no-conflict stimuli will have the same perceived slants when  $S_c = S_{nc}$  and from Equations 7 and 8, we have

$$w_d = \delta / \Delta. \tag{9}$$

Thus, the two-cue experiment yields an estimate of the PSE from which we can determine the weights given to disparity and texture. The assumption that the weights are constant for even small variations is inconsistent with statistically optimal slant estimation, in which the weights vary as a function of slant. However, given the precision of our measurements and the rate of change of cue reliability, the fixed local weight assumption provides a reasonable approximation.

We can also plot the percentage of judgments for which the no-conflict stimulus appeared to have greater slant as a function of its slant. The slopes of such psychometric functions index the discriminability of the stimuli (discussed below in Results: Just-noticeable differences).

## Results

### Specifying the predictions

To quantify the predictions of the MLE model, we need estimates of the variances of the single-cue estimators ( $\sigma_d^2$  and  $\sigma_t^2$ , or equivalently, the reliabilities  $r_d$  and  $r_t$  in Equations 2 and 6). To estimate these variances, we fit the psychometric data with a cumulative Gaussian using a maximum-likelihood criterion. The standard deviations of the resulting functions were divided by  $\sqrt{2}$  (because the psychophysical procedure was 2IFC) to yield estimates of the standard deviations of the underlying slant estimators (Green & Swets, 1974). We call these just-noticeable differences (JNDs) because they represent the slant difference that is correctly discriminated ~76% of the time.

Figure 4 shows the JND estimates for JMH and ACD (JNDs for RM and MSB, whose performance was tested at only one distance, were similar to the those shown here and are plotted in Figure 1S). Each row of panels represents data from one observer. The left column shows the

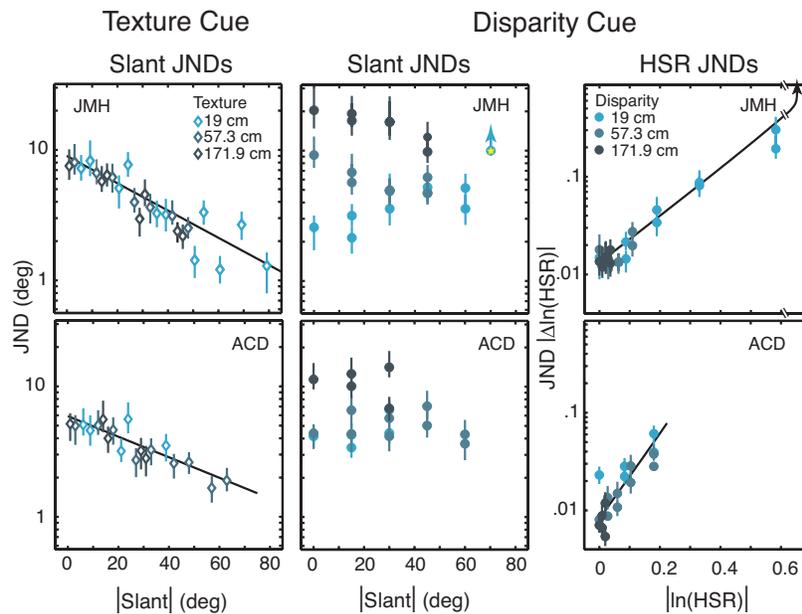


Figure 4. Just-noticeable slant differences (JNDs) for the single-cue experiment. JNDs are plotted as a function of slant or horizontal disparity. Different symbol colors represent data for different viewing distances: 19.1, 57.3, and 171.9 cm. Left and middle columns: JNDs for texture-alone and disparity-alone, respectively, as a function of the absolute value of the slant. Error bars are 95% confidence intervals. Right column: JNDs for disparity alone plotted in terms of  $HSR$ . The ordinate is the difference between the absolute values of the natural log of  $HSR$  for the base slant and the just-noticeably different slant. The abscissa is the absolute value of the natural log of  $HSR$  of the base slant. Error bars are 95% confidence intervals. The lines in the left column (texture-alone) and right column (disparity-alone) are maximum-likelihood curve fits to the data ( $JND = \alpha e^{\beta x}$ ). The break in the curve and the upward pointing arrow in JMH's disparity fit indicates that JNDs go to infinity somewhere between  $\ln(HSR)$  of 0.58 and 0.97 (slants of 60 and 70 deg at 19.1 cm).

texture-alone data: JNDs in units of slant are plotted as a function of the absolute value of base slant (there was no apparent difference in the results for positive and negative slants). Different symbols represent data from different viewing distances. As expected, texture JNDs did not vary systematically with distance. Also as expected (Knill, 1998a), texture JNDs decreased as the absolute value of slant increased.

The middle column of Figure 4 shows the data for the disparity-alone condition: Slant JNDs are plotted as a function of the absolute value of base slant (there were no systematic differences for positive and negative slants). As expected from the viewing geometry (Equations 1 and 2 in Backus et al., 1999), disparity slant JNDs increased systematically with an increase in viewing distance. JNDs also tended to decrease with base slant at the medium and far viewing distances (see also Knill & Saunders, 2003). At the near viewing distance, JNDs tended to increase with base slant. In fact, as indicated by the symbol with the yellow star, JMH's thresholds were infinite at base slants of  $\pm 70^\circ$  and a viewing distance of 19.1 cm because the binocular images could not be fused in this condition. This difference in the trend between near and far distances can be understood in terms of the retinal signal to slant.

The right column of Figure 4 plots the same data as the middle column but in units of relative disparity: specifically, the horizontal-size ratio ( $HSR$ ) (Backus et al., 1999).  $HSR = \alpha_L / \alpha_R$ , where  $\alpha_L$  and  $\alpha_R$  are the horizontal angles subtended by a surface patch in the left and right eyes. Plotted in these units, JNDs do not vary systematically as a function of viewing distance. This implies that the increase in slant-discrimination threshold is caused only by the geometric relationship between distance and disparity and not by greater error in the calculation of disparity nor by greater error in estimates used to scale for distance (such as vergence; Equation 2 in Backus et al., 1999). JNDs plotted in these units increase with increasing  $|\ln(HSR)|$ . This increase may reflect difficulties in solving the binocular-matching problem as the disparity gradient (which is linearly related to  $HSR$ ) increases (Banks, Gepshtein, & Landy, 2004; Burt & Julesz, 1980). The increase may also reflect the fact that surfaces with large  $|\ln(HSR)|$  contain fewer points near the Vieth-Müller Circle where stereoacuity is highest.  $|\ln(HSR)|$  increases more rapidly as a function of slant at near distances (indicated by the fact that JMH's data at high base  $HSRs$  all come from the near viewing distance). For example, a change in slant from 60 to 70 deg results in a change in  $\ln(HSR)$  from 0.58 to 0.97 at 19.1 cm and from 0.06 to 0.1 at 171.9 cm. (We did not plot the point at 70 deg,  $\ln(HSR) = 0.97$ , because thresholds were infinite.) We will return to a discussion of the effects of distance and base slant in Discussion: Comparison of observed and expected effects of slant and distance on disparity- and texture-based JNDs.

To make predictions for the two-cue conditions, we needed estimates of the variances of the disparity and texture estimators at slants between the ones for which we have measurements. For the interpolation, we fit smooth curves to the data ( $JND = \alpha \cdot e^{\beta|x|}$ , where  $x$  is slant and  $JND$  is in deg [texture], or  $x$  is  $\ln(HSR)$  and  $JND$  is  $|\Delta \ln(HSR)|$  [disparity], and  $\alpha$  and  $\beta$  are free parameters). This was done by performing a maximum-likelihood fit to all of the raw psychometric data for a given condition (texture or disparity), varying  $\alpha$  and  $\beta$ . The curves and the data are shown in the left and right columns of Figure 4. The curve fits represent a fit to the data at all three viewing distances. Thus, they give us a way to estimate disparity and texture reliability between slants where we have measurements, and they also allow us to interpolate across distance. While the reliability of the disparity cue to slant,  $HSR$ , does not vary systematically with distance, the relationship between  $HSR$  and slant varies significantly with viewing distance. Figure 5 shows how the reliability of disparity slant estimates varies with slant and distance, based on the curve fits to JMH's data. The reliability of the disparity cue to slant decreases as distance increases and the reliability of the disparity cue varies with base slant in different ways at different distances. At near distances the disparity cue is more reliable than the texture cue (and hence, should be given more weight according to the MLE model).

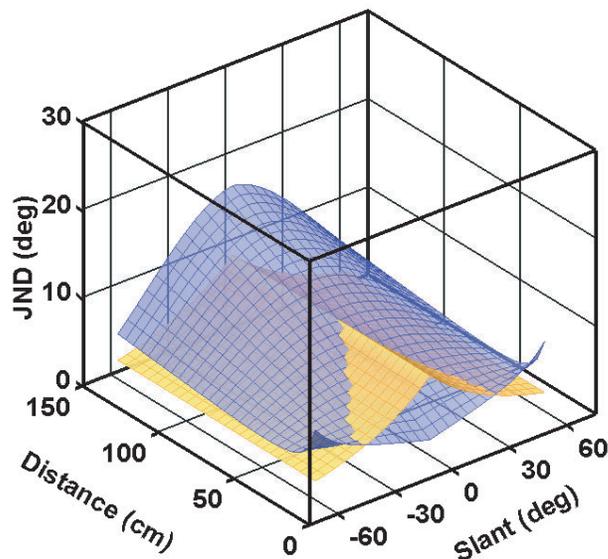


Figure 5. JNDs of texture (orange) and disparity (blue) cues across distance and slant estimated from curve fits to JMH's single-cue data (Figure 4).

Given a pair of JND values for texture and disparity, we can use Equation 2 to calculate optimal weights. Predicted weights determined solely from the standard deviations of cumulative Gaussians fitted to single-cue psychometric data are shown in Figure 6 as data points (based on the raw JND data) and curves (based on the fitted curves in Figure 4).

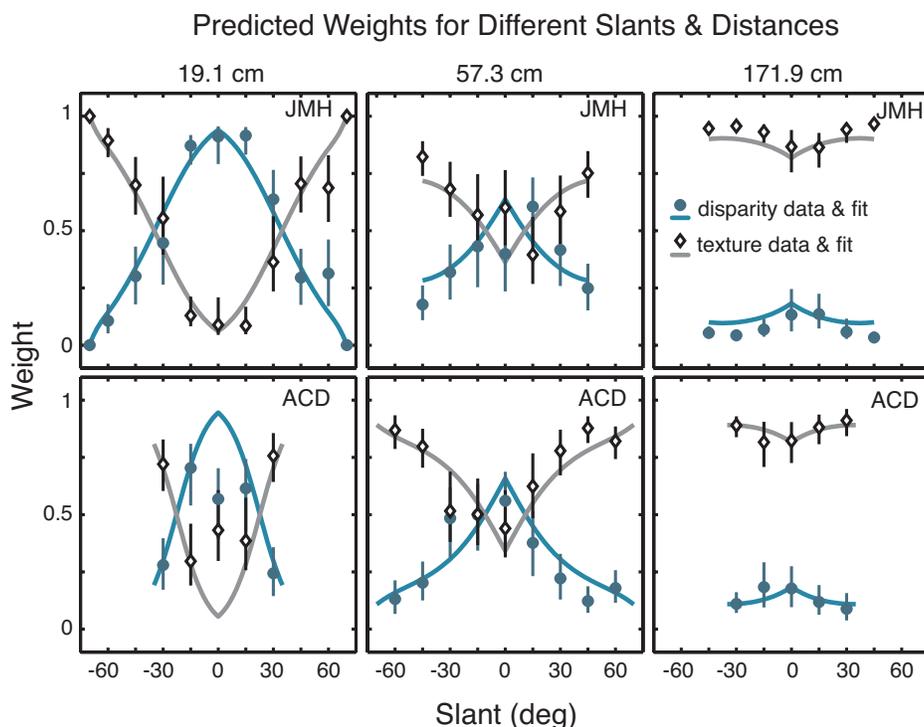


Figure 6. Predicted weights for disparity and texture cues. From left to right, the panels show data from viewing distances of 19.1, 57.3, and 171.9 cm. The weights were calculated using Equation 2 and the single-cue discrimination data and curve fits shown in Figure 4. Unfilled diamonds are predicted weights for the texture cue and filled circles are predicted weights for the disparity cue. The solid lines are predictions calculated from the curve fits in Figure 4. Error bars are 95% confidence intervals.

(Similar plots for the other two observers are shown in Figure 2S. The filled circles and blue curves are the predicted disparity weights and the unfilled diamonds and gray curves are the predicted texture weights. Because JMH's thresholds were infinite at 70 deg in the disparity-alone condition at 19.1 cm, the MLE weight given to disparity in this condition is 0. The curve used to fit JMH's disparity data in Figure 4 does not capture this fact. To incorporate this fact, we smoothly extrapolated the predicted weights curve so that the disparity weight reached 0 at 70 deg.

The predicted weights exhibit two trends. First, with increasing slant, the texture weight increases and the disparity weight decreases (data from RM and MSB showed the same trend; Figure 2S). The reciprocal relationship between the texture and disparity weights occurs because the weights are constrained to sum to 1. The texture weight becomes relatively greater than the disparity weight with increasing slant because it becomes a relatively more reliable estimate (Figure 4). Knill and Saunders (2003) observed a similar effect. Second, with increasing distance, disparity weight decreases (and texture weight increases). Although the reliability of the texture estimator does not change with distance, its relative reliability increases because the reliability of the disparity estimator decreases. Individual differences in disparity and texture estimators (i.e., the single-cue data, Figure 4) are manifest in their predicted weights, a point we will discuss later.

### Points of subjective equality (PSEs)

From the two-cue data, we can derive the weights the observers actually gave the disparity- and texture-specified slants. Figures 3 and 7 illustrate how this was done. The left panel of Figure 7 shows one observer's psychometric data for a base slant of 0 deg and viewing distance of 57.3 cm. It plots the proportion of trials on which the observer indicated that the no-conflict stimulus appeared to have greater slant (right side farther away) than the conflict stimulus. Psychometric data from four cue-conflict conditions are shown. Unfilled diamonds represent data for which the disparity-specified slant was 0 deg and the texture-specified slant was -10 (gray) or +10 deg (black). Filled circles represent data when the texture slant was 0 deg and the disparity slant was -10 (light blue) or +10 deg (dark blue). It is readily apparent that the texture and disparity cues both affected perceived slant because perturbing the texture-specified slant affected judgments (shown by the separation between the gray and black diamonds) and perturbing the disparity-specified slant also affected judgments (the separation between the light and dark blue circles). The effect of disparity perturbation was greater than the effect of texture perturbation, so the weight given to disparity was larger in this condition. PSEs, the no-conflict stimulus values that appeared on average to have the same slant as the conflict stimuli, are indicated by the arrows. The right panel of

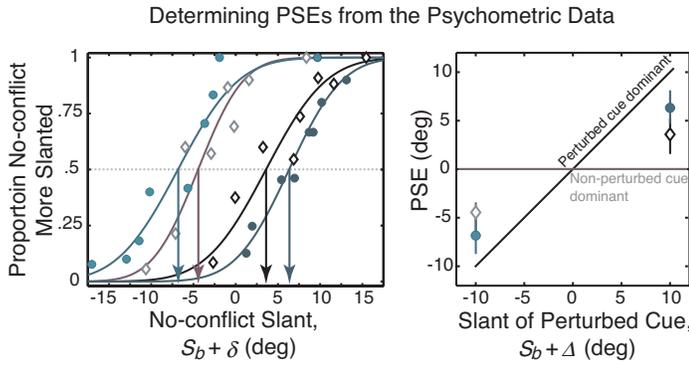


Figure 7. Determination of points of subjective equality (PSEs) from two-cue data. Left panel: one observer's results for four cue-conflict stimuli with  $S_b = 0$  deg,  $\Delta = +/-10$  deg, and distance = 57.3 cm. The conflict stimuli are:  $S_d = 0, S_t = -10$  deg (unfilled gray diamonds),  $S_d = 0, S_t = 10$  deg (unfilled black diamonds),  $S_d = -10, S_t = 0$  deg (filled light-blue circles), and  $S_d = 10, S_t = 0$  deg (filled dark-blue circles). Data represent the proportion of times the observer indicated the no-conflict stimulus was more slanted than the conflict stimulus. Staircase data with fewer than four observations at a given value of the no-conflict stimulus have been removed for clarity. Curves are maximum-likelihood fits of cumulative Gaussians (which used all the points including the ones removed for the clarity). The means of the fits are PSEs, the value of the no-conflict stimulus that on average had the same apparent slant as the conflict stimulus. The PSEs for each of the four conflict stimuli are indicated by the arrows. Right panel: PSEs for the four psychometric functions. Values of the no-conflict stimulus (indicated by arrows in left panel) are plotted as a function of the conflict  $\Delta$ . If perceived slant were determined by one cue only (meaning its weight = 1), the data would lie on the diagonal line labeled "Perturbed cue dominant" when that cue was perturbed and on the horizontal line labeled "Non-perturbed cue dominant" when the other cue was perturbed. Error bars are 95% confidence intervals.

Figure 7 illustrates how those PSEs were used to determine the empirical weights. If the perturbed cue (texture for the diamonds and disparity for the circles) were the sole determinant of perceived slant (meaning that its weight equaled 1; Equation 1), the PSEs would lie along the diagonal line. If the non-perturbed cue were the sole determinant, the PSEs would fall on the horizontal line. The relative location of the PSE data between these two extremes reflects the weight given to the perturbed cue (Equation 9). In the same format as Figure 7, Figures 8-10 compare PSE data from the two-cue conditions (reflecting the weights observers actually gave to the two cues) with MLE predictions based on the single-cue data (Equation 1).

Figure 8 shows the data from observer JMH and Figure 9 the data from ACD. The columns of panels show data, from left to right, for viewing distances of 19.1, 57.3, and 171.9 cm. The rows of panels show data, from top to bottom, for base slants of +60, +30, 0, -30, and -60 deg

(indicated by orange numbers). The abscissa in each panel is the value of the perturbed cue's slant in the conflict stimulus and the ordinate is the PSE. Figure 10 shows data for RM and MSB at the 57.3-cm viewing distance. Here the columns of panels correspond to different observers and the rows are the same as in Figures 7 and 8.

The blue and gray lines are MLE predictions for the disparity-perturbed and texture-perturbed conditions, respectively. For each conflict stimulus, the reliability for each cue was computed based on the fitted curves in Figure 4. The optimal weights were then computed using Equation 2. These weights, together with the displayed slants for each cue, were combined using Equation 1 to predict the PSE (i.e., the perceived slant for the conflict stimulus). The predictions are curved because the relative reliabilities (and hence the cue weights) change as the per-

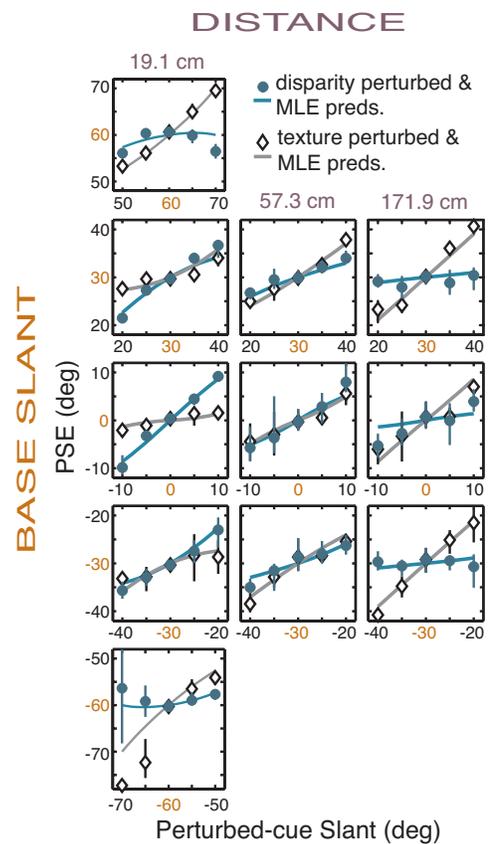


Figure 8. PSE data and predictions for observer JMH. PSE (slant of the no-conflict stimulus perceived on average as the same as conflict stimulus;  $S_b + \delta$ ) is plotted as a function of the value of the perturbed cue in the conflict stimulus ( $S_b + \Delta$ ). The left, middle, and right columns are data from viewing distances of 19.1, 57.3, and 171.9 cm. The rows are for base slants ( $S_b$ ) of -60, -30, 0, 30, and 60 deg. Those base slants are the middle abscissa value in each panel. Blue filled circles are PSEs when the disparity cue was perturbed and black unfilled diamonds are PSEs when the texture cue was perturbed. Blue and gray lines are the predictions based on Equations 1-2 and the curve fits in Figure 4. Error bars are 95% confidence intervals.

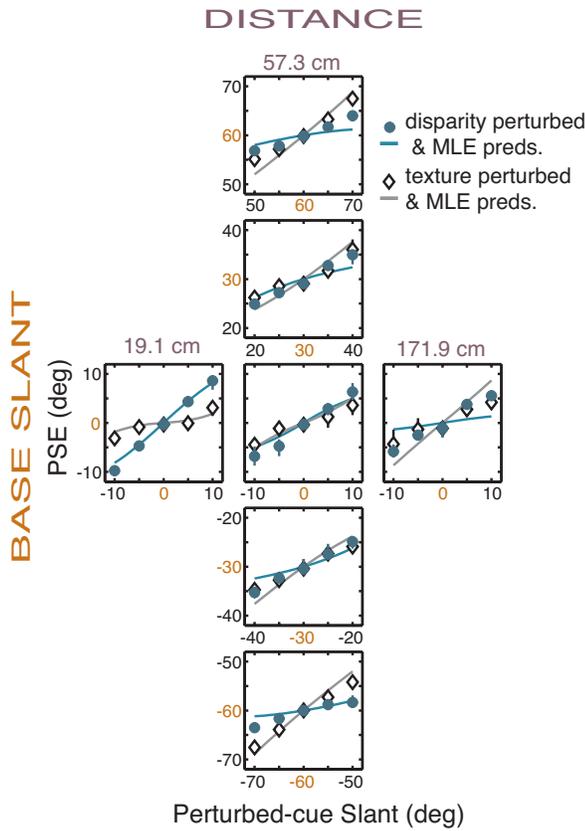


Figure 9. PSE data and predictions for observer ACD. Conventions the same as Figure 8.

turbation is changed (Hillis et al., 2002). We used a shortcut to generate the prediction curves. Specifically, we used the reliability based on the displayed slant to calculate the weight, rather than the reliability based on the observer's estimate of slant from each cue (which varies from trial to trial). Predictions based on a full Monte Carlo simulation in which weights were calculated separately for each simulated trial were, however, indistinguishable from these.

The agreement between the PSE data and predictions is generally excellent. The two main expected trends are observed in the data: The influence of disparity decreases with increasing distance and with increasing slant. We will discuss exceptions to the close agreement in Discussion: Summary of the results. We also plotted the MLE-predicted and actual weights in a similar format to Figures 8-10. These plots are shown in Figures 3S-5S. These plots show that the weights are generally close to the MLE-predicted weights and that the sums of the weights given to texture and disparity do not differ from one.

**Just-noticeable differences (JNDs)**

The estimation model for cues with uncorrelated noises (Equations 1-2) produces the least-variable estimate of slant given the available cues. If observers employ this cue-combination scheme, we should see improvements in JNDs when both cues are available compared to when only

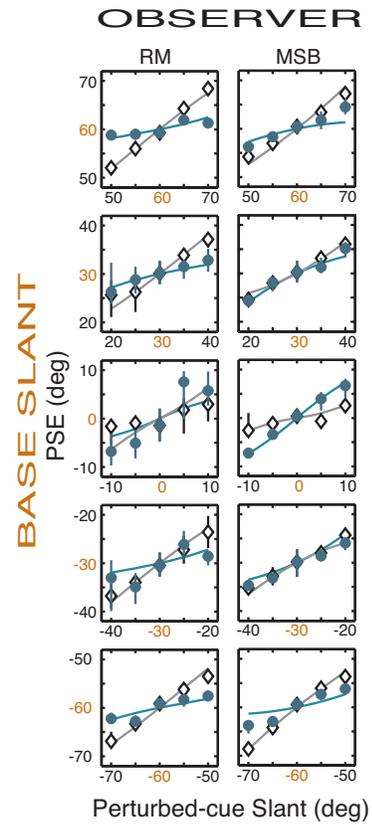


Figure 10. PSE data and predictions for RM and MSB at the 57.3-cm viewing distance. Symbol conventions the same as in Figures 8 and 9.

one cue is available. Equation 6 specifies the variance of the optimal cue-combined estimator, which is lower than either of the single-cue estimators. We used the estimates of JNDs from the single-cue conditions (Figures 4 and 1S) and Equation 6 to calculate the predicted JNDs when both cues were available. Figure 11 shows measured and predicted JNDs for JMH and ACD as a function of base slant for the three distances. The pale symbols represent the single-cue JNDs: diamonds for texture alone and circles for disparity alone. The filled red squares are the observed two-cue JNDs and the shaded red areas contain the 95% confidence intervals for the predictions. With few exceptions (discussed in Summary of results), the two-cue data follow the predictions very closely. Importantly, two-cue JNDs are consistently lower than single-cue JNDs, which shows that the visual system does benefit from having both cues available. Similar JND plots for RM and MSB are shown in Figure 6S.

Earlier we mentioned a test of the assumption that the reliability of the disparity estimator measured in the single-cue experiment with random-dot stimuli is a valid estimate of the estimator's reliability in the two-cue experiment with Voronoi stimuli. We tested the assumption by examining situations in the two-cue experiment in which the texture weight was nearly zero. The texture weight was less than 0.15 in three situations, all with observer JMH: dis-

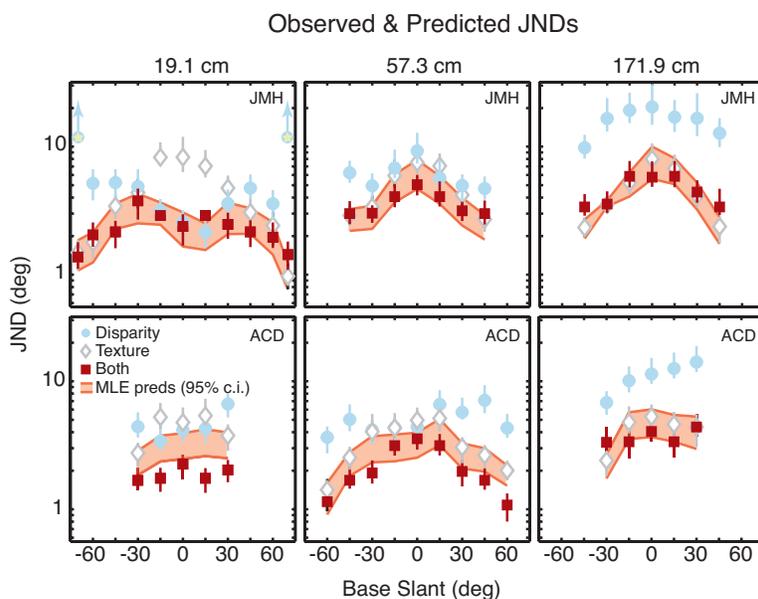


Figure 11. Predicted and observed JNDs. The just-noticeable difference in slant (JND) is plotted as a function of base slant  $S_b$ . JNDs are the sigma parameters for the cumulative normal fits to the psychometric data divided by  $\sqrt{2}$  and represent our estimates of the standard deviation of the slant estimators. Filled red squares are observed JNDs when texture and disparity were both present. Faint gray diamonds are observed JNDs for texture alone (Figure 4, left) and faint blue circles are observed JNDs for disparity alone (Figure 4, middle). Disparity JNDs for  $\pm 70$  deg base slant at 19.1 cm for JMH were infinite (indicated by pale blue symbols with yellow stars). Error bars represent 95% confidence intervals. Red curves represent 95% confidence intervals for the predicted JNDs (Equation 6). Left, middle, and right panels represent the data from viewing distances of 19.1, 57.3, and 171.9 cm.

tance = 19.1 cm and base slants of  $-15$ ,  $0$ , and  $+15$  deg. His two-cue thresholds in those situations were 2.9, 2.4, and 2.9 deg, respectively (Figure 11). His single-cue, disparity-alone thresholds in the same situations were 3.2, 2.6, and 2.1 deg, respectively (Figure 4). The close correspondence supports our assumption that the disparity-alone thresholds provided an estimate of the appropriate reliability for the two-cue experiment.

By similar reasoning, we can test the assumption that the reliability of the texture estimator measured in the single-cue experiment with monocular stimuli is a valid estimate of the estimator's reliability in the two-cue experiment with binocular stimuli. To generate one slant estimate from the texture-specified slants at the two eyes, the visual system should combine the monocular signals in some fashion. The combination could occur in two ways. (1) The visual system might combine the two eyes' images before computing slant. This could be done in principle by averaging the visual directions for each corresponding point in the two images. Then slant would be computed from the combined Cyclopean image. (2) The visual system might estimate eye-centered slants before combining. Specifically, it could estimate the slants from the texture signals received by each eye and then average the two estimates. These two means of combining the monocular images are geometrically equivalent and yield the same slant as would be observed at the Cyclopean eye as long as the coordinate origin is on the Vieth-Müller Circle. At any rate, averaging the two eyes' inputs is a reasonable way to form a texture-based slant estimate. If we assume that the two monocular inputs are

equally informative and that their noises are uncorrelated (perhaps an implausible assumption), the variance of the combined estimate would be half the variance of either monocular estimate. In other words, discrimination thresholds based on the texture information alone would be lower in the binocular than in the monocular case by  $\sqrt{2}$  (Legge, 1984). We tested this possibility by examining situations in the two-cue experiment in which the disparity weight was nearly zero. This occurred for JMH and ACD across all slants at 171.9 cm. It also occurred for observer JMH at 19.1 cm and base slant =  $\pm 70$  deg. JMH's texture-alone JNDs at 171.9 cm for base slants of  $-45$  to  $+45$  deg (the range of tested slants) were 2.3–8.0 deg (the lowest values occurring at the greatest slants; Figure 4). His two-cue JNDs at 171.9 cm for base slants of  $-45$  to  $+45$  deg ranged from 3.4–5.9 deg (again the lowest values occurring at the greatest slants; Figure 11). JMH's texture-alone JNDs at 19.1 cm for base slants of  $-70$  and  $+70$  deg were 1.5 and 1.0 deg, respectively, and his corresponding two-cue JNDs were 1.4 and 1.4 deg. ACD's texture-alone JNDs at 171.9 cm ranged from 2.4–5.3 deg and her two-cue JNDs ranged from 3.3–4.1 deg. Thus, when the disparity weight was low, the texture-alone thresholds were generally similar to the corresponding two-cue thresholds. The good correspondence supports our assumption that the texture-alone thresholds provided an estimate of the appropriate reliability for the two-cue experiment. It also implies that the slant specified by texture is not made more reliable by averaging the two eyes' images, perhaps because the noises are highly correlated.

## Discussion

### Summary of the results

The generally excellent agreement between observed and predicted PSEs and JNDs indicates that humans use a statistically optimal strategy for combining slant information from disparity and texture. There are, however, three cases in which the data deviated from the predictions.

(1) JMH's PSEs in the two-cue condition at 19.1 cm and base slants of  $-60$  and  $+60$  deg (Figure 8). The weight given disparity was lower than predicted when the absolute value of the perturbed-cue slant was greater than 60 deg. In the disparity-alone  $\pm 70$ -deg, 19.1-cm conditions, JMH could not fuse the random dot stimulus (thus, thresholds were infinite). The same was true in the two-cue condition: Slant judgments were made on diplopic images, making the task more complicated. Our model does not consider how depth judgments are made in diplopic conditions. Given this, the discrepancy between observed and two-cue data is understandable.

(2) ACD's JNDs in the two-cue condition for all base slants at 19.1 cm and for the larger base slants at 57.3 cm (Figure 11). Her two-cue thresholds were consistently lower than predicted. Moreover, ACD gave slightly more weight to disparity than predicted for base slants of  $\pm 30$  and  $\pm 60$  deg at 57.3 cm (Figure 9). The most obvious explanation for these discrepancies is that the disparity-alone JNDs (Figure 4) overestimated the variance of ACD's disparity estimator in the two-cue experiment. As described in Methods (Figure 2), ACD may have given some weight to the uninformative texture signal in the disparity-alone experiment for nonzero base slants. This would have caused an overestimate of the variance of the disparity estimator whenever the disparity weight was relatively high in the two-cue experiment (which occurs when the viewing distance is 19.1 or 57.3 cm) and whenever the base slant differed significantly from zero.

(3) JMH's and ACD's disparity weights were higher than predicted at 171.9 cm when the base slant was 0 deg. We think this small discrepancy is caused by variation in binocular fusion at long distances. Both observers reported difficulty fusing the random-dot stimulus in the single-cue experiment when the viewing distance was 171.9 cm (perhaps because of the conflict between vergence and accommodation). Thus, their thresholds at 171.9 cm may have slightly overestimated the variance of the disparity estimator at that distance. (ACD also had difficulty fusing the random-dot stimulus at 19.1 cm, which may have contributed to the apparent overestimate of the variance of the disparity estimator as discussed under #2 above.) Both observers found it easier to fuse the Voronoi stimulus at 171.9 cm, presumably because that stimulus provides contours to guide vergence eye movements. The discrepancy is most likely to show up when the base slant is 0 deg because

the disparity weight is highest in that case. Thus, this discrepancy between predicted and observed behavior is probably caused by fusion difficulties in the single-cue experiment at the long distance.

The great majority of the data is consistent with the MLE predictions and strongly supports the hypothesis that observers combine the slant cues of disparity and texture in a statistically optimal fashion.

### Comparison to other studies

Five studies have examined quantitatively whether cue combination is statistically optimal (Alais & Burr, 2004; Ernst & Banks, 2002; Gepshtein & Banks, 2003; Knill & Saunders, 2003; Landy & Kojima, 2001). In agreement with our results, all five found that combination of cues from different sensory modalities (haptics and vision: Ernst & Banks, 2002; Gepshtein & Banks, 2003; audition and vision: Alais & Burr, 2004) or different visual cues (Knill & Saunders, 2003; Landy & Kojima, 2001) was quite close to MLE predictions.

Knill and Saunders (2003) tested the MLE model for combining texture and disparity cues to surface slant. Their stimuli were slanted about a horizontal axis (tilt = 90 deg). Like us, they took advantage of the fact that the relative reliabilities of texture and disparity vary naturally with viewing geometry. They reported reasonable agreement between observed and predicted behavior. We extended their investigation by examining texture and disparity combination for surfaces slanted about a vertical axis (tilt = 0 deg) at various distances. Our data are similar and dissimilar to Knill and Saunders'. Our texture-alone data exhibited a smaller effect of base slant on JNDs (compare our Figure 4 to their Figure 6). Average texture-alone JNDs in our study were  $\sim 8$  and  $\sim 1.5$  deg at 0 and 70 deg, respectively (a ratio of 5.3). The corresponding JNDs in Knill and Saunders were  $\sim 40$  and  $\sim 2$  deg (a ratio of 20). The fact that our JNDs were generally lower is undoubtedly because our Voronoi patterns were more regular than Knill and Saunders'. The differing effect of base slant is most likely due to differences in how the angular subtense of the stimuli varied with slant: Ours varied with base slant and theirs was constant. To hold angular size constant, Knill and Saunders added texture elements as slant increased, and this adds progressively more information as slant increases.

We also examined how viewing distance affects the weights assigned to texture and disparity and found that the weight assignment is essentially optimal. This result seems to contradict numerous reports of failures to scale veridically for distance in stereoscopic tasks. For example, Johnston (1991), Johnston, Cumming, and Parker (1993), and Bradshaw, Glennerster, and Rogers (1996) had observers judge the amount of depth in disparity-defined cylinders, spheres, and ridges when presented at different distances. Responses were far from veridical, indicating that

depth was overestimated at near and underestimated at far distances. How could we observe optimal weight changes as a function of distance, while previous work showed apparent failures to take distance into account? We think the answer lies in the influence of unmodeled cues. In all but one of the previous experiments (Experiment 1 in Johnston et al., 1993), the texture gradient specified a frontoparallel plane. From our analysis, one would expect observers to report seeing less depth at long distances, not because they failed to take distance into account, but rather because they gave increasing weight to a signal specifying that the stimulus is flat. This claim is supported by the observation that making the texture gradient consistent with the disparity-specified shape generally makes judgments more veridical (Buckley & Frisby, 1993; Johnston et al., 1993). Furthermore, when the task is to adjust the shape of a surface until it appears planar and thereby consistent with the texture-specified shape, observers seem to take distance into account veridically (Rogers & Bradshaw, 1995).

If distance was not taken into account in scaling disparities, it is possible that this mis-scaling could be mistaken for a change in disparity weight with change in distance. This no-scaling hypothesis is considered and rejected in [Appendix C](#).

### Dynamic determination of cue weights

MLE cue combination has the advantage that it produces the least-variable estimate of slant given the available cues. But it requires the observer to choose weights based on the reliability of the cues. In the case of texture, the reliability clearly depends on the slant, which is what the observer is trying to estimate. Thus, the choice of weights must be made dynamically, with the possibility of varying weights from trial to trial (or from location to location within a stimulus, discussed shortly). The model suggests that on each trial the observer makes an estimate of slant from each cue, uses the value of slant for each cue along with other relevant information (“ancillary cues” such as a distance estimate; Landy, Maloney, Johnston, & Young, 1995) to determine that cue’s current reliability. The relative reliabilities are then used to determine the cue weights ([Equation 2](#)), followed by weighted cue combination ([Equation 1](#)). In our experiments, the slant shown to the observer was selected randomly before each trial from the set of possible slants within each block. For performance to approach optimality, the weights must have been determined in a trial-by-trial dynamic fashion. In a previous study we also had clear-cut evidence of weights changing from trial to trial (Hillis et al., 2002). The reader may wonder how such dynamic computation could be accomplished in a biological system without prior knowledge of the likelihood functions associated with each slant cue. Ernst and Banks (2002) outlined a plausible neural model that could carry out the computation automatically.

### Comparison of observed and expected effects of slant and distance on disparity- and texture-based JNDs

We observed three effects in the single-cue experiments—a large improvement in discrimination threshold with decreasing distance with disparity alone, a small improvement in threshold with increasing slant with disparity alone (see also Knill & Saunders, 2003), and a large improvement in discrimination threshold with increasing slant with texture alone (Knill, 1998b; Knill & Saunders, 2003). Here, we ask whether the three observed effects are expected from the slant information in the stimulus.

When the eyes are in forward gaze, as they were in these experiments, the vergence is

$$\mu = 2 \tan^{-1}(i/2d), \quad (10)$$

where  $d$  is viewing distance and  $i$  is the inter-ocular distance. Slant from disparity (for tilt = 0) is given to close approximation by

$$\hat{S}_d \approx -\tan^{-1}\left[\frac{1}{\mu} \ln(HSR)\right]. \quad (11)$$

Thus, errors in the disparity and distance estimates will both yield errors in the estimated slant. We calculated the distribution of slant estimates for different viewing conditions under the assumption that the errors in  $HSR$  and  $\mu$  can be represented by additive, independent noises. Specifically, we conducted a Monte Carlo simulation to determine the standard deviation of slant estimates  $\sigma_{\hat{S}}$  from [Equation 11](#). The noises were Gaussian with mean = 0. We adjusted the noise standard deviations,  $\sigma_{\mu}$  and  $\sigma_{HSR}$ , to obtain simulation JNDs similar to the observed JNDs. The simulation results are displayed in [Figure 12](#). The left panel shows  $\sigma_{\hat{S}}$  as a function of distance (the curves representing different base slants) and the right panel shows  $\sigma_{\hat{S}}$  as a function of base slant (the curves representing different distances).

The standard deviation of the slant estimate,  $\sigma_{\hat{S}}$ , is roughly proportional to viewing distance for all base slants (left panel). This result is expected from [Equation 11](#) because  $d \approx i/\mu$ , so fixed additive noise in  $\mu$  has an increasing effect with distance. We found that  $\sigma_{\hat{S}}$  was proportional to distance for a wide range of  $\sigma_{\mu}$  and  $\sigma_{HSR}$ ; the key assumption is that the noise in disparity normalization is fixed and additive in vergence. The data points in the lower left panel are JNDs from observer JMH; clearly, his discrimination thresholds increased monotonically with increasing distance in much the same way as the simulation. The data from ACD were similar. Thus, the distance effect we observed in the disparity-alone experiment is expected if error in disparity normalization is additive in units of vergence.

The right panel of [Figure 12](#) shows that  $\sigma_{\hat{S}}$  is inversely related to the absolute value of slant. This relationship was observed for all values except when  $\sigma_{\mu} \gg \sigma_{HSR}$ . The

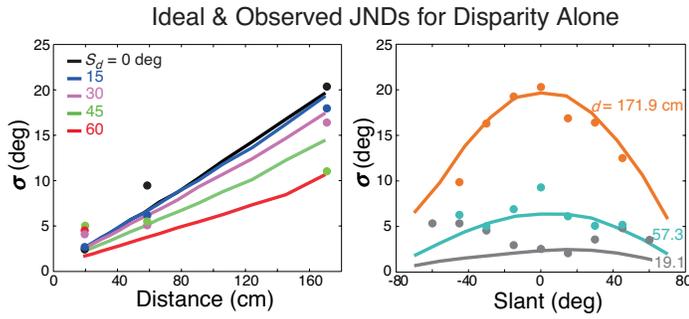


Figure 12. Results of a simulation of slant from disparity estimation. We used a Monte Carlo simulation to calculate the standard deviation of disparity-based slant estimates (Equation 11) for different viewing conditions. We assumed that error in the slant estimates stemmed from noise in *HSR* and  $\mu$  (vergence angle) and that these errors (the variances) were the same for all viewing distances. The noises were additive and Gaussian with mean = 0, and we obtained simulation results for many sets of parameters. The results for  $\sigma_{HSR} = 0.012$  and  $\sigma_{\mu} = 0.012$  radians, which fit the data reasonably well, are displayed in the figure. Left panel: the standard deviations of slant estimates are plotted as a function of distance. Different curves represent different absolute values of base slant. The circles represent the observed JNDs for observer JMH at the various distances. Different colors represent different absolute values of base slant. Right panel: the standard deviations of slant estimates as a function of base slant. Different curves represent different viewing distances. The circles represent the observed JNDs for observer JMH. Different colors represent different distances.

relationship is expected from Equation 11 because  $\hat{S}_d \approx k \ln(HSR)$ , so fixed additive noise in *HSR* has progressively less effect on  $\sigma_{\hat{S}_d}$  as base slant increases. The data points in the lower right panel are JNDs from observer JMH; data were similar for the other three observers. At viewing distances of 57.3 and 171.9 cm, JMH's discrimination thresholds decreased monotonically with slant magnitude much like the simulation's standard deviations.

Thus, the base-slant effect we observed in the disparity-alone condition is expected if error in disparity measurement is additive in *HSR*. Does this assumption make sense? It does when *HSR* is not significantly different from 1, which was true for distances of 57.3 and 171.9 cm (see Figure 4). However, when *HSR* is quite different from 1, points on the surface fall where stereo-acuity is low and problems arise in solving the binocular correspondence problem (Burt & Julesz, 1980). *HSR* and the horizontal gradient of horizontal disparity are closely related,

$$DG \approx 2(1 - HSR)/(1 + HSR), \tag{12}$$

where *DG* is an approximation to the disparity gradient (Howard & Rogers, 2002). From Equations 10 and 11 when *d* is small and *S* is large, *HSR* is quite different from 1 and thus *DG* will be quite different from 0. Burt and Julesz (1980) and others have shown that binocular correspon-

dence becomes difficult when  $|DG|$  deviates significantly from 0 and breaks down altogether when  $|DG| \approx 1$ . Recent results indicate that this is probably a by-product of a matching process that is similar to cross-correlating the two eyes' images to estimate the disparity in a region of the visual field (Banks et al., 2004).

Figure 13 plots the disparity gradient as a function of slant for the three distances we used.  $|DG|$  increases rapidly as a function of slant at the short distance, so we expect performance to be worse at that distance for large slants. JMH's data exhibited this effect. His discrimination thresholds at 19.1 cm increased with slant, which is inconsistent with the assumption that the sole source of error in disparity measurement is additive in *HSR* (Figure 12, lower-right panel, gray curve). They were higher than predicted for  $|slant| \geq 30$  deg which corresponds to a higher disparity gradient ( $|DG| \geq 0.19$ ,  $HSR \geq 1.2$ ) than occurs at 57.3 and 171.9 cm. Thus, the base-slant effect in the disparity-alone experiment is expected if error in disparity measurement is additive in *HSR* except when *HSR* deviates significantly from 1 where problems arise in solving correspondence.

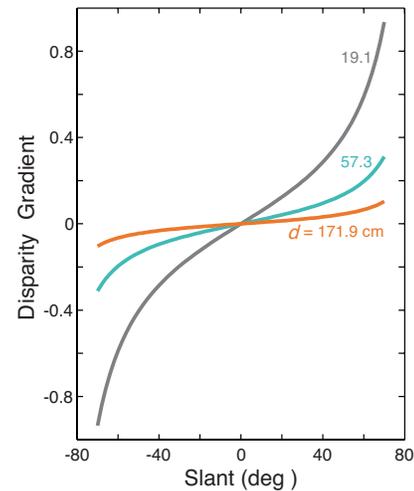


Figure 13. Disparity gradient as a function of slant at different viewing distances. The disparity gradient was calculated from Equation 12.

We also compared our observed texture-alone thresholds with those expected from the information in the various slant cues associated with the texture gradient. Knill (1998a, 1998b) described ideal observers for slant from texture when presented Voronoi stimuli like the stimuli in our experiments. The stimulus parameters in our experiment differed from those in his modeling and experiments in two ways.

First, the Voronoi patterns in our stimuli were more regular than in his. From this one would expect the texture-gradient cue to be more reliable in our experiment than in Knill's.

Second, the angular subtense of our stimuli varied with slant (even though there was a random element to the an-

gular width so as to make the width an unreliable cue to slant), so the average number of texture elements was constant across slant. To keep the angular subtense constant, Knill added texture elements as slant increased, which adds information. This added information probably explains why his observed and predicted discrimination thresholds varied more with slant than ours did.

Despite the differences in stimulus parameters, it is informative to compare ideal thresholds with our observers' thresholds. The curve in Figure 14 shows the standard deviation of slant estimates from Knill's foreshortening ideal observer (Figure 5A in Knill, 1998a). There is a striking effect of base slant. The data points are JMH's thresholds in the texture-alone experiment; data were similar for the other three observers. The data exhibit a base-slant effect like the ideal observer's, but the effect is smaller in our data for reasons described above. Therefore, the variation we observed in texture-based slant thresholds is by and large expected from the information content of the stimulus.

We conclude that the effects of distance and slant on JNDs can be expected from the information present in the stimuli. These effects are summarized in Figure 5.

## What other variables might affect cue weights?

Presumably, the visual system takes the disparity and texture variances into account across many viewing situations. To do so, however, is complex because many viewing properties will affect the likelihood functions associated with disparity and texture cues. Here we list the most obvious properties and suggest how the relative weights assigned to disparity and texture ought to be affected.

### 1. Regularity of texture.

The slant information contained in the texture gradient can be divided into three cues: (1) scaling, the change in the projected sizes of texture elements, (2) foreshortening, the change in projected shapes of texture elements, and (3) density, the change in the number of elements per unit area in the projection (Blake et al., 1993; Cutting & Millard, 1984; Knill, 1998a). The reliability of scaling as a slant signal depends on the variation in the sizes of the texture elements on the surface. With greater size variation, the cue's reliability decreases (Knill, 1998b). The reliability of foreshortening depends on the variation in the shapes of the elements on the surface. For regular shapes, like circles, reliability is greater than for irregular shapes, such as ellipses with variable aspect ratios (Knill, 1998b; Young et al., 1993). The reliability of the density cue depends on the number of elements and the regularity of their positioning on the surface. Presumably, many elements placed regularly (i.e., in a grid) yield more reliable estimates than few elements placed randomly. All three cues are affected by the field of view, particularly in the tilt direction, so slant discrimination from texture is more precise with large than with small stimuli (Blake et al., 1993; Knill, 1998b). If the

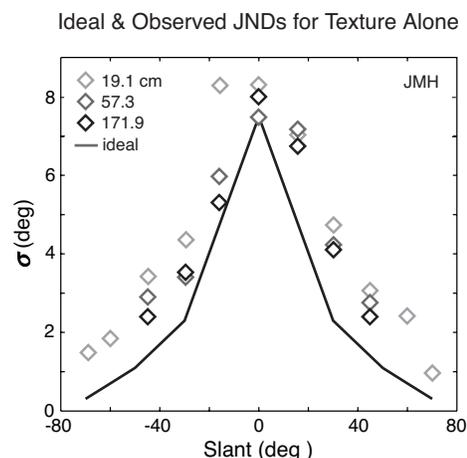


Figure 14. Ideal and measured JNDs for slant from texture as a function of base slant. The solid line represents the standard deviations of the slant estimates of the foreshortening ideal observer for Voronoi stimuli (Figure 5A in Knill, 1998a). The diamonds represent discrimination thresholds in the texture-alone condition for observer JMH. Light gray diamonds are thresholds at 19.1 cm, medium gray at 57.3 cm, and dark at 171.9 cm.

visual system takes the varying reliability of the texture gradient cue into account, all of these stimulus properties will affect the relative weights assigned to disparity- and texture-based signals.

### 2. Surface tilt.

The direction of slant or tilt affects the amount of perceived slant in stereograms (Howard & Rogers, 2002). The disparity signal for surfaces slanted about a vertical axis (tilt = 0 deg) is the horizontal gradient of horizontal disparities. We have quantified this as the horizontal-size ratio (HSR). The disparity signal for surfaces slanted about a horizontal axis (tilt = 90 deg) is the vertical gradient of horizontal disparities. This disparity pattern is often referred to as horizontal-shear disparity (Banks et al., 2001). Random-element stereograms simulating a slanted plane with tilt = 0 deg generally produce less perceived slant than planes with tilt = 90 deg (Gillam & Ryan, 1992). Similarly, the amount of depth seen in curved disparity-defined surfaces varies with tilt (Buckley & Frisby, 1993). These tilt-dependent variations in perceived depth are called *slant anisotropy*. The phenomenon is most striking when the texture gradient specifies a frontoparallel plane, as is usually the case with random-element stereograms. The phenomenon is not observed when disparity and texture signal the same depth variation, as occurs with real surfaces (Bradshaw, Hibbard, van der Willigen, Watt, & Simpson, 2002; Buckley & Frisby, 1993). These observations strongly suggest that slant anisotropy is caused by conflicting disparity and texture signals in conventional random-element stereograms. They also suggest that texture is generally given more weight for tilt 0 (as in our experiments) and less weight for tilt 90 (as in Knill & Saunders, 2003). By the

argument presented here, this may be due to reduced disparity reliability for tilt 0 than for tilt 90 because there is no obvious reason for the reliability of the monocular texture cue to depend on tilt. There may, however, be differences in the steps required to combine the texture and disparity signals for different tilts. The issues involved in transforming texture-gradient signals into the same coordinates for combination with disparity signals are taken up in [Appendix D](#).

### 3. Reliability of estimated distance and azimuth.

To estimate slant from the measured disparities, the visual system must “normalize” the disparities with a distance estimate and “correct” the disparities with an azimuth estimate (Gårding et al., 1995). Relaxing the assumption of forward gaze in [Equation 11](#), slant about a vertical axis (tilt = 0) is

$$S \approx -\arctan\left(\frac{1}{\mu} \ln HSR - \tan \gamma\right), \quad (13)$$

where  $\mu$  is vergence, and  $\gamma$  is azimuth (the angle between the head’s median plane and the Cyclopean line of sight) (Backus et al., 1999).  $\mu$  is estimated both from extra-retinal signals concerning the eyes’ vergence and from the horizontal gradient of vertical disparity (Rogers & Bradshaw, 1995). When vertical disparities are large, as occurs with large stimuli at close range, they are the predominant means for estimating distance. However, when vertical disparities are unreliable because the stimulus is small (Rogers & Bradshaw, 1995), or because the texture contains no horizontal contours (Helmholtz, 1910), the eyes’ vergence becomes the predominant means of estimating distance and the accuracy of disparity normalization drops (Rogers & Bradshaw, 1995).

The azimuth  $\gamma$  is used to correct disparities; it is estimated from extra-retinal, eye-position signals and from the magnitude of vertical disparities (Backus et al., 1999). When vertical disparities are large, as occurs with near stimuli subtending a large angle, they are the predominant means of estimating azimuth. When the stimulus is short or when vertical disparities are unmeasurable, eye position becomes the predominant means and the accuracy of disparity correction suffers (Backus et al., 1999).

Similar arguments apply for slant estimation with tilt = 90 deg. In this case, slant around a surface point is

$$S \approx -\arctan\left(\frac{1}{\mu} \ln HSh - \tau\right), \quad (14)$$

where  $\mu$  is again the vergence angle,  $HSh$  is horizontal shear disparity (Banks et al., 2001) and  $\tau$  is the cyclovergence of the eyes (the difference in the eyes’ torsion).  $HSh$  must be normalized for distance by an estimate of  $\mu$  and corrected for cyclovergence by an estimate of  $\tau$  (Banks et al., 2001; Howard & Kaneko, 1994).

For our present purposes, when the viewing situation reduces the reliability of the estimates of the normalizing and/or correcting signals, the disparity estimate will become more variable. This will occur, for example, when the stimulus subtends a small angle, when the surface markings make the measurement of vertical disparity unreliable, and when the stimulus is distant. If the visual system takes such changes into account, the weight given to disparity should decrease in those circumstances.

### 4. Duration.

Van Ee and Erkelens (1998) showed that the slant perceived from disparity-defined planes increases with stimulus duration. Their random-element stereograms contained the texture gradient associated with a frontoparallel plane, so their results are consistent with a model in which the weight given to disparity relative to texture increases over time. Presumably, disparity and texture estimates both become more precise with increases in stimulus duration, but the increase may be slower for disparity. Thus, stimulus duration may also affect the relative weights given to disparity- and texture-based slant estimates.

### Are cue weights computed locally?

It is interesting to consider whether the visual system determines one set of weights for each surface or whether the weights are calculated locally. That is, can the weights vary from one patch on a surface to another? If they are calculated locally, there are situations in which a cue-conflict stimulus specifying a plane should appear curved. Here we explain why this should happen and report that the predicted curvature is in fact observed.

The left panel of [Figure 15](#) shows how slant and distance vary with azimuth when the surface is a plane. For the part of the plane that lies straight ahead, the slant is  $S$  and the distance is  $d$ ; for the part on the right, it is

$$S_\gamma = S - \gamma, \quad (15)$$

where  $\gamma$  is the azimuth. The distance to the intersection of the line and plane is

$$d_\gamma = \frac{d}{\cos \gamma + \sin \gamma \tan S}. \quad (16)$$

The left and middle panels of [Figure 16](#) show how  $S_\gamma$  and  $d_\gamma$  vary with azimuth for different base slants and  $d = 19.1$  cm. Because the local slant and distance vary with azimuth, the statistically optimal weights for the texture and disparity cues should vary with azimuth.

Now consider the cue-conflict stimulus in the middle panel of [Figure 15](#). For rightward gaze ( $\gamma < 0$ ), slants  $S_{d\gamma}$  and  $S_\gamma$  approach zero and distance  $d_\gamma$  decreases. Our data ([Figure 6](#)) show that texture weight is relatively low when the absolute value of slant is  $\sim 0$  and distance is short. Thus, if the weights used in combining slant estimates are determined locally, one would expect the texture weight in this situation to be lower on the right than straight ahead. (The

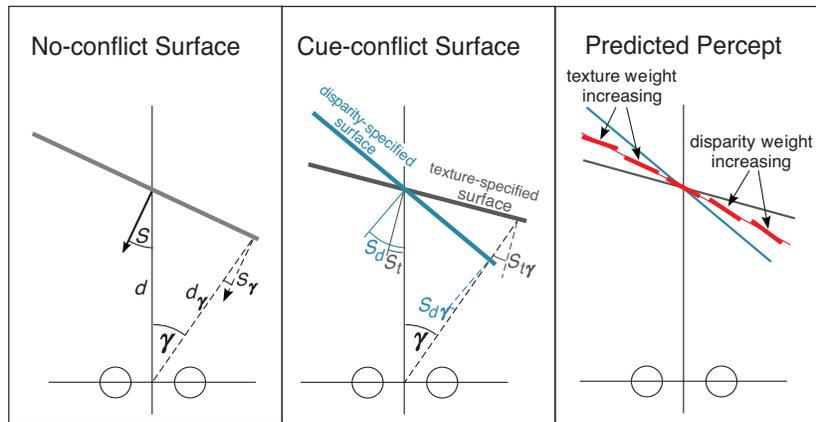


Figure 15. Change in slant with azimuth. Left panel: Definitions of local slant and distance. A plane is positioned straight ahead of the Cyclopean eye. The surface normal at the intersection of the plane and the head’s median plane is shown. Slant  $S$  is the angle between the normal and the Cyclopean line of sight to that point and  $d$  is the distance to that point. The azimuth to another point on the plane is  $\gamma$ . The slant with respect to that point is  $S_\gamma$  and the distance is  $d_\gamma$ . Middle panel: Slants and local slants for a cue-conflict stimulus. The disparity-specified slant of the stimulus is  $S_d = -25$  deg and the texture-specified slant is  $S_t = -10$  deg. The local slants at azimuth  $\gamma$  are  $S_{d\gamma}$  and  $S_{t\gamma}$ . Right panel: Predicted perceived slants across the cue-conflict stimulus if cue weights are determined locally. The thin red line represents the predicted slant if cue weights are fixed across the surface. The thick red line segments represent the predicted changes in slant if the weights are determined locally. The surface should appear concave in this case.

changes in local slant and distance with changes in azimuth are unaffected by the direction in which the eyes are looking; they are determined only by the positions of surface points relative to the head. Thus, when we say “on the right” or “straight ahead,” we refer to the head-centered azimuth of a line of sight from the Cyclopean eye and not necessarily the azimuth of fixation.) For leftward azimuth ( $\gamma > 0$ ), the slants become increasingly negative and distance increases; the texture weight in this situation should be higher on the left than straight ahead.

If the disparity and texture weights change across a surface, the slant estimate should change when the disparity- and texture-specified slants differ. This is illustrated in the right panel of Figure 15. The thin red line represents the slant estimate if the disparity and texture weights were equal throughout. However, if the disparity weight increased with increasingly rightward azimuth, the slant estimate should approach the disparity-specified slant toward the right and the texture-specified slant toward the left; this is depicted by the thick red line segments. As a conse-

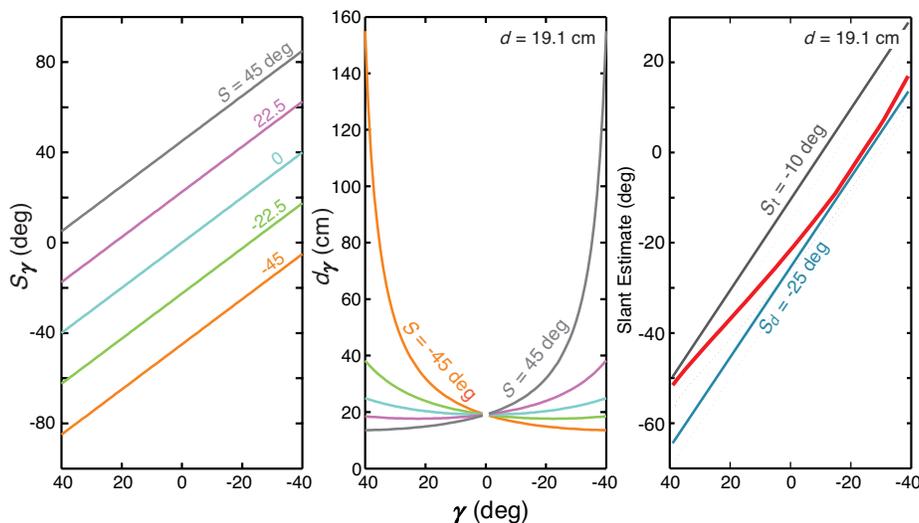


Figure 16. Change in local slant, local distance, and estimated slant with azimuth. Left panel: Local slant  $S_\gamma$  (see Figure 15 for definition) as a function of azimuth  $\gamma$ . The lines from top to bottom show  $S_\gamma$  for different slants  $S$ . Middle panel: Local distance  $d_\gamma$  as a function of azimuth  $\gamma$ . The gray, magenta, blue, green, and orange curves show  $d_\gamma$  for slants of 45, 22.5, 0, -22.5, and -45 deg, respectively. Right panel: Predicted local slant estimates from the weighted sum (Equation 1) if the cue weights are determined locally. Disparity-specified slant ( $S_d$ ; blue) is -25 deg and texture-specified slant ( $S_t$ ; gray) is -10 deg. Distance ( $d$ ) is 19.1 cm. Using the cue weights as a function of base slant at a distance of 19.1 cm for observer JMH, we calculated the estimated slant at each azimuth. Those local slants ( $S_\gamma$ ) are represented by the red curve.

quence, this particular cue-conflict stimulus should appear concave.

We calculated how the estimated local slant  $S_\gamma$  should change with azimuth for observer JMH. The cue-conflict stimulus in the calculation had a disparity-specified slant  $S_d$  of  $-25$  deg and texture-specified slant  $S_t$  of  $-10$  deg. Distance  $d$  to the midpoint was 19.1 cm. If weights are determined locally for each azimuth  $\gamma$ , the observer must associate local disparity-defined slant  $S_{d\gamma}$  with its corresponding variance  $\sigma_{d\gamma}^2$  and the texture-defined  $S_{t\gamma}$  with its variance  $\sigma_{t\gamma}^2$ . The right panel of Figure 16 shows the expected change in slant as a function of azimuth. With leftward azimuth, the slant estimate approaches the texture-specified slant, and with rightward azimuth, it approaches the disparity-specified slant. The result would be an apparently concave surface as schematized in the right panel of Figure 15. If the disparity- and texture-specified surfaces were swapped ( $S_d = -10$ ;  $S_t = -25$  deg), the result would be an apparently convex surface. If the disparity and texture specified the same slant, as they would with most real surfaces, the result should be an apparently planar surface.

In doing these calculations, we assumed that the cue weights were determined by the local disparity and texture slants only and not by the distance (because the texture-specified distance is undefined). The predicted curvature would have been somewhat greater if we had included the changes in disparity-defined distance in the calculation.

Do people see the predicted curvature? To answer this, we back-projected Voronoi patterns onto a large screen. The texture gradient specified different slants relative to the screen. The disparity-defined slant was equal to that of the projection screen (and hence was consistent with other cues such as blur and accommodation). Observers viewed the display binocularly. Azimuth was manipulated by having them stand at an oblique position relative to the screen center. The situation in Figure 15 was recreated as follows. Observers stood 25 deg to the right of center at a distance of 19.1 cm and viewed a stimulus whose texture gradient specified a slant of  $+15$  deg relative to the screen. Thus,  $\gamma = -25$ ,  $S_d = -25$ , and  $S_t = -10$  deg. At this azimuth, the field of view was  $\sim 70$  deg wide. The stimulus was clipped by an elliptical window to make the outline shape an unreliable cue to slant. The room was completely dark except for the display so the screen's frame could not be seen. If weights are set locally, this cue-conflict stimulus should appear concave. We also created a viewing situation in which the surface should appear convex— $\gamma = -25$ ,  $S_d = -25$ , and  $S_t = -40$  deg—and another in which the surface should appear planar— $\gamma = -25$ ,  $S_d = -25$ , and  $S_t = -25$  deg. Seven observers (five naïve) viewed the displays and reported whether they appeared concave, planar, or convex. Five of the seven (three naïve) reported that the stimuli predicted to look concave and convex actually looked that way; the other two said that the stimuli all appeared concave or planar but that the one predicted to look concave appeared the most concave. We asked them to order the three stim-

uli according to the amount of perceived concavity, and all seven ordered them in the predicted order.

We conclude that the weights assigned to disparity and texture are estimated locally. As a consequence, large cue-conflict stimuli in fact can appear curved. In the main experiments, observers did not notice such distortions in the cue-conflict stimuli because the stimuli were small and the conflicts were small.

## Ideal observer models

We have based our modeling on an ideal Bayesian or statistical observer assuming uncorrelated (or conditionally independent) cues and a negligible effect of the prior distribution. Here, we consider the impact and validity of these assumptions.

Oruç et al. (2003) developed an ideal observer that allowed for correlated noise associated with each cue. The resulting cue-combination rule remains linear (a weighted average), but the optimal weights do not satisfy Equation 2. Rather, they must be “corrected” for the amount of cue correlation. We observed excellent agreement between the predicted and observed two-cue data, which indicates that the correlation between the noises of the texture and disparity estimators is probably small. One might expect a strong correlation between the noises associated with disparity- and texture-based slant estimators because the two estimators must share some processing (e.g., noise in the stimulus itself, eye movements, retinal-image formation, and retinal processing). The fact that we observe (as did Knill & Saunders, 2003) the improvement in JNDs expected from combining two conditionally independent estimates implies that the dominant noises are independent. Those noises probably arise in separate processes such as comparing the two eyes' images, normalizing for distance, and correcting for azimuth.

In the modeling we assumed that the prior distribution ( $P(\hat{S})$  in Equation 3) has a negligible effect. This assumption is usually justified (e.g., Ernst & Banks, 2002) by assuming that the variance of the prior is much greater than the variances of the likelihoods. Is this the case in the estimation of surface slant? It is reasonable to assume that the distribution of surface slants in the world is uniform, particularly for tilt = 0. But if that distribution is uniform, the probability of observing slant  $S$  at the retina will be proportional to  $\cos(S)$  because steeply slanted surfaces project to smaller retinal images. Equations 7-9 show how the observers' judgments will be affected by the stimulus values and weights. If we add the prior into those equations, Equation 9 becomes

$$w_d = \frac{(1 - w_p)\delta}{\Delta}. \quad (17)$$

The value of  $w_p$  depends on its inverse variance, or reliability, relative to the estimator reliabilities (Equation 5). As long as the prior's variance is large relative to the estimators' variances,  $w_p$  will be small and will have no discerni-

ble effect on the data. The prior distribution is proportional to  $\cos(S)$ , which is defined from  $-90$  to  $90$  deg. Such a half cosine has a standard deviation of  $\sim 40$  deg. The standard deviations of the disparity and texture estimators (Figure 4) ranged from 1-20 deg. They were the highest when distance = 171.9 cm and slant = 0 deg. Then the standard deviations of the disparity and texture estimators were 20.4 and 8.0 deg for JMH and 16.1 and 5.3 deg for ACD. We can use Equation 5 to calculate the expected  $w_p$  for those conditions:  $w_p = 0.034$  and  $0.016$  for JMH and ACD, respectively. Those represent the largest possible influence of the prior distribution on the results and  $w_p$  is still quite small. Further, the weights given to texture and disparity generally sum to one (Figures 3S-5S), indicating that the prior received little or no weight in all of our conditions. We conclude that the prior had no discernible influence on our results.

**Decision noise**

An important underlying assumption in our analysis is that single-cue thresholds accurately reflect the observer’s uncertainty about slant from the cue in question. In reality, discrimination thresholds are affected by other sources of uncertainty, such as high-level decision noise. Here we examine the consequences of decision noise on the predicted and observed results. We assume that the decision noise is additive, and has a mean of zero and standard deviation of  $\sigma_n$ . We also assume that  $\sigma_n$  has the same value in the single-cue and two-cue experiments.

First consider the PSE data (Figures 8-10). Let  $\tilde{\sigma}_d^2$  and  $\tilde{\sigma}_t^2$  represent the variances we measured in the disparity- and texture-alone experiments:

$$\begin{aligned} \tilde{\sigma}_d^2 &= \sigma_d^2 + \sigma_n^2 \\ \tilde{\sigma}_t^2 &= \sigma_t^2 + \sigma_n^2 \end{aligned} \tag{18}$$

We used those measured variances to generate predictions for the weights given disparity and texture. Thus, Equation 2 becomes

$$\begin{aligned} w_{dP} &= \frac{\tilde{r}_d}{\tilde{r}_d + \tilde{r}_t} \\ w_{tP} &= \frac{\tilde{r}_t}{\tilde{r}_d + \tilde{r}_t} \end{aligned} \tag{19}$$

where  $\tilde{r}_d$  and  $\tilde{r}_t$  represent the measured reliabilities from the disparity-alone and texture-alone experiments (the measured reliabilities include the effects of decision noise). In the two-cue experiment, we measured the weight observers actually assigned to disparity and texture and those weights were presumably affected only by the visual system’s estimates of the uncertainties of the disparity and texture estimators. In other words, Equation 2 rather than Equation 19 describes what the observed weights should be. Decision noise should, therefore, affect the predicted

weights (Equation 19) and not the observed weights in the two-cue experiment. To determine the consequences of decision noise, we calculated the predicted and observed weights for a variety of situations. We set the sum of estimator variances to one ( $\sigma_d^2 + \sigma_t^2 = 1$ ) and varied  $\sigma_n^2$  from  $\sim 0$  to  $\sim 1$ .  $\sigma_n^2$  was set to 0, 0.1, 0.32, or 1. The left panel of Figure 17 shows the results. The predicted texture weight ( $w_{tP}$ ) is plotted as a function of the actual texture weight ( $w_t$ ). Naturally, the prediction (diagonal dashed line) is perfect when  $\sigma_n^2 = 0$  because Equations 2 and 19 are then identical. For  $\sigma_n^2 > 0$ , the predicted weights deviate from the observed. When  $w_t$  is greater than 0.5 (and hence greater than  $w_d$ ),  $w_{tP}$  is less than  $w_t$ . When  $w_t$  is less than 0.5, the opposite occurs. When  $w_t \approx w_d$  ( $w_t = \sim 0.5$ ), the effect of decision noise is negligible. Thus, if decision noise were sufficiently large in our experiments, it should cause error in the PSE data when  $w_t$  is either much larger or much smaller than  $w_d$ . This circumstance occurred when the viewing distance was 19.1 cm and the base slant was 0 deg and when the viewing distance was 171.9 cm (Figures 8 and 9). With the exception of distance = 171.9, base slant = 0, the agreement between predicted and observed PSEs is excellent in these cases. This implies that uncertainty due to decision noise (and other additive noises) was small relative to the uncertainty of the underlying slant estimators.

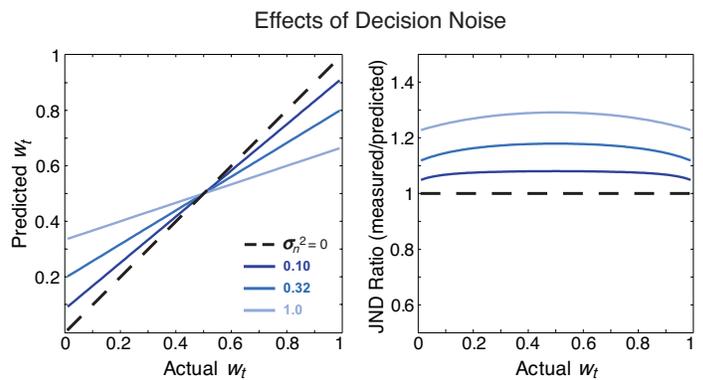


Figure 17. Results of a simulation of the effects of decision noise on PSEs and JNDs. Both graphs plot a measure of performance against the actual value of the texture weight  $w_t$ . We added decision noise to the single-cue estimates (Equation 16). The decision noise was Gaussian with mean = 0 and variances  $\sigma_n^2 = 0, 0.1, 0.32,$  and  $1.0$  where  $\sigma_d^2 + \sigma_t^2 = 1$ . Left panel: The value of  $w_t$  we would predict from the single-cue measurements (corrupted by decision noise) is plotted as a function of the actual value of  $w_t$ . The different lines represent different amounts of decision noise as indicated by the legend. Right panel: The JND ratio—the JNDs we measure in the two-cue experiment divided by the JNDs we predict from the single-cue experiments—plotted as a function of the actual value of  $w_t$ . The different curves represent the results with different amounts of decision noise as indicated by the legend in the left panel. See text for further explanation.

Knill and Saunders (2003) also considered the effects of decision noise on observed and predicted weights. In their simulation, they assumed that  $\sigma_n^2$  was equal to the variance of the combined estimate (Equation 6), so  $\sigma_n^2$  varied with  $w_t$ ; in particular, it had lower values when  $w_t = \sim 0$  or  $\sim 1$ . Thus, their simulations showed very little effect of decision noise on predicted weights. Knill and Saunders also modeled constant-variance noise like we did, but they did not show the results of that analysis.

Now consider the JND data (Figure 11). Again  $\tilde{\sigma}_d^2$  and  $\tilde{\sigma}_t^2$  represent the variances we measured in the disparity-alone and texture-alone experiments, respectively (Equation 18). Using those measured values and Equation 6, we generate a prediction for the two-cue JND (ignoring division by  $\sqrt{2}$  to convert from 2-IFC thresholds to  $\sigma$ 's):

$$\sigma_{cP} = \sqrt{\frac{\tilde{\sigma}_d^2 \tilde{\sigma}_t^2}{\tilde{\sigma}_d^2 + \tilde{\sigma}_t^2}} \tag{20}$$

Now we make the two-cue measurements in order to compare the observed and predicted JNDs. In the two-cue experiment, the visual system would weight the cues as in Equations 1 and 2, and the decision noise would again affect the threshold measurement. Thus, the JND we measure in the two-cue experiment is

$$\tilde{\sigma}_c = \sqrt{\frac{\sigma_d^2 \sigma_t^2}{\sigma_d^2 + \sigma_t^2} + \sigma_n^2} \tag{21}$$

Equations 20 and 21 are equal to one another when  $\sigma_n = 0$ , but when  $\sigma_n > 0$ , the decision noise has different effects on the predicted and observed two-cue JNDs. To determine how additive decision noise could affect the interpretation of the JNDs, we calculated the ratio of observed JND (Equation 21) divided by the predicted JND (Equation 20). In doing the calculations we again set the sum of estimator variances to 1 and varied  $\sigma_t^2$  from  $\sim 0$  to  $\sim 1$ .  $\sigma_n^2$  was again set to 0, 0.1, 0.32, or 1. The right panel of Figure 17 shows the results. The JND ratio is plotted as a function of the texture weight. The dashed horizontal line represents the ratio when  $\sigma_n^2 = 0$ . As  $\sigma_n^2$  increases from 0 to 1, the observed JND becomes larger than the predicted. The ratio is largest at  $\sim 1.3$  when  $w_t = 0.5$  and  $\sigma_n^2 = 1$ . The JNDs we observed in the two-cue experiment were generally quite close to the predicted values (Figure 11), so this analysis suggests that we can rule out the presence of decision noise whose variance is greater than approximately half the sum of the estimator variances.

We conclude that uncertainty due to additive, high-level decision noise was small in comparison to the uncertainties of the underlying slants from the two cues.

## Conclusion

We performed two quantitative tests of a maximum-likelihood estimator model for combining the slant cues of

texture and disparity. Our results indicate that the visual system combines texture and disparity information in a statistically optimal fashion, thereby reducing the variance of slant estimates that could be achieved otherwise. To do this, the relative reliabilities of each cue have to be determined dynamically, on a trial-by-trial basis, suggesting the presence of well-developed circuits for combining depth-cue information. The success of the MLE model in this and other studies indicates that perceptual systems have gone to some trouble to incorporate all available information into the estimation process, and that they have done so in a manner that maximizes the precision of perceptual estimates.

## Appendix A: Dot number for disparity-alone experiment

We wanted to make sure that we presented enough dots in the display for disparity-based thresholds to be as low as possible while still isolating the disparity estimator. For this purpose, we measured slant-discrimination thresholds as a function of the number of dots. We ran three conditions: (1) texture-specified slant = 0 deg and disparity-specified slant = 0 deg +  $\delta$  (where  $\delta$  is the increment or decrement given to the base slant in order to obtain a threshold), (2) texture slant = 0 and disparity-specified slant = 45 +  $\delta$ , and (3) texture slant = 45 and disparity-specified slant = 45 +  $\delta$ . The results are plotted in Figure A1, which shows the just-discriminable change in slant as a function of dot number for the three conditions. For observer JMH, discrimination thresholds decreased as dot number was increased from 2 to 32 and then thresholds reached an asymptote by 32-64 dots. With 64 dots,

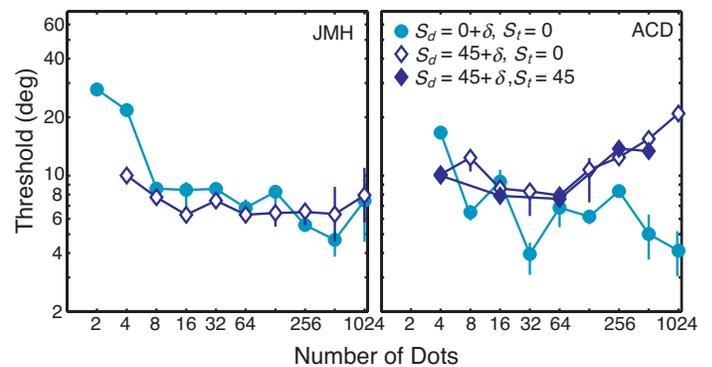


Figure A1. Slant discrimination thresholds for the disparity-alone condition as a function of the number of dots on the surface for the 57.3-cm viewing distance. Filled blue circles represent discrimination thresholds for a base slant of 0 deg ( $S_d = 0 + \delta$ ;  $S_t = 0$ ). Unfilled diamonds represent thresholds for a base slant of 45 deg when the texture specifies a slant of 0 deg ( $S_d = 45 + \delta$ ;  $S_t = 0$ ). Filled diamonds represent thresholds for a base slant of 45 deg when the texture specifies a fixed slant of 45 deg ( $S_d = 45 + \delta$ ;  $S_t = 45$ ). Error bars are 95% confidence intervals.

disparity-based thresholds were essentially as low as they could be. The results were more complicated for observer ACD. When the disparity-specified slant was 0 deg, her data were quite similar to JMH's. However, when the disparity slant was 45 deg, thresholds decreased from 4-64 dots and then increased with more than 64 dots. One would observe such an effect if the conflicting texture signal, which was not informative for the discrimination task, was given increasing weight with increasing dot number. Thus, ACD may have given some weight to the uninformative texture signal when shown the random-dot stimulus at base slants different from 0 deg. We discuss this point in regard to her two-cue data in [Summary of results](#) in [Discussion](#).

## Appendix B: Validity of single-cue measurements for two-cue experiment

To combine two cues for slant, the cues must be promoted to the same units. Disparity signals alone do not provide a slant estimate because they must be scaled or normalized for distance (Gårding et al., 1995). We were concerned that observers might perform the slant-discrimination task in the single-cue, disparity-alone case by comparing only the disparity gradients in the two stimulus intervals. Said another way, they could in principle perform the task without normalizing the disparity signals into slant estimates. To test the MLE model, we must acquire valid measures of the reliabilities of single-cue *slant* estimates. In the disparity-alone condition, this means our measure must reflect the process of scaling the disparity signal into units of slant. If the task in the disparity-alone condition were done without normalizing the disparity signal, the psychometric data would not reflect errors introduced by the scaling process (which, within the framework of weighted-linear cue combination, is essential for combining disparity and texture signals), and we would underestimate the variance of disparity-based *slant* estimates. For this reason, we looked for evidence that observers scale the disparity signal for distance in a discrimination task with our disparity-alone stimuli. We did so by having observers perform the slant-discrimination task with the disparity-alone stimulus with the comparison stimuli appearing at different distances relative to the standard stimulus.

Five observers participated. Two were unaware of the experimental hypotheses. All had normal stereopsis and did not manifest eye misalignment in normal viewing situations.

A different apparatus was used than the one in the main experiment. Stimuli were displayed on a CRT at a distance of 57 cm. Dichoptic presentation of the left and right eye's images was achieved using CrystalEyes™ liquid-crystal shutter glasses. Left- and right-eye images were displayed on alternate frames so each eye's image was drawn only when the corresponding shutter was open. The moni-

tor refresh rate was 100 Hz, so each eye's image was redrawn at 50 Hz. The stimuli were drawn using the red phosphor only because this minimized cross-talk through the shutter glasses. The room was otherwise dark. Precise reproduction of visual directions was achieved using the same anti-aliasing and spatial calibration techniques as in the main experiment. The same bite-bar set up was used to position and stabilize the observer's head.

The stimuli were virtual planes slanted about a vertical axis. They were very similar to the stimuli used in the disparity-alone condition in the main experiment ([Figure 1, top](#)): random-dot stereograms with the same dot density and size as in the main experiment. Stimulus size for a given presentation was drawn randomly from a uniform distribution from 12.5-17.5 deg. The stimuli were generated taking each observer's interpupillary distance into account. Changes in simulated distance were achieved by shifting the two eyes' images laterally on the CRT; this technique produces the correct eye vergence and horizontal gradient of vertical disparity for the simulated distance. The size of the stimulus, dot density, and dot size were the same in angular terms at each simulated distance.

As in the main experiment, two stimuli were presented sequentially and observers indicated the one containing the apparently greater slant. No feedback was provided. The standard stimulus was always presented at a simulated distance of 57 cm, and the comparison stimulus was presented at one of several simulated distances (selected randomly before each trial). Thus, observers had to judge relative slant even when the standard and comparison were presented at different simulated distances. There were three comparison distances (45.2, 57.0, and 71.8 cm) and two base slants ( $\pm 30$  deg). Each stimulus was presented for 1 s with an interstimulus interval of 1.5 s. To facilitate fusion, a fixation marker appeared 500 ms prior to each stimulus presentation at the distance of the upcoming stimulus. Observers were all able to make a vergence eye movement and fuse the fixation marker before the stimulus appeared. 1-down/2-up and 2-down/1-up staircases were used to vary the slant of the comparison stimulus. For each base-slant/distance combination, both reversal rules were used to sample points on the psychometric function either side of the 50% point. At least four staircases were employed for each psychometric function, corresponding to approximately 200 trials per function. We fit the resulting psychometric data with a cumulative Gaussian using a maximum-likelihood criterion.

In [Figure B1](#), we plot the slants of the comparison stimulus that had the same perceived slant on average as the standard stimulus as a function of the simulated distance of the comparison. There were no significant differences between base slants of  $-30$  and  $+30$  deg, so we pooled the data from the two slants. If observers performed the task by comparing slants (as they were instructed), they would have to take distance into account ([Equation 11](#)). If they did so veridically, the PSEs would all have the same slant as the standard stimulus (horizontal dashed line). If,

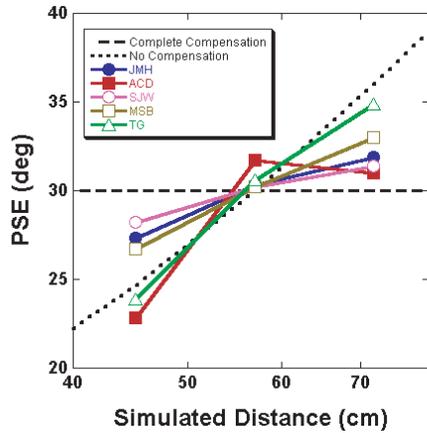


Figure B1. PSEs as a function of distance. The slant of the comparison stimulus that was on average perceived as the same slant as a standard at  $\pm 30$  deg slant and a distance of 57 cm is plotted as a function of the distance of the comparison stimulus. The dashed horizontal line represents the predicted PSEs if observers compensated for the change in distance veridically. The dotted diagonal curve represents the predicted PSEs if observers made their judgments from the disparity gradients of the two stimuli only; that is, if they failed to compensate for distance. The filled symbols are the data from the two observers whose data are presented in the main experiment. The unfilled symbols are the data from three who did not participate in the main experiment.

on the other hand, observers performed the task by comparing disparity gradients, they would not need to take distance into account. In this case, the PSEs would have the same disparity gradients (*HSRs*), but different slants (diagonal dotted line).

The PSEs for four of the five observers were not consistent with the no-compensation predictions; thus, they *did* take the change in stimulus distance into account. However, none exhibited complete compensation, which means that the distance compensation was not veridical. The data from JMH and ACD, the observers who participated in all conditions in the main experiments, are represented by filled symbols. The results of this control experiment show that most observers (importantly including observers JMH and ACD) do not perform the slant-discrimination task by only comparing the disparity gradient in the two stimulus intervals.

## Appendix C: Scaling of disparity signal for distance

To generate the PSE predictions in Figures 8-10, we assumed that slant estimates from texture and disparity are unbiased. This assumption is invalid for disparity if the distance estimates used to “scale” *HSRs* were biased. Other reports (Johnston, 1991; Johnston et al., 1993; Bradshaw, et al., 1996) and our control experiment (Appendix B) suggest that observers do not scale the disparity signal veridi-

cally for distance; that is, they tend to use a farther estimate than the actual distance for near viewing and a near estimate than the actual distance for far viewing. As we pointed out in the Discussion (Comparison to other studies), this result could be due to the influence of unmodeled cues. Here we consider the possibility that the observed changes in PSEs with distance resulted from mis-scaling of the *HSR* signal rather than changes in cue weighting with changes in distance.

We considered two models of distance scaling. (1)  $\hat{d} = C$ ; one fixed distance is used to scale the *HSR* signal at all distances. (2)  $\hat{d} = \alpha \cdot d + \beta$ ; distance estimates are a linear function of distance. There were three main steps to fitting the distance-scaling models: (1) calculate slant estimates from disparity,  $\hat{S}_d$ , from the modeled distance estimate and *HSRs* for the cue-conflict and no-conflict PSE stimuli; (2) calculate the cue-combined slant estimate (assuming  $\hat{S}_i$  was unbiased) for the PSE and conflict stimuli using Equations 1 and 2 and the weights determined from fits in Figure 6; (3) find the parameters for the distance model that yield the smallest squared difference between the cue-combined slant estimates for the cue-conflict and PSE stimuli.

Here we work through an example of these calculations. Consider JMH’s PSE data for a standard slant of 30 deg (second row of Figure 10) and texture perturbed by +10 deg (rightmost black diamonds in the three panels of the second row;  $S_d = 30$  deg and  $S_t = 40$  deg). *HSRs* for cue-conflict stimulus at 19.1, 57.3, and 171.9 cm were 1.22, 1.07, and 1.02, respectively. If a single distance estimate were used to scale these *HSRs*, the slant estimates derived from the disparity cue (which objectively specifies 30 deg at the three distances) would be quite different. For example, if a distance of 57.3 cm were used in all three cases ( $\hat{d} = C$ ), the perceived slant from disparity  $\hat{S}_d$  would be 60, 30, and 10.9 deg, respectively. To get the combined slant estimate, we used Equation 1 with the weights determined from the fits to the 57.3-cm data in Figure 6 (in other words, the weights were determined only by *HSR* and the texture gradient and not by distance). The perceived slants for the cue-conflict stimuli at these three distances would be  $\hat{S}_{19}^c$ ,  $\hat{S}_{57}^c$ , or  $\hat{S}_{172}^c$ . The objective slants of the no-conflict stimuli that appeared, on average, the same as the conflict stimuli were 34.1, 37.8, and 40.6 deg; these correspond to *HSRs* of 1.26, 1.09, and 1.03. If these *HSRs* were scaled by the distance estimate of 57.3 cm,  $\hat{S}_d$ , would be 63.8, 37.8, and 15.9 deg. The cue-combined slant estimate for these no-conflict stimuli would be  $\hat{S}^{nc} = \hat{S}_{19}^{nc}$ ,  $\hat{S}_{57}^{nc}$ , or  $\hat{S}_{172}^{nc}$ . The error, minimized across all conditions simultaneously, was  $\epsilon = \sum_i (\hat{S}_i^c - \hat{S}_i^{nc})^2$ .

Both models of distance scaling provided reasonable fits to the PSE data. We can, however, rule out the  $\hat{d} = C$  model from three lines of evidence. (1) In the condition in which texture was perturbed and disparity-specified slant

was 0 deg (black open diamonds in center row of panels in Figures 10 and 11), the  $\hat{d} = C$  model predicts that distance should have no effect on PSEs. According to this model, a single distance estimate is used to scale HSRs, so HSRs of the PSEs should be the same at all three distances. Figure C1 plots HSR as a function of conflict for the  $S_d = 0$ , texture-perturbed condition. There is clearly a systematic effect of distance on the HSRs of the PSEs, which indicates that there was some compensation for distance. (2) For the distance that provided the best fit to the data, a stimulus that had an objective slant of 30 deg would have appeared to have a slant of 65 deg. This is inconsistent with the phenomenology. (3) In a control experiment (Appendix B), we found that observers take distance into account when performing the slant-discrimination task.

Thus, two models provide a good fit to the data reported in these experiments: the optimal cue re-weighting model (with veridical scaling for distance) and the incomplete-scaling model ( $\hat{d} = \alpha \cdot d + \beta$ ). The former has no free parameters and the latter has two free parameters ( $\alpha$  and  $\beta$ ). By Occam's Razor, the optimal weighting model is preferred, but we hasten to point out that the latter cannot be rejected without further experimentation.

### Appendix D: Computing texture slant from two eyes' images

The texture gradient specifies different surface orientations at the two eyes and the differences depend on surface tilt. Here we derive the differences in texture gradients at the two eyes as a function of slant and tilt.

To express surface orientation, we need a coordinate system. The convention in the stereo literature is to place the origin at the Cyclopean eye, the point on the Vieth-Müller Circle, half way between the eyes. The  $x$ -axis is parallel to the interocular axis. The  $z$ -axis lies in the plane of fixation and is perpendicular to the interocular axis. The  $y$ -axis is perpendicular to the other two axes.

Consider a homogeneously textured frontoparallel plane in front of the observer. The eyes are converged on the plane at the midline and the vergence angle is  $\mu$ . The plane's slant is  $S_C = 0^\circ$ , and its tilt  $\tau_C$  is undefined. From the viewpoint of the left eye (equivalent to placing the origin there),

$$S_L = \frac{\mu}{2} \text{ and } \tau_L = 0^\circ. \tag{D1}$$

From the right eye's viewpoint,

$$S_R = -\frac{\mu}{2} \text{ and } \tau_R = 0^\circ \tag{D2}$$

or, equivalently,

$$S_R = \frac{\mu}{2} \text{ and } \tau_R = 180^\circ. \tag{D3}$$

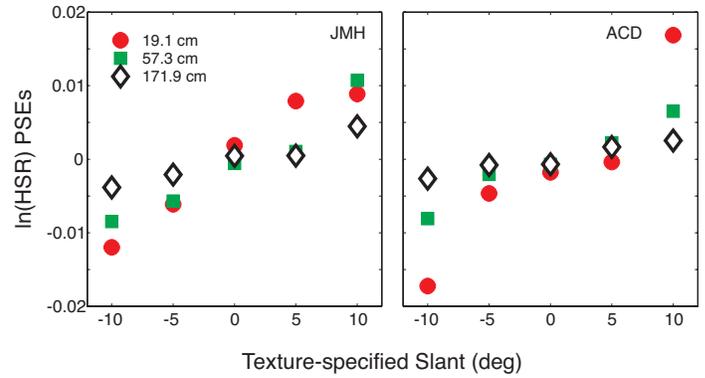


Figure C1. HSRs for PSEs as a function of the texture-specified slant. These data are from the condition in which  $S_d = 0$  and texture was perturbed. Left panel: observer JMH. Right panel: observer ACD.

Hence, the texture gradient specifies different surface orientations at the two eyes. The disparity-specified slant is, of course, 0 deg with the origin at the Cyclopean eye. The texture-gradient signals must be converted into the same Cyclopean coordinates as the disparity signal. The transformation required to convert the texture-gradient signals from left- and right-eye coordinates into Cyclopean coordinates is a potential source of error. That error is presumably not measured in the single-cue, texture-alone condition because monocular stimuli were used in that condition. Here, in Figure D1, we examine the properties of the required transformation and show that they vary with tilt.

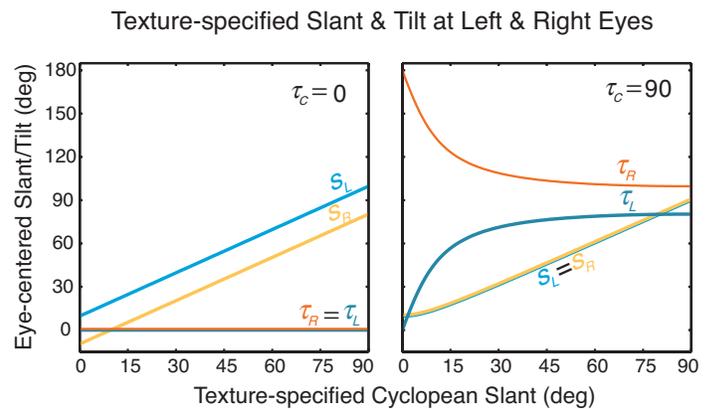


Figure D1. Monocular texture-specified slants and tilts as a function of Cyclopean slant for a 19-cm viewing distance and 6.5-cm interpupillary distance. The left panel shows left and right eye slants and tilts when the Cyclopean tilt is 0 ( $\tau_C = 0^\circ$ ). The right panel shows left and right eye slants and tilts when the Cyclopean tilt is 90 deg ( $\tau_C = 90^\circ$ ).

When a frontoparallel plane is rotated about the vertical axis ( $S_C = \Delta$  and  $\tau_C = 0^\circ$ ), the texture-specified slants and tilts in the left and right eyes vary, but the tilts do not. Technically, for rotations about a vertical axis, the tilt pa-

parameter for eye-centered viewpoints flips between 0 and 180°. This corresponds to a sign change in slant. For simplicity, we used a sign change in slant, rather than changing the tilt parameter, for vertical-axis rotations. The eye-centered slants specified by the texture gradients are

$$S_L = S_C + \frac{\mu}{2} \text{ and } S_R = S_C - \frac{\mu}{2}, \quad (\text{D4})$$

so they differ by  $\mu$ . The eye-centered, texture-specified tilts are  $\tau_L = \tau_R = 0$ . The left panel of Figure A1 shows eye-centered slants and tilts as a function of Cyclopean slant for  $\tau_C = 0^\circ$ .

Now consider a frontoparallel plane rotated about the horizontal axis ( $S_C = \Delta$  and  $\tau_C = 90^\circ$ ). For simplicity, we now use positive slants and let the tilt flip from 0 to 180° when necessary. The texture-specified slants are equal in the two eyes:

$$S_L = S_R = \cos^{-1}[\cos(\frac{\mu}{2})\cos(S_C)]. \quad (\text{D5})$$

But the tilts differ:

$$\tau_L = \cos^{-1}\left[\frac{\sin(\frac{\mu}{2})}{\sin(S_L)}\right] \text{ and } \tau_R = \cos^{-1}\left[\frac{\sin(-\frac{\mu}{2})}{\sin(S_R)}\right]. \quad (\text{D6})$$

The right panel of Figure A1 shows these eye-centered slants and tilts.

When the Cyclopean tilt is between 0 and 90 deg, the eye-centered slants and tilts will both differ.

In summary, the transformation required to convert texture signals into Cyclopean coordinates depends on tilt. This fact raises the interesting possibility that different errors occur in that transformation for one tilt as opposed to another. Perhaps an explanation for slant anisotropy could be derived from an understanding of these transformations.

## Acknowledgments

This research was supported by National Institutes of Health Research Grant EY12851 (MSB), Air Force Office of Scientific Research Grant F49620-01-1-0417 (MSB), and National Institutes of Health Research Grant EY08266 (MSL). We thank Alison Dilworth for spending many hours as an observer. Part of this work was presented at the European Conference on Visual Perception in 2002.

Commercial relationships: none.

Corresponding author: Martin S. Banks.

Email: marty@john.berkeley.edu.

Address: Vision Science Program, Department of Psychology, & Wills Neuroscience Institute, University of California, Berkeley, CA 94720, USA.

## References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257-262. [PubMed]
- Backus, B. T., & Banks, M. S. (1999). Estimator reliability and distance scaling in stereoscopic slant perception. *Perception*, *28*, 217-242. [PubMed]
- Backus, B. T., Banks, M. S., van Ee R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, *39*, 1143-1170. [PubMed]
- Banks, M. S., Gepshtein, S., & Landy, M. S. (2004). Why is spatial stereoresolution so low? *Journal of Neuroscience*, *24*, 2077-2089. [PubMed]
- Banks, M. S., Hooge, I. T. C., & Backus, B. T. (2001). Perceiving slant about a horizontal axis from stereopsis. *Journal of Vision*, *1*(2), 55-78, <http://journalofvision.org/1/2/1/>, doi:10.1167/1.2.1. [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*, *20*, 1391-1397. [PubMed]
- Blake, A., Bülthoff, H. H., & Sheinberg, D. (1993). Shape from texture: Ideal observers and human psychophysics. *Vision Research*, *33*, 1723-1737. [PubMed]
- Bradshaw, M. F., Glennerster, A., & Rogers, B. J. (1996). The effect of display size on disparity scaling from differential perspective and vergence cues. *Vision Research*, *36*, 1255-1264. [PubMed]
- Bradshaw, M. F., Hibbard, P. B., van der Willigen, R., Watt, S. J., & Simpson, P. J. (2002). The stereoscopic anisotropy affects manual pointing. *Spatial Vision*, *15*, 443-458. [PubMed]
- Buckley, D., & Frisby, J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Research*, *33*, 919-933. [PubMed]
- Burt, P., & Julesz, B. (1980). A disparity gradient limit for binocular fusion. *Science*, *208*, 615-617. [PubMed]
- Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society*, *4*(Suppl.), 102-118.
- 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]
- de Berg, M., van Kreveld, M., Overmars, M., & Schwarzkopf, O. (2000). *Computational geometry: Algorithms and applications* (2nd Ed.). New York: Springer-Verlag.

- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429-433. [PubMed]
- Frisby, J. P., Buckley, D., & Horsman, J. M. (1995). Integration of stereo, texture, and outline cues during pinhole viewing of real ridge-shaped objects and stereograms of ridges. *Perception*, *24*, 181-198. [PubMed]
- Gårding, J., Porrill, J., Mayhew, J. E. W., & Frisby, J. P. (1995). Stereopsis, vertical disparity and relief transformations. *Vision Research*, *35*, 703-722. [PubMed]
- Gepshtein, S., & Banks, M. S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology*, *13*, 483-488. [PubMed]
- Gillam, B., & Ryan, C. (1992). Perspective, orientation disparity, and anisotropy in stereoscopic slant perception. *Perception*, *21*, 427-439. [PubMed]
- Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. New York: Robert E. Krieger.
- Heath, G. G. (1956). The influence of visual acuity on accommodative responses of the human eye. *American Journal of Optometry*, *33*, 513-523. [PubMed]
- Helmholtz, H. (1910). *Physiological optics*. New York: Dover.
- Hillis, J. M., & Banks, M. S. (2001). Are corresponding points fixed? *Vision Research*, *41*, 2457-2473. [PubMed]
- Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science*, *298*, 1627-1630. [PubMed]
- Howard, I. P., & Kaneko, H. (1994). Relative shear disparities and the perception of surface inclination. *Vision Research*, *34*, 2505-2517. [PubMed]
- Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth. Volume 2: Depth perception*. Toronto: I Porteous.
- Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*, 3621-3629. [PubMed]
- Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research*, *31*, 1351-1360. [PubMed]
- Johnston, E. B., Cumming, B. G., & Parker, A. J. (1993). Integrations of depth modules: Stereopsis and texture. *Vision Research*, *33*, 813-826. [PubMed]
- Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271-304. [PubMed]
- Knill, D. C. (1998a). Surface orientation from texture: Ideal observers, generic observers and the information content of texture cues. *Vision Research*, *38*, 1655-1682. [PubMed]
- Knill, D. C. (1998b). Discrimination of planar surface slant from texture: Human and ideal observers compared. *Vision Research*, *38*, 1683-1711. [PubMed]
- 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]
- Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*, 244-247. [PubMed]
- Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A*, *18*, 2307-2320. [PubMed]
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. J. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, *35*, 389-412. [PubMed]
- Legge, G. E. (1984). Binocular contrast summation I. Detection and discrimination. *Vision Research*, *24*, 373-383. [PubMed]
- Mamassian, P., Landy, M. S., & Maloney, L. T. (2002). Bayesian modeling of visual perception. In R. P. N. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabilistic models of the brain: perception and neural function* (pp. 13-36). Cambridge, MA: MIT Press.
- Mather, G., & Smith, D. R. (2002). Blur discrimination and its relation to blur-mediated depth perception. *Perception*, *31*, 1211-1219. [PubMed]
- Ogle, K. N. (1950). *Researches in binocular vision*. Philadelphia, PA: W. B. Saunders.
- Orcu, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*, 2451-2468. [PubMed]
- Rogers, B. J., & Bradshaw, M. F. (1995). Disparity scaling and the perception of frontoparallel surfaces. *Perception*, *24*, 155-179. [PubMed]
- van Beers, R. J., Sittig, A. C., & Denier van der Gon, J. J. (1998). The precision of proprioceptive position sense. *Experimental Brain Research*, *122*, 367-377. [PubMed]
- van Beers, R. J., Wolpert, D. M., & Haggard, P. (2002). When feeling is more important than seeing in sensorimotor adaptation. *Current Biology*, *12*, 834-837. [PubMed]
- van Ee, R., & Erkelens, C. J. (1998). Temporal aspects of stereoscopic slant estimation: An evaluation and extension of Howard and Kaneko's theory. *Vision Research*, *38*, 3871-3882. [PubMed]
- 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]